Not Always Buried Deep

Selections from Analytic and Combinatorial Number Theory

Paul Pollack

2003 Summer Course Notes
Ross Summer Mathematics Program

First Draft: June 9, 2003
Second Draft: November 16, 2003
Last updated: May 9, 2004

In memory of Professor Arnold Ephraim Ross (1906-2002),
whose example continues to be a teacher for all of us.

Preface

> The gold in 'them there hills' is not always buried deep. Much of it is within easy reach. Some of it is right on the surface to be picked up by any searcher with a keen eye for detail and an eagerness to explore. As in any treasure hunt, the involvement grows as the hunt proceeds and each success whether small or great adds the fuel of excitement to the exploration. – Arnold Ross, *Prolog* (1978)

This manuscript grew out of notes for an eight-week summer course offered to counselors at the 2003 Ross Summer Mathematics Program.

As the title suggests, all of the topics discussed lie close to the surface in that their study requires no significant technical preparation. Solid undergraduate courses in number theory, abstract algebra and single-variable calculus should enable one to appreciate bulk of the text.

Outline of Contents

Our first chapter treats elementary prime number theory in $\mathbf{Z}, \mathbf{Z}[i]$ and $\mathbf{F}_q[T]$. We begin with a topical survey of several proofs of the infinitude of the primes (over $\mathbf{Z}$). We then digress briefly to discuss how Gauss was led to conjecture the prime number theorem (whose proof we defer to Chapter 4) before we tackle some of the simplest estimates for $\pi(x)$, such as Euler's estimate $\pi(x) = o(x)$. The next two sections are devoted to the work of Chebyshev and Mertens, including Chebyshev's determination of $x/\log x$ as the correct order of magnitude of $\pi(x)$ and Ramanujan's proof of the asymptotic Bertrand's postulate. Some analogous results in $\mathbf{Z}[i]$ are discussed before we turn our attention to prime number theory in the ring of polynomials over a finite field. There we focus on a classical analog of the prime number theorem, due essentially to Gauss, and a recently established analog of the twin prime conjecture, due (for $q \neq 3$) to C. Hall [Hal03, Corollary 19]: if $\alpha \in \mathbf{F}_q^*$, where $q > 2$, then there are infinitely many twin prime pairs $h, h + \alpha \in \mathbf{F}_q[T]$.

Our heuristic approach to the prime number theorem also motivates several as-yet unproved conjectures, such as the Goldbach and twin prime conjectures. Such consequences are explored in the final section of Chapter 1. A further example is the conjecture of Schinzel [SS58] that if $f_1, \dots, f_k \in \mathbf{Z}[T]$ are distinct irreducible polynomials (with positive leading coefficient), then there are infinitely many integers n for which each $f_i(n)$ is prime, unless there is a "local obstruction," i.e., a prime p dividing the product $f_1(n) \dots f_k(n)$ for every value of n. We adapt our heuristic to motivate a stronger, quantitative form of this conjecture put forward by Bateman & Horn [BH62], special cases of which had previously been examined by Hardy & Littlewood.

The special case of Schinzel's conjecture when $k = 1$ and f_1 is linear is a classical theorem of Dirichlet whose proof is taken up in Chapter 2. The theorem is well known to follow from the nonvanishing of $L(1, \chi)$ for every character χ of $\mathbf{Z}/q\mathbf{Z}^*$, where $L(s, \chi)$ is the associated Dirichlet series $\sum_n \chi(n)n^{-s}$. Both this implication and the stated non-vanishing are usually proved using tools from complex analysis, but we follow H.N. Shapiro [Sha50] in giving elementary proofs of these facts. The major point of departure from Shapiro's treatment is our proof of the non-vanishing of $L(1, \chi)$ for real χ, which follows P. Monsky [Mon93]. We conclude the chapter with the characterization of the integers expressible as a sum of three squares, following Ankeny [Ank57] and Mordell [Mor58].

Chapter 3 discusses elementary sieve methods in number theory. We begin with a discussion of the sieve of Eratosthenes and some corollaries, e.g., that almost all numbers do *not* admit a representation as a sum of two squares. We then turn to Viggo Brun's elementary "pure sieve," which is used to deduce that the twin primes are sparse in the set of all primes, in that $\pi_2(x) = O(x(\log\log x)^2/\log^2 x)$. One famous consequence is that $\sum_p 1/p < \infty$, where p runs over the twin primes. The strong form of the twin prime conjecture would imply the more precise result $\pi_2(x) = O(x/\log^2 x)$, which was obtained unconditionally by Brun and which we establish by a recent method Hooley [Hoo94]. The Brun-Hooley sieve also permits lower estimates, and in this way we prove the following two approximations to the Goldbach and twin prime conjectures:

i. every large even N is the sum of two positive integers, neither of which has more than 7 prime factors,

ii. there are infinitely many positive even integers N for which $N, N+2$ each have no more than 7 prime divisors (counted with multiplicity).

Brun obtained results of this form (but with the worse constant 9 in place of 7) in 1919 ([Bru19], [Bru20]), but by a much more complicated method.

In Chapter 4 we keep a promise made in Chapter 1: we describe an elementary proof of the prime number theorem, following A.J. Hildebrand [Hil86]. This is one of only two elementary routes to the PNT known that differs essentially from the original 1949 Erdős-Selberg approach. In order to keep the treatment self-contained, we include Landau's proof that the prime number theorem is equivalent to the estimate $\sum_{n \le x} \mu(n) = o(x)$ (which is what we directly establish), a proof of the Turán-Kubilius inequality, and a brief account of Selberg's sieve (needed to derive an upper bound for the number of primes in a short interval).

At this point, we switch focus from analytic number theory to additive and combinatorial number theory.

We commence our study anew in Chapter 5, with a survey of results from additive number theory. If $\mathcal{A}_1, \ldots, \mathcal{A}_h$ are subsets of an additive semigroup S, we define

$$\mathcal{A}_1 + \cdots + \mathcal{A}_h := \{a_1 + \cdots + a_h : a_i \in \mathcal{A}_i\}.$$

The set S is called an additive basis of finite order if $h\mathcal{A} = S$ for some positive integer h, where $h\mathcal{A}$ denotes the h-fold sumset. S is called an asymptotic basis of finite order if $S \setminus h\mathcal{A}$ is finite for some positive integer h. We will be particularly concerned with the cases when $S = \mathbf{Z}/m\mathbf{Z}$ and $S = \mathbf{Z}$; in these situations we attack the problem by studying how the size (respectively "thickness," appropriately defined) of a sumset $\mathcal{A} + \mathcal{B}$ compares to that of the summands. In these respective cases, we prove the important sumset theorems of Cauchy, Davenport, Chowla [Cho35] and Mann [Man42]. We then combine our results on additive bases with a sieve result from Chapter 3 to prove Schnirelmann's theorem [Sch33] that every integer $n > 1$ can be written as the sum of at most C primes for some absolute constant C.

An intermediate result obtained in the proof of Schnirelmann's theorem is that $\{p+q : p, q \text{ prime}\}$ has positive lower density. We prove the same for the set $\{p + a^k : p \text{ prime}, k \geq 1\}$, for every fixed integer $a \geq 2$. This is due to Romanov [Rom34]. When $a = 2$, Romanov asked whether this set contained all sufficiently large odd numbers. This was answered in the negative by Erdős [Erd50], who exhibited an infinite arithmetic progression of counterexamples. We discuss his results and a related theorem of Crocker before turning to our final topic, Schur's regularity lemma. Schur proved [Sch16] that if the set $\{1, 2, \ldots, \lfloor k!e \rfloor\}$ is partitioned into k subsets, then one of the subsets contains a solution to $x - y = z$. While appearing as a lemma in a paper on Fermat's Last Theorem, Schur's theorem is a combinatorial gem and an important ancestor of the results discussed in the next chapter.

Chapter 6 treats the subject of sequences without arithmetic progressions. The founding result in this area is the theorem of van der Waerden: in any finite partition of the positive integers, (at least) one set in this partition contains arbitrarily long arithmetic progressions [vdW27]. We present a short proof of this, due to Graham & Rothschild [GR74], as exposited by Pomerance & Sárközy [PS95]. One might suspect what is behind the truth of van der Waerden's theorem is that in any finite partition of the positive integers into k sets, (at least) one of the sets is "thick," in fact has upper density not less than $1/k$. This suspicion was confirmed by Szemerédi [Sze75], who proved that any subset of the positive integers with positive upper density contains arbitrarily long arithmetic progressions. The easier special case of 3-term progressions had been treated earlier by K.F. Roth [Rot52], and we present two proofs of Roth's theorem. The first argument is purely combinatorial and is due to Szemerédi; our treatment follows Graham [Gra81, Chapter 5]. The second argument is analytic, being an instance of the circle method, and in the form presented here is due to D.J. Newman ([New81], [New98]).

In the opposite direction, one can ask for a lower bound on $r_3(n)$, defined as the size of the largest subset of $\{1, 2, \ldots, n\}$ with no three terms in arithmetic progression. Behrend [Beh46] (improving on earlier work of Salem & Spencer [SS42]) proved

$$r_3(n) \geq N^{1-(2\sqrt{2\log 2}+o(1))/\sqrt{\log N}},$$

and apart from the constant $2\sqrt{2\log 2}$ this remains the best known result. The closing of Chapter 6 is devoted to a presentation of Behrend's original elegant construction.

The final chapter marks our return to the realm of additive number theory, but our energies are now concentrated on a single problem: Waring's conjecture that for every positive integer k, the set of nonnegative kth powers forms an additive basis of finite order (for $\mathbf{N}$). The proof we present is an adaptation of the Hardy-Littlewood circle method (see [Vau97, Chapter 2]) due to D.J. Newman ([New60], [New98, Chapter V]). The circle method is usually used to establish an asymptotic formula for the number of representations of an integer as a sum of s kth powers, valid as $n \to \infty$ for fixed $s > s_0(k)$. From this asymptotic formula, Waring's conjecture is immediate. However, we take an easier route, proving only a rough upper bound for the number of representations which, in combination with Schnirelmann's results from Chapter 5, is enough to imply the conjecture.

The Appendix reviews definitions and results from asymptotics; familiarity with this material is assumed throughout the text.

Notation and Conventions

Much of our notation is standard and should be familiar to students of elementary number theory. However, the reader should be aware that we take the set of natural numbers $\mathbf{N}$ to include 0; the positive integers are denoted by $\mathbf{Z}^+$.

By an *arithmetic function* we mean a function $f : \mathbf{Z}^+ \to \mathbf{C}$. An arithmetic function is said to be *multiplicative* if f is not identically 0 and

$$f(ab) = f(a)f(b) \qquad \text{whenever } \gcd(a,b) = 1.$$

Of the two common notations for the number of divisors of n, we choose $\tau(n)$. We also use $\nu(n)$ to denote the number of distinct prime divisors of n, since the usual notation $\omega(n)$ is needed for a different purpose in Chapter 3.

The notation $f = O(g)$ means that $|f| \leq C|g|$ for some constant C. An equivalent but more suggestive notation, due to Vinogradov, is $f \ll g$. The notation $f \gg g$ means $g \ll f$. If $f \ll g$ and $g \ll f$, we say that f and g have *the same order of magnitude* and write $f \asymp g$. The notation $f = o(g)$ as $x \to a$ (respectively $f \sim g$ as $x \to a$) means that $\lim_{x\to a} f(x)/g(x) = 0$ (respectively $= 1$); if no a is specified, then $a = \infty$ is assumed. See Appendix A for more details.

Subsets of $\mathbf{Z}, \mathbf{Z}/m\mathbf{Z}$ or $\mathbf{F}_q$ are often denoted with script letters, e.g., $\mathcal{A}$. The number of *positive* elements of a subset of $\mathbf{Z}$ that do not exceed N (the so-called *counting function* of the set) is denoted with the corresponding roman letter, e.g.,

$$A(N) := |\{1 \leq n \leq N : n \in \mathcal{A}\}|.$$

We define the *lower* and *upper asymptotic densities* of $\mathcal{A} \subset \mathbf{Z}$ by

$$\underline{d}(\mathcal{A}) := \liminf_{x\to\infty} A(x)/x, \quad \overline{d}(\mathcal{A}) := \limsup_{x\to\infty} A(x)/x; \tag{0.1}$$

if $\underline{d}(\mathcal{A}) = \overline{d}(\mathcal{A})$, then the common value is referred to as the *natural density* and denoted $d(\mathcal{A})$. Note that since $A(x) = A(\lfloor x \rfloor)$, one obtains equivalent definitions if x in (0.1) is restricted to the set of positive integers, which we shall sometimes assume. A statement about positive integers is said to hold for "almost all n" if the set of exceptions has density 0.

Sums and products indexed over p are restricted to primes p, unless otherwise specified. The condition $n \leq x$ and its variants indicates that the sum (or product) is taken over all *positive* integral $n \leq x$. The empty sum and the empty product are taken to be 0 and 1, respectively.

A *partition* of a set is a decomposition into disjoint subsets; we do *not* require that the component sets be nonempty.

A Word on the Exercises

There are several exercises scattered throughout the text, usually discussing applications or extensions of the main results. Exercises with a common theme are often grouped into their own subsection, while an isolated exercise or two is often left in the subsection treating the relevant subject matter. Most of the exercises are *not* strictly necessary for an understanding of the text, and those that are have been explicitly marked.

Acknowledgements

I would like to thank Matthew Baker for allowing me to produce a preliminary version of these notes as my senior thesis. His numerous suggestions along the way helped shape these notes into their present form.

Andrew Granville is responsible for stirring my interest in these subjects during my early undergraduate years, when I was his "lab assistant" as part of the Undergraduate Research Apprenticeship Program at UGA. In addition to teaching me much of the number theory I know, the flavor, arrangement and selection of the material of Chapter 1 largely derives from two courses I took with him as an undergraduate. Sections §1.4 and §1.8 (including the exercises therein) are examples of where I have followed Granville particularly closely. An exercise marked with his name but with no further citation indicates that it also derives from one of these courses, or from personal communication.

Keith Conrad kindly pointed out errors in an earlier draft of Chapter 1. Several mistakes were also caught by the students who attended the eight-week course these notes were written for; all of them have my sincere thanks.

On the TEXnical side of things, I am grateful for the experience typesetting garnered at A.J. Hildebrand's 2002 UIUC Number Theory REU.

References

[Ank57] N. C. Ankeny, *Sums of three squares*, Proc. Amer. Math. Soc. **8** (1957), 316–319. MR 19,15d

[Beh46] F. A. Behrend, *On sets of integers which contain no three terms in arithmetical progression*, Proc. Nat. Acad. Sci. U. S. A. **32** (1946), 331–332. MR 8,317d

[BH62] Paul T. Bateman and Roger A. Horn, *A heuristic asymptotic formula concerning the distribution of prime numbers*, Math. Comp. **16** (1962), 363–367. MR 26 #6139

[Bru19] V. Brun, *Le crible d'Eratosthène et le théorème de Goldbach*, C. R. Math. Acad. Sci. Paris **168** (1919), 544–546.

[Bru20] ——, *Le crible d'Eratosthène et le théorème de Goldbach*, Christiania Vidensk. Selsk. Skr (1920), no. 3, 36pp.

[Cho35] Inder Chowla, *A theorem on the addition of residue classes: Application to the number $\Gamma(k)$ in Waring's problem.*, Proc. Indian Acad. Sci., Sect. A **2** (1935), 242–243 (English).

[Erd50] P. Erdös, *On integers of the form $2^k + p$ and some related problems*, Summa Brasil. Math. **2** (1950), 113–123. MR 13,437i

[GR74] R. L. Graham and B. L. Rothschild, *A short proof of van der Waerden's theorem on arithmetic progressions*, Proc. Amer. Math. Soc. **42** (1974), 385–386. MR 48 #8257

[Gra81] Ronald L. Graham, *Rudiments of Ramsey theory*, CBMS Regional Conference Series in Mathematics, vol. 45, American Mathematical Society, Providence, R.I., 1981. MR 82j:05018

[Hal03] Christopher J. Hall, *L-functions of twisted Legendre curves*, Ph.D. thesis, Princeton University, 2003.

[Hil86] Adolf Hildebrand, *The prime number theorem via the large sieve*, Mathematika **33** (1986), no. 1, 23–30. MR 88a:11085

[Hoo94] C. Hooley, *On an almost pure sieve*, Acta Arith. **66** (1994), no. 4, 359–368. MR 95g:11092

[Man42] Henry B. Mann, *A proof of the fundamental theorem on the density of sums of sets of positive integers*, Ann. of Math. (2) **43** (1942), 523–527. MR 4,35e

[Mon93] Paul Monsky, *Simplifying the proof of Dirichlet's theorem*, Amer. Math. Monthly **100** (1993), no. 9, 861–862. MR 95b:11085

[Mor58] L. J. Mordell, *On the representation of a number as a sum of three squares*, Rev. Math. Pures Appl. **3** (1958), 25–27. MR 23 #A117

[New60] Donald J. Newman, *A simplified proof of Waring's conjecture*, Michigan Math. J **7** (1960), 291–295. MR 22 #10967

[New81] ______, *Sequences without arithmetic progressions*, Analytic number theory (Philadelphia, Pa., 1980), Lecture Notes in Math., vol. 899, Springer, Berlin, 1981, pp. 311–314. MR 83h:10102

[New98] ______, *Analytic number theory*, Graduate Texts in Mathematics, vol. 177, Springer-Verlag, New York, 1998. MR 98m:11001

[PS95] Carl Pomerance and András Sárközy, *Combinatorial number theory*, Handbook of combinatorics, Vol. 1, 2, Elsevier, Amsterdam, 1995, pp. 967–1018. MR 97e:11032

[Rom34] N.P. Romanov, *Über einige Sätze der additiven Zahlentheorie*, Math. Annalen **109** (1934), 668–678.

[Rot52] Klaus Roth, *Sur quelques ensembles d'entiers*, C. R. Acad. Sci. Paris **234** (1952), 388–390. MR 13,724d

[Sch16] I. Schur, *Über die Kongruenz* $x^m + y^m \equiv z^m \pmod{p}$, Jahresber. d. Deutschen Math. Verein. **25** (1916), 114–117.

[Sch33] L.G. Schnirelmann, *Über additive Eigenschaften von Zahlen*, Math. Ann. **107** (1933), 649–690.

[Sha50] Harold N. Shapiro, *On primes in arithmetic progression. II*, Ann. of Math. (2) **52** (1950), 231–243. MR 12,81b

[SS42] R. Salem and D. C. Spencer, *On sets of integers which contain no three terms in arithmetical progression*, Proc. Nat. Acad. Sci. U. S. A. **28** (1942), 561–563. MR 4,131e

[SS58] A. Schinzel and W. Sierpiński, *Sur certaines hypothèses concernant les nombres premiers*, Acta Arith. 4 (1958), 185–208; erratum **5** (1958), 259. MR 21 #4936

[Sze75] E. Szemerédi, *On sets of integers containing no k elements in arithmetic progression*, Acta Arith. **27** (1975), 199–245, Collection of articles in memory of Juriĭ Vladimirovič Linnik. MR 51 #5547

[Vau97] R. C. Vaughan, *The Hardy-Littlewood method*, second ed., Cambridge Tracts in Mathematics, vol. 125, Cambridge University Press, Cambridge, 1997. MR 98a:11133

[vdW27] B. L. van der Waerden, *Beweis einer Baudetschen Vermutung*, Nieuw Arch. Wiskd. **15** (1927), 212–216.

Contents

Part I

Analytic Number Theory

Chapter 1

Elementary Prime Number Theory

> Mathematicians have tried in vain to this day to discover some order in the sequence of prime numbers, and we have reason to believe that it is a mystery into which the human mind will never penetrate. – Leonhard Euler

1.1 Introduction

Few topics in number theory attract more attention, popular or professional, than the theory of prime numbers. It is not hard to see why; the study of the distribution of the primes possesses in abundance the very features that draw so many of us to number theory in the first place: intrinsic beauty, accessible points of entry, and a lingering sense of mystery embodied in numerous unprententious but infuriatingly obstinate open problems.

The first significant result here is, of course, the infinitude of the primes. A considerable portion of this opening chapter is devoted to surveying the many arguments that have been brought to bear on this theme.

After we know the prime counting function $\pi(x)$ tends to infinity, it is natural to wonder if something more precise can be asserted. In the late 1700s and early 1800s, Legendre and Gauss independently conjectured the *prime number theorem*:

$$\pi(x) := |\{p \leq x : p \text{ prime}\}| \sim x/\log x.$$

Gauss' motivation for making this conjecture is described in §1.3.

Though unable to prove the prime number theorem, the 19th century mathematician Chebyshev made several important contributions to prime number theory. He proved that $x/\log x$ was the correct order of magnitude of $\pi(x)$, in that the ratio $\pi(x)\log x/x$ is bounded between two positive constants for $x \geq 2$.

He also showed

$$\liminf \frac{\pi(x)}{x/\log x} \leq 1, \quad \limsup \frac{\pi(x)}{x/\log x} \geq 1,$$

so that *if* $\pi(x)\log x/x$ tends to a limit, then that limit must be 1. Thus, the surprise (and the difficulty!) inherent in the prime number theorem is not in the *rate* at which $\pi(x)$ grows, but in the *regularity* of its growth. Finally, he proved (a strengthening of) *Bertrand's postulate*, that there is always a prime between n and $2n$. After a brief discussion of simpler estimates for $\pi(x)$ in §1.4, we take up Chebyshev's ideas in §1.5, together with an application of Sierpiński to primes represented by quadratic polynomials.

Later in the 19th century, Mertens' took up Chebyshev's results and used them to make rigorous certain earlier assertions of Euler and Gauss. In particular, he obtained quite precise estimates for $\sum_{p\leq x} 1/p$ and $\prod_{p\leq x}(1-1/p)$. These estimates are useful in many contexts, for example in our proof of Dirichlet's Theorem (Chapter 2) and in our discussion of sieve methods (Chapter 3).

The estimates of Chebyshev and Mertens are the zenith in our development of prime number theory over $\mathbf{Z}$. At this point, we leave the realm of what we can prove to be true to discuss what, if the universe is just, *should* be true. For example, modifying Gauss' heuristic motivating the prime number theorem, we derive quantitative versions of the twin prime and Goldbach conjectures.

In the final two sections, we take a brief look at prime number theory in two other contexts: the ring of Gaussian integers and the ring of polynomials over a finite field. In the former case, we content ourselves with stating the analog of the prime number theorem and generalizing the previously established estimates of Chebyshev. In the latter case, we are able to do much more; it is possible to prove the prime number theorem in an elementary manner, and to settle one version of the "twin prime conjecture."

1.2 There are Infinitely Many Primes

Recall that a *prime number* is an integer greater than 1 admitting no nontrivial factorization. Much of classical number theory centers around the study of the counting function

$$\pi(x) := |\{p \leq x : p \text{ prime}\}|.$$

In this chapter we explore some of the more elementary results in this direction, beginning with a deservedly famous theorem from antiquity:

Theorem 1.2.1. *There are infinitely many primes, i.e.,* $\pi(x) \to \infty$ *as* $x \to \infty$.

In this section we describe several methods for proving Theorem 1.2.1. Most of these proofs may be found in [Dic66, Chapter XVIII] or [Nar00, §1.1]. For other compilations, see [Rib96, Chapter 1] and [Moh79]. Some of the methods discussed here yield the stronger result that $\sum 1/p$ diverges; other elementary proofs of that fact may be found in [Bel43], [Mos58], and the survey [VE80].

1.2.1 Euclid and his Imitators

We begin with the well-known classical proof found in Euclid's *Elements* (from circa 300 BC):

Proof. Suppose $p_1, p_2, \ldots, p_n$ is any finite list of primes. Let P denote the product of the p_i and consider the integer $P+1$. Since $P+1 \equiv 1 \pmod{p_i}$ for each prime p_i in our list, $P+1$ cannot be divisible by any of the p_i. But as an integer larger than 1, it has some prime divisor p. It follows that there is always a prime missing from any finite list, or as Euclid put it, prime numbers are more than any assigned multitude of primes. □

There are many trivial variants; for instance, we can easily show that for every integer m there is a prime $p > m$ by taking as p any prime divisor of $m! + 1$.

We will give several similar proofs in this section: all of these begin with a finite list of primes and produce an integer > 1 that is coprime to all of them. Stietjes' proof is a quintessential example:

Stieltjes' proof, 1890. Suppose $p_1, \ldots, p_n$ is a finite list of distinct primes with product P and let $P = AB$ be any decomposition of P into positive factors.

Let p be one of the p_i. Then p divides AB, so that either $p \mid A$ or $p \mid B$. If p were to divide both, then p^2 would divide P, and this would violate unique factorization; consequently p divides exactly one of A and B. It follows that $p \nmid A+B$. So $A+B$ is divisible by none of the p_i; but as $A+B \geq 2$, it has some prime divisor. So again we've discovered a prime not on our original list. □

Euler's second proof (published posthumously). The proof is based on the multiplicativity of the ϕ-function; Let $p_1, \ldots, p_n$ be a list of distinct primes with product P. By said multiplicativity,

$$\phi(P) = \prod (p_i - 1) \geq 2^{n-1} \geq 2,$$

provided our list contains at least two primes (as we may assume). This inequality says there exists an integer in the range $[2, P]$ that is coprime to P, but such an integer has a prime factor necessarily different from any of the p_i. □

Proof of Braun (1897), Métrod (1917). Let $p_1, \ldots, p_n$ be a list of $n \geq 2$ distinct primes and let $P = p_1 p_2 \cdots p_n$ as usual. Consider the integer

$$N := P/p_1 + P/p_2 + \cdots + P/p_n.$$

For each i, $1 \leq i \leq n$,

$$N \equiv P/p_i = \prod_{j \neq i} p_j \not\equiv 0 \pmod{p_i},$$

so that N is divisible by none of the p_i. But $N \geq 2$, so must possess a prime factor not on our list. □

Exercise 1.2.1. Adapt Euclid's proof of the infinitude of primes to demonstrate that for every integer $m \geq 3$, there exist infinitely many primes $p \not\equiv 1 \pmod{m}$.

Exercise 1.2.2 *(A. Granville, cf. [Has50, p. 168]).* Generalizing the result of the previous exercise, show that if H is a proper subgroup of $\mathbf{Z}/m\mathbf{Z}^*$, then there exist infinitely many primes p with $p \pmod{m} \notin H$.

1.2.2 Coprime Integer Sequences

Suppose we know an infinite sequence of pairwise relatively prime positive integers

$$2 \leq n_1 < n_2 < \dots.$$

Then we may define a sequence of primes p_i by selecting arbitrarily a prime divisor of the corresponding n_i; the terms of this sequence are pairwise distinct because the n_i are pairwise coprime.

If we can exhibit such a sequence of n_i without invoking the infinitude of the primes, we have a further proof of Theorem 1.2.1. This was accomplished by Goldbach:

Proof (Goldbach, 1730*).* Let $n_1 = 3$, and inductively define

$$n_i = 2 + \prod_{1 \leq j < i} n_j$$

for $i > 1$. Then the following assertions are all easily verified in succession:

i. Each n_i is odd.

ii. When $j > i$, $n_j \equiv 2 \mod n_i$.

iii. We have $\gcd(n_i, n_j) = 1$ for $i \neq j$.

Theorem 1.2.1 now follows from the above remarks. □

A straightforward induction shows

$$n_i = 2^{2^{i-1}} + 1, \tag{1.1}$$

and this is how Goldbach presented the proof. Of course we could have chosen the n_i in a different fashion above, taking for instance a different value for n_1 and/or adding 1 instead of adding 2. See Exercise 1.2.3.

Before proceeding, we pause to note that the above proof implies more than simply the infinitude of the primes. First, it gives us an upper bound for the nth prime: $2^{2^{n-1}} + 1$; this translates into a lower bound of the shape

$$\pi(x) \gg \log\log x \quad (x \to \infty).$$

Second, it may be used to prove that certain arithmetic progressions contain infinitely many primes. To see this, let $p \mid n_i$ and note that by (1.1), we have

$$2^{2^{i-1}} \equiv -1 \pmod{p}, \quad 2^{2^i} \equiv 1 \pmod{p}.$$

Hence the order of 2 $\pmod p$ is precisely 2^i. It follows that 2^i divides the order of the multiplicative group mod p, so that $2^i \mid (p-1)$ or, otherwise stated, $p \equiv 1 \pmod{2^i}$. As a consequence, for any fixed k, there are infinitely many primes $p \equiv 1 \pmod{2^k}$: choose a prime p_i dividing n_i for each $i \geq k$. The subject of primes in prescribed arithmetic progressions will be revisited in Chapter 2, when we prove Dirichlet's celebrated theorem on the subject.

A related method of proving the infinitude of the primes is as follows: Let $a_1 < a_2 < a_3 < \dots$ be a sequence of positive integers with the property that

$$\gcd(i, j) = 1 \quad \Longrightarrow \quad \gcd(a_i, a_j) = 1.$$

Moreover, suppose that for some prime p, the integer a_p has at least two distinct prime divisors. Then if $p_1, \dots, p_k$ were a list of all the primes, the integer

$$a_{p_1} a_{p_2} \cdots a_{p_k}$$

would possess least $k+1$ prime factors: indeed, each factor exceeds 1, the factors are pairwise relatively prime, and one of the factors is divisible by two distinct primes. So there are $k + 1 > k$ primes, a contradiction.

It remains to construct such a sequence. We leave to the reader the easy exercise of showing that $a_n = 2^n - 1$ has the desired properties (note that $a_{11} = 23 \cdot 89$). A similar proof taking instead a_n as the nth Fibonacci number was given by M. Wunderlich [Wun65] and was abstracted much as above by Hemminger [Hem66].

Exercise 1.2.3 *(A. Granville).* Let $n_1 = 2$, and for $i > 1$ define

$$n_i = 1 + \prod_{j<i} n_j.$$

Thus the sequence $\{n_i\}$ begins $2, 3, 7, 43, \dots$. Show that

a) The n_i are pairwise relatively prime.

b) For $i > 1$, $n_i = f(n_{i-1})$, where $f(T) = T^2 - T + 1$.

c) If $p \mid f(n)$ for some integer n, then -3 is a square $\pmod p$. Taking a prime divisor of each n_i, conclude that there are infinitely many primes $p \equiv 1 \pmod 3$.

Exercise 1.2.4 *(Mohanty [Moh78]).* Let m be a positive integer. Let A_1 be any integer with $\gcd(A_1, m) = 1$, and for $n \geq 1$ define

$$A_{n+1} = A_n^2 - mA_n + m.$$

a) Show that $\gcd(A_i, A_j) = 1$ for $i \neq j$. Suggestion: Show that if $p \mid A_i$ for some i, then for all $j > i$ we have $p \nmid A_j$.

b) Show that if $A_1 > m$, one has $A_n > m \geq 1$ for every n. Use this to give another demonstration that there are an infinite number of primes.

c) Taking $A_1 = 3$ and $m = 2$, show that $A_n = 2^{2^{n-1}} + 1$; thus we have recovered Goldbach's proof. Note that the choice $A_1 = 2$, $m = 1$ corresponds to Exercise 1.2.3.

Exercise 1.2.5 *(Harris [Har56]).* Let b_0, b_1, b_2 be positive integers with b_0 coprime to b_2. Define A_k for $k = 0, 1$ and 2 as the numerator when the finite continued fraction

$$b_0 + \cfrac{1}{b_1 + \cfrac{1}{\ddots + \cfrac{1}{b_k}}}$$

is put in lowest terms. For $k = 3, 4, \ldots$, inductively define b_k and A_k by

$$b_k = A_0 A_1 \ldots A_{k-3}$$

and A_k by the rule given above. Prove that the A_i form an increasing sequence of pairwise coprime positive integers.

Exercise 1.2.6 *(Aldaz & Bravo [AB03]).* Let p_i denote the ith prime. Euclid's argument shows that for each r, there is a prime in the interval $(p_r, \prod_1^r p_i + 1]$. Prove that the number of primes in the (smaller) interval $(p_r, \prod_2^r p_i + 1]$ tends to infinity with r.

Suggestion: With $P = \prod_2^r p_i$, show that $P - 2, P - 2^2, \ldots, P - 2^k$ are > 1 and pairwise coprime for fixed k and large r; then choose a prime factor of each.

1.2.3 The Riemann Zeta Function

Define the Riemann zeta function (for $\Re(s) > 1$, to ensure convergence) by

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

A mere glance at any analytic number theory text (such as [Dav00]) will reveal that this function occupies a central position in the subject, particularly in investigations concerning the distribution of primes. We shall more or less avoid it in these notes, but in the deeper investigations the zeta function is indispensable.

Riemann introduced the study of $\zeta(s)$ as a function of a complex variable in an 1859 memoir on the distribution of primes [Rie59]. But the connection between the zeta function and the prime numbers goes back earlier. Over a hundred years prior, Euler had looked at the same series for real s and had had shown that[Eul37, Theorema 8]

$$\sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_p \frac{1}{1 - \frac{1}{p^s}} \quad (s > 1). \tag{1.2}$$

This is often called an analytic statement of unique factorization. To see why, notice that formally (i.e, disregarding matters of convergence)

$$\prod_p \left(1 + \frac{1}{p^s} + \frac{1}{p^{2s}} + \cdots\right) = \sum_{n=1}^{\infty} \frac{a_n}{n^s},$$

where a_n counts the number of factorizations of n into prime powers. Thus, unique factorization (the statement that $a_n = 1$ for all n) is equivalent to the statement that (1.2) holds as a formal product of Dirichlet series, which is itself equivalent to the fact that (1.2) holds for all real $s > 1$ (or even a sequence of s tending to ∞) by a standard uniqueness theorem for Dirichlet series (for a proof see, e.g., [Apo76, Theorem 11.3]).

Euler's product expansion of ζ is the first example of what is now called an "Euler factorization." We now prove (following [Hua82]) a theorem giving general conditions for the validity of such factorizations.

Theorem 1.2.2 (Euler Factorization). *Let f be a multiplicative function. Then*

$$\sum_{n=1}^{\infty} f(n) = \prod_p \left(1 + f(p) + f(p^2) + \cdots\right) \tag{1.3}$$

provided either of the following two conditions holds:

i. $\sum_{n=1}^{\infty} |f(n)|$ *converges.*

ii. $\prod_p \left(1 + |f(p)| + |f(p^2)| + \cdots\right)$ *converges.*

If f is completely multiplicative, the factors in (1.3) form a geometric series whose convergence is implied by either of the above conditions. Thus we have the following consequence:

Corollary 1.2.3. *Let f be a completely multiplicative function. Then*

$$\sum_{n=1}^{\infty} f(n) = \prod_p \frac{1}{1 - f(p)}$$

subject to either of the two convergence criteria above.

Observe that (1.2) is immediate from this corollary. One takes $f(n) = 1/n^s$ and observes that for $s > 1$, condition i) holds (for example) by the integral test.

Proof of Theorem 1.2.2. Suppose i) holds above, and set $S_0 = \sum_{n=1}^{\infty} |f(n)|$. Observe that for each prime p, the series $\sum_{k=0}^{\infty} f(p^k)$ converges absolutely, since each term of the series $\sum_{k=0}^{\infty} |f(p^k)|$ appears in the sum defining S_0. Therefore

$$P(x) = \prod_{p \le x} \left(1 + f(p) + f(p^2) + \cdots\right)$$

is a finite product of absolutely convergent series; consequently,

$$P(x) = \sum_{n:p|n \Rightarrow p \le x} f(n).$$

If we now set $S = \sum_{n=1}^{\infty} f(n)$ (which converges absolutely), we have

$$S - P(x) = \sum_{n:p|n \text{ for some } p > x} f(n),$$

which shows

$$|S - P(x)| \le \sum_{n>x} |f(n)| \to 0$$

as $x \to \infty$. Thus $P(x) \to S$ as $x \to \infty$, which is the assertion of (1.3).

Now suppose that ii) holds above. We shall show that i) holds as well, so the result follows from what we have just done. To see this, let

$$P_0 = \prod_p \bigl(1 + |f(p)| + |f(p^2)| + \cdots\bigr),$$

and let

$$\begin{aligned} P_0(x) &:= \prod_{p \le x} \bigl(1 + |f(p)| + |f(p^2)| + \cdots\bigr) \\ &= \sum_{n:p|n \Rightarrow p \le x} |f(n)| \ge \sum_{n \le x} |f(n)|. \end{aligned}$$

Since $P_0(x) \le P_0$ for all positive x, the partial sums $\sum_{n \le x} |f(n)|$ form a bounded increasing sequence. Thus $\sum |f(n)|$ converges, proving i). □

We can now present Euler's proof of the infinitude of the primes.

Euler's first proof of Theorem 1.2.1. Let f be defined by $f(n) = 1/n$ for every n. Assuming there are only finitely many primes p, condition ii) of Theorem (1.3) is trivially satisfied, as the product in question only has finitely many terms. It follows that

$$\sum_{n=1}^{\infty} \frac{1}{n} = \prod_p \left(1 + \frac{1}{p} + \frac{1}{p^2} + \cdots\right) < \infty,$$

in contradiction with the divergence of the harmonic series. □

It was known to Euler that this proof actually gave more:

Theorem 1.2.4. *The series $\sum 1/p$ diverges, where the sum is over all primes p.*

Proof. Suppose not and let $C = \sum 1/p$. As in the last proof, take $f(n) = 1/n$ in Theorem 1.2.2. We check that condition ii) of that theorem holds in this circumstance as well. For nonnegative x,

$$\prod_{p\le x}\left(1+\frac{1}{p}+\frac{1}{p^2}+c\dots\right)=\prod_{p\le x}\frac{1}{1-\frac{1}{p}}=\prod_{p\le x}\left(1+\frac{1}{p-1}\right)\le\prod_{p\le x}\left(1+\frac{2}{p}\right).$$

Now recall that $e^t \ge 1+t$ for every nonnegative t; this is clear from truncating the Taylor expansion $e^t = 1+t+t^2/2!+\cdots$. It follows that

$$\prod_{p\le x}\left(1+\frac{2}{p}\right)\le\prod_{p\le x}e^{2/p}=e^{\sum_{p\le x}2/p}\le e^{2C}.$$

Consequently, the partial products

$$\prod_{p\le x}\left(1+\frac{1}{p}+\frac{1}{p^2}+\dots\right)$$

form a bounded, increasing sequence, which implies ii). Thus

$$\sum_{n=1}^{\infty}\frac{1}{n}=\prod_{p}\frac{1}{1-\frac{1}{p}}\le e^{2C},$$

a contradiction. □

Similar methods give an explicit lower bound on the partial sums $\sum_{p\le x}1/p$: Note that for $x\ge 2$,

$$\prod_{p\le x}\frac{1}{1-\frac{1}{p}}=\sum_{n\le x:p|n\Rightarrow p\le x}\frac{1}{n}\ge\sum_{n\le x}\frac{1}{n}\ge\log x. \tag{1.4}$$

Proceeding as above, we deduce that $\sum_{p\le x}(p-1)^{-1}\ge\log\log x$. To derive a lower bound for $\sum_{p\le x}1/p$ from this, note that

$$\begin{aligned}\sum_{p\le x}\frac{1}{p}&=\sum_{p\le x}\frac{1}{p-1}-\sum_{p\le x}\left(\frac{1}{p-1}-\frac{1}{p}\right)\\&\ge\sum_{p\le x}\frac{1}{p-1}-\sum_{n\ge 2}\left(\frac{1}{n-1}-\frac{1}{n}\right)=\left(\sum_{p\le x}\frac{1}{p-1}\right)-1\ge\log\log x-1.\end{aligned}\tag{1.5}$$

Exercise 1.2.7. Say that a number n is *squarefull* if $p^2\mid n$ whenever $p\mid n$, i.e., if every prime showing up in the factorization of n occurs with multiplicity larger than 1. Using Euler factorizations (Theorem 1.2.2), show that $\sum n^{-1}$ converges when the sum is restricted to squarefull numbers n. Determine all real α for which $\sum n^{-\alpha}$ converges.

Exercise 1.2.8 *(Ramanujan).* Taking as known

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6} \qquad \text{and} \qquad \sum_{n=1}^{\infty} \frac{1}{n^4} = \frac{\pi^4}{90},$$

show that $\sum_n 1/n^2 = 9/(2\pi^2)$, where the sum ranges over positive squarefree integers n with an *odd* number of prime divisors. Hint: $|\mu| - \mu = 2\chi$, where χ is the characteristic function of the integers in question.

The next few proofs also make use of the ζ-function and its Euler factorization, but in a decidedly different manner.

Proof of J. Hacks. We need the well-known result that $\zeta(2) = \pi^2/6$; several proofs of this may be found in [Cha02]. We put $s = 2$ in the Euler factorization of $\zeta(s)$ to obtain

$$\frac{\pi^2}{6} = \zeta(2) = \prod_p \frac{1}{1 - \frac{1}{p^2}}.$$

If there are only finitely many primes, the rightmost product is a finite product of rational numbers, implying that $\pi^2/6$, so also π^2, is rational. But π^2 is irrational since π is transcendental. $\square$

Exercise 1.2.9. In this exercise we present a similar proof based instead on the irrationality of π. Let

$$\chi(n) = \begin{cases} (-1)^{(n-1)/2} & \text{if } 2 \nmid n, \\ 0 & \text{otherwise.} \end{cases}$$

a) Show that $\chi(n)$ is a completely multiplicative function, i.e.,

$$\chi(ab) = \chi(a)\chi(b)$$

for every pair of positive integers a, b.

b) Assume there are only finitely many primes. Show that this assumption implies that for every $s > 0$,

$$\sum_{n=1}^{\infty} \frac{\chi(n)}{n^s} = \prod_p \left(1 - \frac{\chi(p)}{p^s}\right)^{-1}.$$

c) Take $s = 1$ and obtain a contradiction to the irrationality of π. You may assume that $\pi/4 = 1 - 1/3 + 1/5 - 1/7 + \cdots$.

Hacks' demonstration relies on knowing the irrationality of π^2; we can avoid direct considerations of irrationality provided we also assume as known that $\zeta(4) = \pi^4/90$:

Proof. We begin with the observation that

$$\frac{\zeta(2)^2}{\zeta(4)} = \frac{\pi^4/36}{\pi^4/90} = \frac{5}{2}.$$

We now use the Euler factorization of ζ to get $5/2$ as an infinite product of primes. Since

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

we find

$$\zeta(2)^2/\zeta(4) = \prod_p (1-p^{-4})(1-p^{-2})^{-2} = \prod_p \frac{p^4-1}{p^4}\frac{p^4}{(p^2-1)^2} = \prod_p \frac{p^2+1}{p^2-1},$$

so that

$$\frac{5}{2} = \frac{5}{3}\frac{10}{8}\frac{26}{24}\cdots.$$

If there are only finitely many primes, then the product on the right hand side is a finite one and can be written as M/N, where $M = 5\cdot 10\cdot 26\cdots$ and $N = 3\cdot 8\cdot 24\cdots$. Then $M/N = 5/2$, so $2M = 5N$; since $3 \mid N$, necessarily $3 \mid M$. But this cannot be: M is a product of numbers of the form k^2+1, and no such number is a multiple of 3. $\square$

The above was inspired by Wagstaff's (open) question [Guy94, B48] as to whether there exists an elementary proof that $\frac{5}{2} = \prod_p \frac{p^2+1}{p^2-1}$.

1.2.4 Squarefree Numbers

By unique factorization there is a bijection

$$\{\text{subsets of the primes}\} \longleftrightarrow \{\text{squarefree positive integers}\},$$

given by taking

$$S \longmapsto \prod_{p\in S} p.$$

So to prove the infinitude of the primes, it suffices to prove there are infinitely many positive squarefree integers.

J. Perott's proof, 1881. We sieve out the non-squarefree integers from $1,\ldots,N$ by removing those divisible by 1^2, then those divisible by 2^2, etc. The number of removed integers is bounded above by

$$\sum_{k=2}^{\infty} \lfloor N/k^2 \rfloor \le N\sum_{k=2}^{\infty} k^{-2} = N(\zeta(2)-1),$$

so that the number of squarefree integers up to N, say $A(N)$, satisfies

$$A(N) \ge N - N(\zeta(2)-1) = N(2-\zeta(2)).$$

Because $1/t^2$ is a decreasing function of t,

$$\zeta(2) = 1 + \sum_{n=2}^{\infty} \frac{1}{n^2} < 1 + \sum_{n=1}^{\infty} \int_n^{n+1} \frac{dt}{t^2} = 1 + \int_1^{\infty} \frac{dt}{t^2} = 2.$$

It follows that $A(N)/N$ is bounded below by a positive constant, so that the squarefree numbers have positive lower density. In particular, there are infinitely many of them. $\square$

Remark. We have modified Perrott's proof by eliminating the use of $\zeta(2) = \pi^2/6$.

As observed by Dressler[Dre75], this proof also yields a lower bound on $\pi(N)$. One has only to take logarithms in the estimate $2^{\pi(N)} \geq A(N) \gg N$ to find

$$\pi(N) \geq \log A(N)/\log 2 \geq \log N/\log 2 + O(1).$$

Every positive n can be decomposed as the product of a square and a squarefree integer, for dividing n by its largest square factor can only yield a squarefree quotient. This simple observation plays a key role in the following short proof of Theorem 1.2.1:

Erdős' proof. Let N be a positive integer. There are at most $\sqrt{N}$ squares not exceeding N and at most $2^{\pi(N)}$ squarefree integers below this bound. Since every integer not exceeding N is a product of a square not exceeding N and a squarefree number not exceeding N, we must have

$$2^{\pi(N)}\sqrt{N} \geq N.$$

Dividing by $\sqrt{N}$ and taking logarithms, we get the bound

$$\pi(N) \geq \log N/\log 4. \qquad \square$$

Actually we can get considerably more mileage from this idea, namely the divergence of the series $\sum 1/p$:

Erdős' proof of Theorem 1.2.4. Suppose $\sum 1/p$ converges; then the contribution from the tail of the sum tends to 0, so that there exists some real number M with

$$\sum_{p>M} \frac{1}{p} < 1/2. \tag{1.6}$$

Keep this M fixed.

Now let N be an arbitrary positive integer. The estimate (1.6) implies that most integers up to N factor completely over the primes not exceeding M. Indeed, the number of integers not exceeding N that have a prime factor $p > M$ is bounded above by

$$\sum_{M<p\leq N} \left\lfloor \frac{N}{p} \right\rfloor \leq N \sum_{p>M} \frac{1}{p} < N/2,$$

so that at least $N/2$ integers not exceeding N are divisible only by primes $p \leq M$.

We now show that there are too few integers divisible only by primes $p \leq M$ for this to be possible. There are at most $\sqrt{N}$ squares not exceeding N, and at most $C := 2^{\pi(M)}$ squarefree numbers composed only of primes not exceeding M. Thus there are at most $C\sqrt{N}$ integers not exceeding N all of whose prime factors do not exceed M. But $C\sqrt{N} < N/2$ whenever N is large, in fact as soon as $N > 4C^2$. □

It is remarkable that this method of proving the infinitude of the primes (in contrast with Euclid's, for instance) is independent of the additive structure of the integers. This allows us to carry over the idea to certain commutative semigroups:

Exercise 1.2.10. Let S be a countable set equipped with a binary operation satisfying the usual axioms for an abelian group with the (possible) exception of the existence of inverses. Moreover, suppose that there is a nonempty set of "primes" $\mathcal{P} \subset S$ with the property that every $s \in S$ admits a factorization

$$s = p_1^{e_1} p_2^{e_2} \cdots p_k^{e_k},$$

where each $p_i \in \mathcal{P}$ and the e_i are natural numbers. (Note that we do *not* require uniqueness of this representation.) Furthermore, suppose there is a "norm"

$$\| \cdot \| : S \to \mathbf{Z}^+$$

with the following properties:

i. $\| \cdot \|$ is totally multiplicative, i.e., $\|ab\| = \|a\|\|b\|$ for every $a, b \in S$.

ii. There exist constants $c_1, c_2 > 0$ such that

$$c_1 x \leq |\{s \in S : \|s\| \leq x\}| \leq c_2 x \tag{1.7}$$

whenever x is sufficiently large.

Show that $\sum_{p \in \mathcal{P}} \|p\|^{-1}$ diverges.

Exercise 1.2.11 *(continuation).* We can recover the divergence of $\sum 1/p$ by taking for S the semigroup of positive integers, $\mathcal{P}$ the set of primes, and $\| \cdot \|$ the usual absolute value. Here we present an analogous application to the Gaussian integers:

a) Show that the conditions i) and ii) of the preceding exercise are satisfied with S the set of nonzero Gaussian integers, $\mathcal{P}$ the set of Gaussian primes together with $\{1, -1, i, -i\}$, and $\| \cdot \|$ the usual norm map. Conclude that $\sum 1/\mathcal{N}\pi$ diverges, where the sum is over all Gaussian primes π.

   Hint: Verifying (1.7) amounts to estimating the number of lattice points in the circle of radius $\sqrt{x}$; for this, consider appropriate circumscribed and inscribed squares.

b) Show that the sum $\sum 1/\mathcal{N}\pi$, taken over the Gaussian primes π associated to rational primes $p \equiv 3 \pmod 4$, converges. By combining this with part a), show that $\sum_{p\equiv 1 \pmod 4} 1/p$ diverges.

Remark. Suppose that in Exercise 1.2.10 we assume that elements of S factor uniquely (up to order) as products of elements of $\mathcal{P}$. Furthermore, suppose we replace (1.7) with the tighter hypothesis that

$$|\{s \in S : \|s\| \leq x\}| = cx + O(x^\theta) \quad (x \to \infty) \tag{1.8}$$

for some positive constant c and some $\theta < 1$. Then Knopfmacher [Kno75] has shown

$$\pi_{\mathcal{P}}(x) := \sum_{p\in\mathcal{P}, \|p\|\leq x} 1 = (c + o(1))\frac{x}{\log x}; \tag{1.9}$$

this generalizes the prime number theorem.

For the reader who has seen some algebraic number theory, we remark that all these hypotheses are satisfied with $\mathfrak{O}_K$ the ring of integers of a number field, $\mathcal{P}$ the set of this ring's maximal ideals and $\|\cdot\|$ as the usual norm map on ideals. In this case (1.9) amounts to the "prime ideal theorem" (see the brief discussion at the start of §1.9). However, verifying condition (1.8), or even the weaker (1.7), is difficult. An instructive exercise is to check the latter for $\mathbf{Q}(\sqrt{2})$; note that this implies $\sum_{p\equiv \pm 1 \pmod 8} 1/p$ diverges. The material from the final section of Appendix A may prove useful.

1.2.5 Smooth Numbers

A theme in the last few proofs is that for fixed M, there are not enough primes $p \leq M$ to account for the primes forming a "multiplicative basis" for the integers. A particularly direct proof along the same lines can be constructed by counting *smooth numbers*, numbers all of whose prime factors are small.

Define

$$\psi(x, y) := \#\{n \leq x : p \mid n \Longrightarrow p \leq y\}. \tag{1.10}$$

In other words, $\psi(x, y)$ counts the number of positive integers not exceeding x all of whose prime factors do not exceed y; such integers are called *y-smooth.* We need only the following trivial upper bound:

Lemma 1.2.5. *For $x \geq 1, y \geq 2$, we have*

$$\psi(x, y) \leq \left(1 + \frac{\log x}{\log 2}\right)^{\pi(y)}.$$

Proof. Let $k = \pi(y)$. By unique factorization, calculating $\psi(x, y)$ is equivalent to counting the number of k-tuples of nonnegative integers $e_1, \ldots, e_k$ with

$$p_1^{e_1} p_2^{e_2} \cdots p_k^{e_k} \leq x.$$

This inequality requires $p_i^{e_i} \le x$, so that

$$e_i \le \log x/\log p_i \le \log x/\log 2.$$

This forces e_i to lie in a set of $1 + \lfloor \log x/\log 2 \rfloor$ integers. $\square$

Since every positive integer not exceeding N is a (possibly empty) product of primes not exceeding N,

$$N = \psi(N, \pi(N)) \le (1 + \log N/\log 2)^{\pi(N)}.$$

It follows that

$$\pi(N) \ge \frac{\log N}{\log(1 + \log N/\log 2)}.$$

Taking some care to estimate the denominator, we obtain the lower bound

$$\pi(N) \ge (1 + o(1)) \frac{\log N}{\log\log N},$$

which of course tends to infinity. Proofs of Theorem 1.2.1 based on similar ideas were given by Thue in 1897 and Auric in 1915.

Exercise 1.2.12 *(M. Hirschhorn [Hir02]).* Let $p_1, p_2, \ldots$ denote the sequence of *odd* primes.

a) Reasoning as before, prove that the number of odd positive integers not exceeding N which can be written in the form $p_1^{e_1} \cdots p_k^{e_k}$ does not exceed

$$\prod_{i=1}^{k} \left(\frac{\log N}{\log p_i} + 1 \right) = \prod_{i=1}^{k} \frac{\log N p_i}{\log p_i} < (\log p_k N)^k < \sqrt{2k!}\sqrt{p_k N}.$$

b) Supposing $p_1, \ldots, p_k$ exist, prove that p_{k+1} exists and satisfies $p_{k+1} \le 4(2k!)p_k + 1$. (Of course this is far weaker than Bertrand's postulate, which asserts $p_{k+1} < 2p_k$.)

Exercise 1.2.13 *(A Bit More Psixyology).* We now develop a sharper estimate of $\psi(x, y)$ when y is fixed and x is tending to infinity.

Suppose that $x \ge 1$ and $y \ge 2$, and let $p_1, \ldots, p_k$ be the list of all primes not exceeding y (so that $k = \pi(y)$).

a) Show that $\psi(x, y)$ is given by

$$\#\{(x_1, \cdots, x_k) \in \mathbf{N}^k : x_1 \log p_1 + x_2 \log p_2 + \cdots + x_k \log p_k \le \log x\}.$$

Hence, for fixed y, $\psi(x, y)$ counts the number of "first quadrant" lattice points contained in a simplex expanding homothetically with x.

b) Recall that for positive numbers $a_1, \cdots, a_k$,

$$\mathrm{vol}(\{(x_1, \cdots, x_k) : a_1 x_1 + a_2 x_2 + \cdots a_k x_k \le X, \text{each } x_i \ge 0\}) = X^k/(k! a_1 \cdots a_k).$$

Referring to Appendix A, Theorem A.3.1, prove that for fixed $y \ge 2$,

$$\psi(x, y) \sim \frac{1}{\pi(y)!} \frac{\log^{\pi(y)} x}{\log p_1 \log p_2 \cdots \log p_{\pi(y)}} \quad (x \to \infty).$$

For another proof of this formula, see [Rub93]. By considering appropriate inscribed and circumscribed tetrahedrons, one may show that the difference of the two sides in this asymptotic formula is $O(\log^{\pi(y)-1} x)$ (again for fixed y, as $x \to \infty$).

1.2.6 The Heavy Machinery

We now present three high-powered proofs of the infinitude of the primes, utilizing (respectively) nonstandard analysis, topology and commutative algebra.

None of these proofs really yield any new insights. But for the sensibly sophisticated party host, they are indispensable.

Euclid's proof revisited (Non-standard Analysis, [Gol98]). Let $\mathcal{P}$ be the set of primes, which we suppose finite. Let $*\mathbf{Z}^+$ denote the hyperreal extension of the positive integers, and $*\mathcal{P}$ the extension of $\mathcal{P}$. $\mathcal{P}$ finite implies $*\mathcal{P} = \mathcal{P}$. Choose $n \in *\mathbf{Z}^+$ divisible (in $*\mathbf{Z}^+$) by every standard positive integer. Let $p \in *\mathcal{P}$ be a divisor (in $*\mathbf{Z}^+$) of $n+1$. (That n, p exist follows from transfer.) As $*\mathcal{P} = \mathcal{P}$, p is a standard positive integer, so $p \mid n$. Hence $p \mid (n+1) - n = 1$, which is absurd. $\square$

Furstenberg's proof (Point-set Topology, [Fur55]). We put a topology on the integers by taking as basic open sets all arithmetic progressions (infinite in both directions). Then each arithmetic progression is both open and closed: it is open by choice of the basis, and it is closed since its complement is the union of the other arithmetic progressions with the same common difference. For each prime p, let $A_p = p\mathbf{Z}$, and define $A := \cup_p A_p$. The set $\{-1, 1\} = \mathbf{Z} \setminus A$ is not open. (Indeed, every open set is either empty or contains an arithmetic progression, so infinite.) It follows that A is not closed. On the other hand, if there are only finitely many primes then A is a finite union of closed sets, hence itself closed. $\square$

Remark. Golomb [Gol59] studies the topology on the positive integers generated by taking as a basis the (restricted) arithmetic progressions

$$(q\mathbf{Z} + a) \cap \mathbf{Z}^+ = \{n \ge 1 : n \equiv a \pmod{q}\},$$

where

$$a, q \in \mathbf{N}, \quad \gcd(a, q) = 1, \quad \text{and} \quad q > 0.$$

This topology possesses a number of interesting properties; e.g., Dirichlet's theorem on the infinitude of primes in arithmetic progressions is equivalent to the statement that the set of primes is dense. See also [KP97], which considers similar questions in a more general algebraic context.

Washington's proof (Commutative Algebra). We use the result that a Dedekind domain with finitely many nonzero prime ideals is a principal ideal domain (see, e.g., [Lor96, Proposition III.2.12]), hence also a unique factorization domain.

The ring of integers of any number field is always a Dedekind domain; consequently, if K is a number field for which $\mathfrak{O}_K$ does not possess unique factorization, then $\mathfrak{O}_K$ has infinitely many nonzero prime ideals. Each such prime ideal lies above a rational prime p, and for each prime p there are at most $[K:\mathbf{Q}]$ prime ideals lying above it. It follows that there are infinitely many primes p, provided there exists a single example of a number ring $\mathfrak{O}_K$ without unique factorization. And there does: we may famously take $K = \mathbf{Q}(\sqrt{-5})$, as the factorization

$$6 = (1+\sqrt{-5})(1-\sqrt{-5})$$

is a well-known instance of the failure of unique factorization there. $\square$

1.2.7 Exercises

Exercise 1.2.14 *(Schur).* For $f(T) \in \mathbf{Z}[T]$, let $\mathcal{P}(f)$ be the set of primes p which divide $f(n)$ for some integer n. Show that if f is nonconstant, $\mathcal{P}(f)$ is infinite.

The next two exercises outline an elementary proof that there are infinitely many primes $p \equiv 1 \pmod{m}$ for every positive integer m. We apply the result of Exercise 1.2.14 to the so-called *cyclotomic polynomials*. Recall that the mth cyclotomic polynomial is defined by

$$\Phi_m(T) = \prod_{\substack{1\le k\le m\\ \gcd(k,m)=1}} \left(T - e^{2\pi i k/m}\right),$$

i.e., $\Phi_m(T)$ is the monic polynomial in $\mathbf{C}[T]$ whose roots are precisely the primitive mth roots of unity, each occurring with multiplicity 1. To even to hope apply the preceding exercise, we need that the coefficients of these polynomials are not merely complex numbers, but in fact integers. We sketch the proof of this now:

For any m, we have

$$T^m - 1 = \prod_{d\mid m} \Phi_d(T); \tag{1.11}$$

to see this, note that $T^m - 1$ is the product of $T - \zeta$ where ζ ranges over all mth roots of unity. But the mth roots of unity are the disjoint union of the primitive dth roots of unity, taken over those d dividing m, which implies (1.11). Applying Möbius inversion to (1.11) yields

$$\Phi_m(T) = \prod_{d\mid m} \left(T^d - 1\right)^{\mu(m/d)} = F/G,$$

where
$$F = \prod_{d\mid m,\, \mu(m/d)=1} (T^d - 1), \quad G = \prod_{d\mid m,\, \mu(m/d)=-1} (T^d - 1).$$

Now F and G are *monic* polynomials in $\mathbf{Z}[T]$ with $G \neq 0$, so we can write

$$F = GQ + R, \tag{1.12}$$

where $Q, R \in \mathbf{Z}[T]$ and $\deg R < \deg Q$. Of course (1.12) remains valid over $\mathbf{C}[T]$ and expresses in that ring one result of division by Q. But we know that over $\mathbf{C}[T]$, we have $F = G\Phi_m$ with no remainder; by the uniqueness of the division algorithm for polynomials, it follows that $R = 0$ above. Consequently,

$$\Phi_m = F/G = Q \in \mathbf{Z}[T].$$

Exercise 1.2.15. Let m be a positive integer and suppose that $p \mid \Phi_m(n)$ for the integer n. Show that either $p \mid m$ or the order of $n \pmod p$ is exactly m. Proceed as follows:

a) Using (1.11), show that
$$p \mid \Phi_m(n) \mid n^m - 1.$$
Thus $d := \mathrm{ord}_p n \mid m$.

b) Show that $p \mid \Phi_d(n)$.

c) Suppose that $p \mid \Phi_m(n)$ and that d is a proper divisor of m; show that $T^m - 1$ has a multiple root over $\mathbf{F}_p$.

d) Show that if $T^m - 1$ has a multiple root over $\mathbf{F}_p$, then $p \mid m$. (Hint: What are the roots of f'?)

Exercise 1.2.16. Combining the result of the preceding exercise with that of Exercise 1.2.14, complete the proof that there are infinitely many primes $p \equiv 1 \pmod m$.

Exercise 1.2.17 *(A. Granville, cf. [Has50, IV., p. 171]).* Using the results of Exercises 1.2.2 and 1.2.16, show that at least three of the four residue classes $1, 5, 7, 11 \pmod{12}$ contain infinitely many primes.

Exercise 1.2.18. Let $0 \neq f \in \mathbf{Z}[T]$. Prove the following result of Bauer (as given in [Nag64, pp. 168-169]):

Theorem. If $0 \neq f(T) \in \mathbf{Z}[T]$ *is a polynomial with at least one real root, then for every* $m \geq 3$, *there exist infinitely many primes* $p \in \mathcal{P}(f)$ *with* $p \not\equiv 1 \pmod m$.

Fix $m \geq 3$ and proceed by showing that each of the following conditions on a polynomial $f(T) \in \mathbf{Z}[T]$ is sufficient for the conclusion of the theorem to hold:

i. f has a positive leading coefficient and constant term -1.

ii. f has a positive leading coefficient and negative constant term.

iii. f has a positive leading coefficient and $f(a) < 0$ for some $a \in \mathbf{Z}$.

iv. f has a positive leading coefficient and $f(a) < 0$ for some $a \in \mathbf{Q}$.

v. f has a positive leading coefficient and $f(a) < 0$ for some $a \in \mathbf{R}$.

vi. f has a positive leading coefficient and $f(a) = 0$ for some $a \in \mathbf{R}$. Hint: Show that this is implied by v) provided f has no multiple roots; then argue that we can always make this assumption. Gauss' lemma may be of help.

Exercise 1.2.19 *(Goldbach).* Prove that there is no nonconstant polynomial $f(T) \in \mathbf{Z}[T]$ with the property that $f(n)$ is prime for all natural numbers n.

1.3 Discovering the Prime Number Theorem

> ...Even before I had begun my more detailed investigations into higher arithmetic, one of my projects was to turn my attention to the decreasing frequency of primes, to which end I counted the primes in several chiliads [intervals of length 1000] and recorded the results on the attached white pages. I soon recognized that behind all of its fluctuations, this frequency is on average inversely proportional to the logarithm, so that the number of primes below a given bound n is approximately equal to
>
> $$\int \frac{dn}{\log n},$$
>
> where the logarithm is understood to be hyperbolic. – Gauss, Christmas Eve letter to Enke, 1849 (excerpted from [Gol73])

We know from the preceding section that $\pi(x) \to \infty$ as $x \to \infty$. But how fast and how regularly does it do so? We shall examine this question the way Gauss did, empirically, and in this way we shall be led to conjecture the result now known as the *prime number theorem*.

1.3.1 An Empirical Approach

We begin by studying the "density" of primes around a given point x; it is not clear how this should be defined, other than that it should be something obtained by counting a set of primes near x and dividing by the "size" of the set. We quantify this with the choice

$$\Delta(x) = \frac{\pi(x) - \pi(x - 1000)}{1000},$$

so that $\Delta(x)$ measures the proportion of primes in the interval $(x-1000, x]$. If we make a table of values of x against values of $\Delta(x)$, we end up with a table resembling Table 1.1 (a similar table may be found in Gauss' collected works [Gau73a]): The first two rows of Table 1.1 suggest that $\Delta(x)$ is a slowly decreasing function of x.

It was Gauss' genius to invert $\Delta(x)$ and look for approximations by elementary functions. In this way he discovered empirically that $\Delta(x) \approx 1/\log x$. Since $\Delta(x)$ is defined as the slope of a chord on the graph of $y = \pi(x)$, it is natural to think that one could recover $\pi(x)$ by integrating this approximation. This suggests

$$\pi(x) \approx \mathrm{li}(x) := \int_2^x \frac{dt}{\log t}.$$

Table 1.2 shows a comparison of $\pi(x)$ and $\mathrm{li}(x)$ for powers of 10 up to 10^{12}. Gauss had access to a similar table (with values in a more limited range), and he no doubt noticed that while the absolute difference between $\mathrm{li}(x)$ and $\pi(x)$ appears large for large x, the relative difference is small in that the ratio $\pi(x)/\mathrm{li}(x)$ seems to be tending to 1. Reading the approximation sign above as asymptotic equality, our conjecture takes the following shape:

Prime Number Theorem. *As $x \to \infty$, we have the asymptotic formula*

$$\pi(x) \sim \int_2^x \frac{dt}{\log t}. \tag{1.13}$$

Gauss conjectured this sometime in the (closed) interval between 1792 and 1793, when he was still a teenager. However, the first *published* record of a conjecture of this form occurs later in 1798, in Legendre's *Essai sur la Théorie des Nombres*.

The most important step towards a proof of the prime number theorem was taken by Riemann [Rie59], who in 1859 devoted his only memoir on number theory to outlining a possible attack on the prime number theorem through a study of the analytic properties of the function $\zeta(s)$. It was to take 40 more years for complex analysis to develop to the point where Riemann's plan could be carried through. This was independently accomplished in in 1896 by Hadamard and de la Vallée Poussin. There are still no simple proofs, although there are short proofs which require only a modicum of familiarity with complex analysis (e.g. [Zag97]). In Chapter 4, we will give a (long) proof of the prime number theorem free of any complex analysis.

Table 1.1: Comparison of $\Delta(x)$ and $1/\log x$, rounded to nearest thousandth

x	1000	2000	3000	4000	5000	6000	7000	8000	9000
$\Delta(x)$	.168	.135	.127	.120	.119	.114	.117	.107	.110
$1/\log x$	.145	.132	.125	.121	.117	.115	.113	.111	.110

Table 1.2: Comparison of $\pi(x)$ and $\mathrm{li}(x)$

x	$\pi(x)$	$\mathrm{li}(x)$	$\mathrm{li}(x) - \pi(x)$	$\pi(x)/\mathrm{li}(x)$
10^3	168	177	9	.9514937
10^4	1,229	1245	16	.9870757
10^5	9,592	9,629	37	.9961820
10^6	78,498	78,627	129	.9983658
10^7	664,579	664,917	338	.9994909
10^8	5,761,455	5,762,208	753	.9998694
10^9	50,847,534	50,849,234	1700	.9999665
10^{10}	455,052,512	455,055,614	3103	.9999933
10^{11}	4,118,054,813	4,118,066,400	11587	.9999972
10^{12}	37,607,912,018	37,607,950,280	38263	.9999988

The prime number theorem is often stated in the form

$$\pi(x) \sim \frac{x}{\log x}. \tag{1.14}$$

It is not difficult to show that the two statements (1.13) and (1.14) are equivalent: if we integrate (1.13) by parts, we see

$$\begin{aligned} \mathrm{li}(x) = \int_2^x \frac{dt}{\log t} &= \left.\frac{t}{\log t}\right|_2^x + \int_2^x \frac{dt}{\log^2 t} \\ &= \frac{x}{\log x} - \frac{2}{\log 2} + \int_2^x \frac{dt}{\log^2 t}. \end{aligned} \tag{1.15}$$

If we can show that

$$\mathrm{li}(x) \sim x/\log x, \tag{1.16}$$

then the equivalence of (1.13) and (1.14) follows by transitivity. The constant term on the right of (1.15) is unimportant, so we focus on showing the integral appearing here is of smaller order of magnitude than $x/\log x$.

First, a plausibility argument: most values of $1/\log^2 t$ in $[2, x]$ should look roughly like $1/\log^2 x$; we thus expect a contribution of roughly $x/\log^2 x$ from the integral. To make this rigorous, we think of x as large and split the range of integration at $\sqrt{x}$. The first integral, between 2 and $\sqrt{x}$, is estimated trivially from above, noting that the denominator is bounded below by the positive constant $1/\log^2 2$. For the latter integral, taken between $\sqrt{x}$ and x, we note that the denominator is at least $\log^2 \sqrt{x} \gg \log^2 x$. These observations lead to the estimate

$$\begin{aligned} \int_2^x \frac{dt}{\log^2 t} &= \int_2^{\sqrt{x}} \frac{dt}{\log^2 t} + \int_{\sqrt{x}}^x \frac{dt}{\log^2 t} \\ &\ll \sqrt{x} + \frac{x}{\log^2 x} \ll \frac{x}{\log^2 x} = o(x/\log x). \end{aligned}$$

Now (1.16) follows easily.

Exercise 1.3.1. Use L'Hôpital's rule to give another proof that $\mathrm{li}(x) \sim x/\log x$.

1.3.2 Exercises: Some Consequences of the PNT

Exercise 1.3.2. Assuming the prime number theorem, prove that the nth prime p_n satisfies $p_n \sim n \log n$ as $n \to \infty$.

Suggestion: Show that if $a_n \sim b_n$ and $a_n \to \infty$ (both as $n \to \infty$), then $a_n \log a_n \sim b_n \log b_n$. Apply this with $a_n = p_n/\log p_n$ and $b_n = n$.

Exercise 1.3.3. We give two applications of the preceding exercise:

a) Prove that $p_{n+1}/p_n \to 1$.

   Consequently, to each $\epsilon > 0$ there corresponds an x_0 with
   $$\pi((1+\epsilon)x) - \pi(x) > 0$$
   whenever $x > x_0$. We will prove this elementarily for $\epsilon = 1$ later in the chapter (Theorem 1.5.3).

b) Prove that the set $\{p/q : p, q \text{ prime}\}$ is dense in the set of nonnegative real numbers.

   Hint: what is the limit of p_{an}/p_{bn} as $n \to \infty$?

The prime number theorem in its most basic form (1.13) is equivalent to the relation
$$\pi(x) = \mathrm{li}(x) + o(x/\log x). \tag{1.17}$$
Now $\mathrm{li}(x)$ can be expanded through repeated integration by parts:
$$\mathrm{li}(x) = \frac{x}{\log x} + 1!\frac{x}{\log^2 x} + 2!\frac{x}{\log^3 x} + \cdots + (k-1)!\frac{x}{\log^k x} + O_k\left(\frac{x}{\log^{k+1} x}\right). \tag{1.18}$$

We would like to substitute the expansion (1.18) into (1.17). However, as things stand now it is impossible for this process to produce a sharper estimate than $\pi(x) = x/\log x + o(x/\log x)$, because the higher order terms of (1.18) are absorbed into the error $o(x/\log x)$. In order to preserve the usefulness of this expansion we need a better error term in the prime number theorem, such as that provided by the following theorem (see, e.g., [Dav00]):

Prime Number Theorem (with error term). *For a certain absolute constant $c > 0$, we have*
$$\pi(x) = \mathrm{li}(x) + O(x \exp(-c(\log x)^{1/2}) \tag{1.19}$$
for all $x \geq 2$.

The quantity inside the O-term in (1.19) is easily checked to be $O_A(x/\log^A x)$ for every A, but is not $O(x^\delta)$ for any $\delta < 1$. Nevertheless, its is suspected that $\pi(x) - \mathrm{li}(x) = O(x^\delta)$ for every $\delta > 1/2$; this is equivalent to the most famous conjecture in all of number theory, the *Riemann hypothesis*.

In any event, the fact that the error term is $O(x/\log^A x)$ for arbitrary A implies

$$\pi(x) = \frac{x}{\log x} + 1!\frac{x}{\log^2 x} + \cdots + (k-1)!\frac{x}{\log^k x} + O_k\left(\frac{x}{\log^{k+1} x}\right), \qquad (1.20)$$

which is already exceedingly useful.

Exercise 1.3.4. In 1798, Legendre conjectured $\pi(x)$ was of the form $x/(A\log x + B)$ for some constants A and B. In 1808, he refined this conjecture, claiming that

$$\pi(x) = \frac{x}{\log x - A(x)}$$

where $A(x)$ is "approximately $1.08366\cdots$." Presumably he meant that if we solve the above equation for $A(x)$, so that

$$A(x) = -(x/\pi(x) - \log x),$$

then $A(x) \to 1.08366\cdots$ as $x \to \infty$. Assuming (1.20), show that that Legendre was wrong: $A(x) \to 1$ as $x \to \infty$.

Actually Legendre was proven wrong in regard to this matter fifty years before the prime number theorem was established: in 1848, Chebyshev showed that if $A(x)$ tends to a limit, the limit has to be 1. An alternate means of establishing Chebyshev's claim, due to Pintz, is indicated in Exercise 1.7.2.

Exercise 1.3.5 *(Landau).* Use (1.20) to show that for large enough x, there are eventually more primes in the interval $(1, x]$ than in the interval $(x, 2x]$. Show that in fact

$$\pi(2x) - 2\pi(x) \to -\infty \quad (x \to \infty).$$

1.4 The Simplest Estimates for $\pi(x)$

In this chapter, we content ourselves with estimates that can be established relatively easily by elementary means. Here we take up two such estimates, both of which can be considered corollaries to the divergence of $\sum 1/p$, and both of which (in some nebulous form) find their origins in remarks of Euler.

1.4.1 The Primes are Infinitely Fewer than the Integers

We begin with a simple but nevertheless decidedly nontrivial result:

Theorem 1.4.1. *As $x \to \infty$, we have $\pi(x)/x \to 0$.*

Remark. This theorem is suggested by Euler's claim [Eul37, Theorema 7, Corollarium 3] that the "primes are infinitely fewer than the integers." However, Euler does not prove Theorem 1.4.1; his justification for this assertion is that

$$\prod_p \frac{p}{p-1} = \log\infty, \quad \text{while} \quad \prod_{n=2}^{\infty} \frac{n}{n-1} = \infty.$$

Indeed, the former product "is" (formally) the harmonic series, which Euler interpreted as $\log\infty$ (motivated by the growth of its partial sums). By contrast, the partial products $\prod_{2\le n\le x} n/(n-1)$ telescope to $\lfloor x\rfloor$, which perhaps motivates Euler's labelling the latter product "absolute" ∞.

The proof of Theorem 1.4.1 proceeds along rather intuitive lines. We already know how to get upper bounds on the proportion (i.e., upper density) of the primes: half of all numbers are even, so already the proportion of primes cannot exceed 1/2. Another way of viewing this argument is that all but finitely many primes fall into the unique reduced residue class $\pmod 2$; from this perspective, the correct generalization is in plain sight. Namely, take any integer q; then every prime $p \nmid q$ has to fall into one of the $\phi(q)/q$ reduced residue classes $\pmod q$, and this forces the proportion of primes to be at most $\phi(q)/q$. If we can make $\phi(q)/q$ arbitrarily small, this will imply our theorem.

We now formalize these ideas, first proving that it is possible to find such q.

Lemma 1.4.2. *Let $\epsilon > 0$. Then there exists q with $\phi(q)/q < \epsilon$.*

Proof. Note that $\phi(q)/q = \prod_{p|q}(1-1/p)$, which is small when q is a product of many small primes. Guided by this, we let $q := q_x = \prod_{p\le x} p$ be the product of all primes not exceeding x. Then from (1.4) we know that for $x \ge 2$, one has

$$\phi(q_x)/q_x = \prod\nolimits_{p\le x}(1-1/p) \le (\log x)^{-1},$$

which yields the desired inequality as soon as $x > \max\{2, e^{1/\epsilon}\}$. $\square$

Proof of Theorem 1.4.1. Let q be a positive integer (and p a prime). If p and q have a nontrivial common factor, necessarily $p \mid q$. Hence

$$\pi(x) \le |\{1 \le n \le x : \gcd(n,q) = 1\}| + \nu(q),$$

where $\nu(q)$ denotes the number of distinct prime divisors of q. The interval $[1,x]$ is contained in the first $\lceil x/q\rceil$ blocks of q consecutive integers (beginning at 1), each of which contains $\phi(q)$ numbers relatively prime to q. It follows that

$$\frac{\pi(x)}{x} \le \frac{\phi(q)(x/q+1)}{x} + \frac{\nu(q)}{x}.$$

Now let $\epsilon > 0$ be given. Choose q with $\phi(q)/q < \epsilon/2$; so that

$$\frac{\pi(x)}{x} \le \frac{\phi(q)}{q} + \frac{\phi(q)}{x} + \frac{\nu(q)}{x} < \epsilon/2 + \frac{\phi(q)}{x} + \frac{\nu(q)}{x} < \epsilon$$

for all large x. Hence $\pi(x)/x \to 0$, as claimed. $\square$

1.4.2 More Primes than Squares

We have just shown that there are not too many primes. On the other hand, we have some reason to think that there are not too few primes. For instance, the divergence of $\sum 1/p$ versus the convergence of $\sum 1/n^2$ suggests there are "more primes than squares." Alternately, following Euler [Eul37, Theorema 7, Corollarium 2], we could compare products instead of sums (for the first product, cf. (1.4); note the second product below telescopes so is easy to compute):

$$\prod_p \frac{p}{p-1} = \log\infty \quad \text{while} \quad \prod_{n=2}^{\infty} \frac{n^2}{n^2-1} = 2.$$

To make precise the claim that there are more primes than squares (which Euler did not), apply summation by parts (Theorem A.2.3):

$$\begin{aligned} \sum_{p\le x} \frac{1}{p} &= \int_{3/2}^{x} \frac{d\pi(t)}{t} \qquad (1.21)\\ &= \frac{\pi(x)}{x} - \frac{\pi(3/2)}{3/2} + \int_{3/2}^{x} \frac{\pi(t)}{t^2}\,dt = \int_2^x \frac{\pi(t)}{t^2}\,dt + O(1). \end{aligned}$$

If $\pi(x) \ll x^{1/2}$, the final integral here is $\ll \int_2^x t^{-3/2}\,dt \ll 1$. This in turn implies $\sum 1/p$ has bounded partial sums, so (as a series of positive terms) is convergent, an absurdity. This gives us one way of making our claim precise: for any positive constant c and any x_0, there is an $x > x_0$ with $\pi(x) > c\sqrt{x}$. (However, notice that we have *not* proved $\pi(x) \gg \sqrt{x}$.)

We can get more impressive results by replacing $x^{1/2}$ with certain faster-growing functions. Note for instance that

$$\int_2^{\infty} \frac{dt}{t\log^{1+\epsilon} t} = \int_{\log 2}^{\infty} \frac{1}{e^u u^{1+\epsilon}} e^u\,du = \int_{\log 2}^{\infty} \frac{du}{u^{1+\epsilon}} < \infty$$

for every positive ϵ. It follows by the same argument as above that for no $\epsilon > 0$ is $\pi(x) \ll x/\log^{1+\epsilon} x$. We have thus proven the following: For every $\epsilon > 0$,

$$\limsup_{x\to\infty} \frac{\pi(x)}{x/\log^{1+\epsilon} x} = \infty.$$

Exercise 1.4.1. Show that one can replace $x/\log^{1+\epsilon} x$ with

$$\frac{x}{\log x(\log\log x)^{1+\epsilon}}.$$

Generalize.

Exercise 1.4.2. Show that if we assume as known not only that $\sum 1/p$ diverges, but that (cf. (1.5))

$$\sum_{p\le x} \frac{1}{p} \ge \log\log x - 1,$$

then the above method yields

$$\limsup_{x\to\infty} \frac{\pi(x)}{x/\log x} \geq 1.$$

It is important to emphasize that we have not proved any lower estimate for $\pi(x)$; we have merely disproved an upper estimate. In fact, from the mere divergence of $\sum 1/p$ it is *impossible* to obtain any lower bound of the form $\pi(x) \gg f(x)$ for a function $f(x) \to \infty$:

Exercise 1.4.3. Let f be a nonnegative-valued function defined for $x > x_0$ with the property that $f(x) \to \infty$ as $x \to \infty$. Show that there exists a set $\mathcal{A}$ of positive integers with the following two properties:

$$\sum_{a\in A} \frac{1}{a} = \infty, \qquad \liminf_{x\to\infty} \frac{A(x)}{f(x)} = 0.$$

Hint: Use the divergence of $\sum_{n=N}^{\infty} n^{-1}$ for every N to show we can satisfy the first criterion while including enough "gaps" to force the second to hold as well.

1.5 Chebyshev's Work on $\pi(x)$

> [Chebyshev] was the only man ever able to cope with the refractory character and erratic flow of prime numbers and to confine the stream of their progression with algebraic limits, building up, if I may so say, banks on either side which that stream, devious and irregular as are its windings, can never overflow. – J.J. Sylvester

The first significant results on $\pi(x)$ since Euclid were published by Chebyshev in two important 1851-1852 papers ([Che51], [Che52]). We shall focus our attention on three of his results:

Theorem 1.5.1. *Suppose that*

$$\lim_{x\to\infty} \frac{\pi(x)}{x/\log x}$$

tends to a limit as $x \to \infty$. Then that limit equals 1.

Theorem 1.5.2. *There exist positive constants c_1, c_2 and a real number x_0 such that*

$$c_1 \frac{x}{\log x} \leq \pi(x) \leq c_2 \frac{x}{\log x}$$

for $x > x_0$.

Theorem 1.5.3 (Bertrand's postulate, asymptotic form). *For all sufficiently large x, there exists a prime in the interval $(x, 2x]$.*

This section contains simple proofs of each of these assertions; the methods of proof are related to Chebyshev's, but are not identical. For more faithful renderings of Chebyshev's work, see [Nar00, Chapter 3], [Sha83, Chapter 9] or [Dia82].

We begin by introducing certain auxiliary functions studied by Chebyshev, namely:

$$\theta(x) := \sum_{p \le x} \log p, \qquad \psi(x) := \sum_{n=1}^{\infty} \theta(x^{1/n}). \tag{1.22}$$

The sum defining ψ appears to be infinite, but is essentially finite since $\theta(x^{1/n})$ vanishes whenever $x^{1/n} < 2$.

From the analytic point of view, these functions turn out to be better-behaved and more natural objects of study than $\pi(x)$. But estimates for $\pi(x)$ can be easily deduced from estimates for θ or ψ. For example, partial summation shows that

$$\theta(x) = \int_{3/2}^{x} \log t \, d\pi(t) = \pi(x) \log x - \int_{2}^{x} \frac{\pi(t)}{t} \, dt.$$

Because $\pi(t)/t = o(1)$ (Theorem 1.4.1), we have $\int_2^x \pi(t)/t \, dt = o(x)$, whence

$$\theta(x) = \pi(x) \log x + o(x),$$

and

$$\frac{\theta(x)}{x} = \frac{\pi(x)}{x/\log x} + o(1). \tag{1.23}$$

The same estimate holds with ψ in place of θ, the reason being that ψ and θ differ by only a small amount. To quantify this, write

$$\psi(x) - \theta(x) = \theta(x^{1/2}) + \theta(x^{1/3}) + \dots.$$

Now $\theta(x^{1/n})$ vanishes whenever $x^{1/n} < 2$, i.e., as soon as $n > \log x/\log 2$. Consequently, only $O(\log x)$ of the terms in the right-hand sum are nonzero. Because of the trivial bound $\theta(x) \le x \log x$, we see that

$$\psi(x) - \theta(x) \ll x^{1/2} \log x + (x^{1/3} \log x) \log x \ll x^{1/2} \log x. \tag{1.24}$$

Thus replacing θ with ψ in equation (1.23) results in an extra error term $O(\log x/x^{1/2})$, which can be absorbed into the existing $o(1)$ error term. We have thus shown:

Proposition 1.5.4. *As $x \to \infty$, we have both*

$$\frac{\theta(x)}{x} = \frac{\pi(x)}{x/\log x} + o(1), \tag{1.25}$$

$$\frac{\psi(x)}{x} = \frac{\pi(x)}{x/\log x} + o(1). \tag{1.26}$$

Consequently:

Corollary 1.5.5. *If any of $\theta(x)/x, \psi(x)/x, \pi(x)/(x/\log x)$ tends to a limit as $x \to \infty$, then all of them do, and the limit in each case is the same.*

In particular, the prime number theorem is equivalent to either of the assertions $\theta(x) \sim x$, $\psi(x) \sim x$.

Indeed, (1.25) and (1.26) imply together that

$$\liminf_{x\to\infty} \frac{\theta(x)}{x} = \liminf_{x\to\infty} \frac{\pi(x)}{x/\log x} = \liminf_{x\to\infty} \frac{\psi(x)}{x},$$

and similarly for the lim sup.

The definition of ψ we gave above is useful for comparison with θ, but it masks the arithmetic information that ψ encodes. If we observe that for any fixed positive integer k,

$$\theta(x^{1/k}) = \sum_{p \le x^{1/k}} \log p = \sum_{p^k \le x} \log p,$$

then we may rewrite

$$\psi(x) = \theta(x) + \theta(x^{1/2}) + \cdots = \sum_{p^k \le x} \log p, \tag{1.27}$$

where the final sum is over *all* pairs (p, k) where p is prime, k is a positive integer and $p^k \le x$. Now introduce the von-Mangoldt function $\Lambda(n)$, defined by

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

This is well-defined by unique factorization, and equation (1.27) says

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

Lemma 1.5.6. *For every positive integer n,*

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

Proof. Write $n = \prod_{p|n} p^{e_p}$. Then

$$\begin{aligned} \sum_{d|n} \Lambda(d) = \sum_{p^k|n} \log p &= \sum_{p|n} \sum_{k=1}^{e_p} \log p \\ &= \sum_{p|n} e_p \log p = \sum_{p|n} \log p^{e_p} = \log \prod_{p|n} p^{e_p} = \log n. \end{aligned}$$ □

We may finally introduce our main tool for estimating $\psi(x)$. Following Chebyshev, let

$$\begin{aligned} T(x) :&= \sum_{n \le x} \log n \\ &= x \log x - x + O(\log x), \end{aligned}$$

where the final line is the weak Stirling approximation to $\log [x]!$ (which can easily be obtained from Appendix A, Corollary A.2.2). The following lemma allows us to obtain estimates for ψ from our estimate of T:

Lemma 1.5.7. *For every positive x, we have*

$$T(x) = \sum_{n \le x} \psi(x/n).$$

Proof. Observe

$$\begin{aligned} \sum_{n \le x} \psi(x/n) = \sum_{n \le x} \sum_{m \le x/n} \Lambda(k) &= \sum_{nm \le x} \Lambda(m) \\ &= \sum_{N \le x} \sum_{m \mid N} \Lambda(m) = \sum_{N \le x} \log N = T(x). \end{aligned}$$ □

1.5.1 Proof of Theorem 1.5.1

If we combine Lemma 1.5.7 with the estimate for $T(x)$, we see

$$\sum_{n \le x} \psi(x/n) \sim x \log x \quad (x \to \infty).$$

This is the same estimate one would obtain if one substituted x/n for $\psi(x/n)$, which can be considered heuristic support for the prime number theorem in the form $\psi(x) \sim x$. Moreover, this idea can be used to prove the following proposition:

Proposition 1.5.8. *Suppose that $\psi(x)/x \to C$ as $x \to \infty$. Then $C = 1$.*

Proof. It suffices to show the hypothesis implies $\sum_{n \le x} \psi(x/n) \sim Cx \log x$. By hypothesis, we may write $\psi(x) = Cx + g(x)$ where $g(x) = o(x)$. Then

$$\begin{aligned} \sum_{n \le x} \psi(x/n) &= Cx \sum_{n \le x} n^{-1} + \sum_{n \le x} g(x/n) \\ &= Cx \log x + o(x \log x) + \sum_{n \le x} g(x/n). \end{aligned} \tag{1.28}$$

We would like to show that the final summand can be absorbed into the error term $o(x \log x)$. To see this, let $\epsilon > 0$ be given and choose N large enough that

$x > N$ implies $|g(x)|/x < \epsilon/2$. Let M be an upper bound for $|g|$ on $[1, N]$. Then

$$\left|\sum_{n\le x} g(x/n)\right| \le \sum_{\substack{n\le x\\ x/n<N}} |g(x/n)| + \sum_{\substack{n\le x\\ x/n\ge N}} |g(x/n)|$$
$$\le Mx + \frac{\epsilon}{2}x\sum_{n\le x} n^{-1} < \epsilon x\log x$$

for sufficiently large x. It follows that $\sum_{n\le x} g(x/n)$ is $o(x\log x)$, which inserted in (1.28) completes the proof of the claim. □

By Corollary 1.5.5, we see that if $\pi(x)/(x/\log x)$ tends to a limit, then that limit is also necessarily 1. That is, we have proved Theorem 1.5.1.

Exercise 1.5.1. Modify the proof of Proposition 1.5.8 to show

$$\liminf_{x\to\infty} \frac{\pi(x)}{x/\log x} \le 1, \quad \limsup_{x\to\infty} \frac{\pi(x)}{x/\log x} \ge 1.$$

1.5.2 Proof of Theorem 1.5.2

One theorem down, two to go. We begin by noticing that for $x \ge 4$, say, we have the estimate

$$T(x) - 2T(x/2) = x\log x - x + O(\log x) - 2\left(\frac{x}{2}\log\frac{x}{2} - \frac{x}{2} + O\left(\log\frac{x}{2}\right)\right)$$
$$= x\log 2 + O(\log x).$$

On the other hand, by Lemma 1.5.7 we can also write

$$T(x) - 2T(x/2) = \sum_{n\le x}\psi(x/n) - \sum_{n\le x} 2\psi(x/2n)$$
$$= \psi(x) - \psi(x/2) + \psi(x/3) - \psi(x/4) + \dots$$

Since ψ is a nondecreasing function, this is an alternating series of terms nonincreasing in absolute value. It follows that for any even k, we have

$$T(x) - 2T(x/2) \ge \psi(x) - \psi(x/2) + \cdots + \psi(x/(k-1)) - \psi(x/k), \qquad (1.29)$$

while for any odd k, we have

$$T(x) - 2T(x/2) \le \psi(x) - \psi(x/2) + \dots - \psi(x/(k-1)) + \psi(x/k). \qquad (1.30)$$

Take $k = 1$; this gives the lower bound

$$\psi(x) \ge T(x) - 2T(x/2) = x\log 2 + O(\log x).$$

Getting an upper bound on $\psi(x)$ is a tad bit trickier. We first take $k = 2$ in (1.29), which gives us

$$\psi(x) - \psi(x/2) \le T(x) - 2T(x/2) = x\log 2 + O(\log x).$$

Replacing x with $x/2, x/4$, etc., we obtain estimates for $\psi(x/2) - \psi(x/4)$, then $\psi(x/4) - \psi(x/8)$, etc. We then add the estimates. For this, think of x as large, and choose k with $8 \leq x/2^k < 16$. For each $1 \leq j \leq k$,

$$\psi(x/2^{j-1}) - \psi(x/2^j) \leq \frac{x}{2^j}\log 2 + O\left(\log \frac{x}{2^{j-1}}\right) = \frac{x}{2^j}\log 2 + O(\log x).$$

Noting that $k \ll \log x$ and summing the expressions for $j = 1, 2, \ldots, k$ gives the upper estimate

$$\begin{aligned}\psi(x) - \psi(x/2^k) &\leq x \log 2 \left(1 + \frac{1}{2} + \cdots + \frac{1}{2^j}\right) + O(\log x \log x), \\ &\leq 2x \log 2 + O(\log^2 x).\end{aligned}$$

Thus

$$\psi(x) \leq 2x\log 2 + O(\log^2 x) + \psi(16) \leq 2x \log 2 + O(\log^2 x).$$

We assumed that x was sufficiently large when deriving this estimate, but it is clear now that it holds (with perhaps a different implied constant) for $x \geq 4$, say. Thus we have shown:

Proposition 1.5.9. *For $x \geq 4$, we have*

$$x \log 2 + O(\log x) \leq \psi(x) \leq 2x\log 2 + O(\log^2 x). \tag{1.31}$$

This implies Theorem 1.5.2: Recall (1.26), which states that as $x \to \infty$,

$$\frac{\pi(x)}{x/\log x} = \frac{\psi(x)}{x} + o(1).$$

Substituting the estimates of (1.31) yields

$$\log 2 + o(1) \leq \frac{\pi(x)}{x/\log x} \leq 2\log 2 + o(1) \tag{1.32}$$

as $x \to \infty$. It follows that in the statement of Theorem 1.5.2 we may take for c_1 and c_2 any constants less than $\log 2$ and greater than $2 \log 2$ respectively. Note that if the ratio $2\log 2/\log 2$ were any smaller, this would yield a proof of Bertrand's postulate!

Exercise 1.5.2. Show that there are positive constants c_1, c_2 such that for every $x \geq 2$, we have

$$c_1 \frac{x}{\log x} \leq \pi(x) \leq c_2 \frac{x}{\log x}.$$

1.5.3 Proof of Bertrand's Postulate

We begin with the observation that there is a prime in the interval $(x, 2x]$ precisely when $\theta(2x) - \theta(x) > 0$. Our strategy is to obtain a lower estimate for the related quantity $\psi(2x) - \psi(x)$ and then to transition to $\theta(2x) - \theta(x)$ using the estimate for the difference $\psi - \theta$ given by (1.24).

Our first instinct is perhaps to take $k = 2$ in (1.29); this immediately gives us a bound on $T(x) - 2T(x/2)$, namely

$$\psi(x) - \psi(x/2) \leq T(x) - 2T(x/2).$$

Unfortunately, the inequality is going the wrong way for our purposes. Instead, take $k = 3$ in (1.30); we then have

$$\psi(x) - \psi(x/2) + \psi(x/3) \geq T(x) - 2T(x/2) = x \log 2 + O(\log x).$$

But for large x, one has

$$\psi(x/3) \leq x\frac{2\log 2}{3} + O\left(\log^2 \frac{x}{3}\right) = x\frac{2\log 2}{3} + O(\log^2 x).$$

Consequently,

$$\psi(x) - \psi(x/2) \geq x\frac{\log 2}{3} + O(\log^2 x).$$

But since $\psi(u) - \theta(u) \ll u^{1/2} \log u$ for $u \geq 2$, this implies

$$\theta(x) - \theta(x/2) \geq x\frac{\log 2}{3} + O(x^{1/2} \log x) \qquad (x \to \infty). \tag{1.33}$$

As the right hand side is positive for large enough x, Theorem 1.5.3 follows.

In fact, we can get a lower bound on $\pi(x) - \pi(x/2)$ of the same order of magnitude as the lower bound for $\pi(x)$ that we derived in the previous section. Note that

$$\theta(x) - \theta(x/2) = \sum_{x/2 < p \leq x} \log p \leq \log x \left(\pi(x) - \pi(x/2)\right),$$

so that by (1.33) one has

$$\pi(x) - \pi(x/2) \geq \frac{\log 2}{3}\frac{x}{\log x} + O(x^{1/2}) = \left(\frac{\log 2}{3} + o(1)\right)\frac{x}{\log x} \quad (x \to \infty). \tag{1.34}$$

The argument of this section is due to Ramanujan [Ram19]. For other proofs of Bertrand's postulate, see §9.3 of H.N. Shapiro's superb introductory text [Sha83].

1.5.4 Exercises

Exercise 1.5.3 *(M. Nair [Nai82]).*

a) Show that
$$e^{\psi(x)} := \mathrm{lcm}[1, 2, \ldots, \lfloor x \rfloor].$$

b) As a corollary, prove that if $f(T) \in \mathbf{Z}[T]$ is a polynomial of degree d, then
$$e^{\psi(d+1)} \int_0^1 f(x)\, dx \in \mathbf{Z}.$$

c) Take $f(T) = T^n(1-T)^n$. Prove that
$$0 < \int_0^1 f(x)\,dx \le 4^{-n}.$$

d) Deduce from b), c) and the minimality of 1 among the positive integers that $\psi(2n+1) \ge 2n\log 2$. Conclude from this that as $x \to \infty$, we have $\psi(x) \ge x(\log 2 + o(1))$, so that $\pi(x) \ge (\log 2 + o(1))x/\log x$.

Exercise 1.5.4. Using the method of the previous exercise but considering instead $f(T) = (T-T^2)^{2n}(1-2T)^{2n}$, show that $\pi(x) \ge (C+o(1))x/\log x$, where
$$C := -\frac{\log m}{3} = .7803552047\ldots,$$
$$m := \frac{1}{3}\left(\frac{1}{2} - \frac{1}{6}\sqrt{3} - \left(\frac{1}{2} - \frac{1}{6}\sqrt{3}\right)^2\right)\sqrt{3}.$$

Combining this lower bound for $\pi(x)$ with the upper bound of (1.32), show that $\pi(2x) > \pi(x)$ whenever x is sufficiently large. We thus have another proof of Bertrand's postulate for large x.

Exercise 1.5.5. Let $P(n)$ denote the largest prime power divisor of n so that, e.g., $P(4) = 4$ and $P(100) = 25$.

a) Prove that $P(n) \to \infty$ as $n \to \infty$ without using Chebyshev's theorems.

b) From the inequality
$$\log n \le \sum_{p^k \le P(n)} \log p = \psi(P(n)),$$
deduce that $P(n) \gg \log n$ as $n \to \infty$.

c) Let $N_k := \prod_{i=1}^k p_i$ be the product of the first k primes. Show that $P(N_k) \ll \log N_k$ for $k = 1, 2, \ldots$.

Hint: For c), the bound $\theta(x) \gg x$ for $x \ge 2$ (which follows from (1.31) and (1.24)) may prove useful.

Exercise 1.5.6 *(Richert [Ric49]).* Assume Bertrand's postulate is known in its full form: for *every* $x \ge 1$, there is a prime in the interval $(x, 2x]$. Show that every integer $n > 6$ can be written as a sum of distinct primes.

Suggestion: The assertion holds for all integers $n \le 13$ using only primes not exceeding 11. Thus it is true for all $6 < n \le 26$ using only primes not exceeding 13, etc… A more general sufficient condition for all large integers to be the sum of distinct terms of a given sequence, together with some interesting applications, can be found in [Bro76].

Exercise 1.5.7. Let m and M be defined by

$$m := \liminf \frac{\pi(x)}{x/\log x}, \quad M := \limsup \frac{\pi(x)}{x/\log x};$$

so that according to Chebyshev's Theorem 1.5.2 we have $0 < m \le M < \infty$. Define d_n by $d_n = p_{n+1} - p_n$ (where p_i denotes the ith prime), so that the sequence $\{d_n\}$ begins $1, 2, 2, 4, 2, 4, 2, \ldots$. Prove that

$$\liminf d_n/\log p_n \le \frac{1}{m}, \quad \limsup d_n/\log p_n \ge \frac{1}{M}.$$

Exercise 1.5.8 *(Erdős & Turán [ET48], continuation).* Show that both $d_n < d_{n+1}$ and $d_n > d_{n+1}$ hold for infinitely many values of n. To establish the second of these, proceed as follows: Choose a positive constant C so that $d_n < C \log p_n$ holds infinitely often (as is possible by the last exercise). Supposing contrariwise that $d_n \le d_{n+1}$ holds for all $n \ge n_0$, consider the infinite chain $d_{n_0} \le d_{n_0+1} \le d_{n_0+2} \le \ldots$.

a) Show that for each positive integer k, one has $d_n = d_{n+1} = k$ for at most k contiguous pairs above.

b) Now choose $m > n_0$ with $d_m < C \log p_m$, and use a) to show

$$m - n_0 \le 1 + 2 + \cdots + d_m \le (C \log p_m)^2.$$

Hence obtain an upper bound for $m = \pi(p_m)$ contradicting (for large m) the lower bound of Theorem 1.5.2.

It is not known whether $d_n = d_{n+1}$ holds infinitely often.

Exercise 1.5.9. Chebyshev's 1851 paper contained a proof of the following result, versions of which may be found in [Nar00, §3.1], [Lan53, Chapter 10].

Theorem. For every $k = 2, 3, \ldots$ and every positive constant C, there exist infinitely many positive integers m with

$$\pi(m) < \mathrm{li}(m) + C \frac{m}{\log^k m},$$

and infinitely many positive integers n with

$$\pi(n) > \mathrm{li}(n) - C \frac{n}{\log^k n}.$$

In this exercise we give two of Chebyshev's applications:

a) Deduce that if $\pi(x)/\mathrm{li}(x)$ tends to a limit as $x \to \infty$, then that limit is necessarily 1. Since $\mathrm{li}(x) \sim x/\log x$, this implies Theorem 1.5.1.

b) Deduce that if $x/\pi(x) - \log x$ tends to a limit as $x \to \infty$, then that limit is necessarily -1. This disproves Legendre's conjecture mentioned in Exercise 1.3.4.

1.6 Polynomials that Represent Many Primes

Euler observed that $n^2 - n + 41$ is prime for each of $n = 0, 1, 2, \ldots, 40$. We know from Exercise 1.2.19 that no nonconstant polynomial can represent only primes, so that the pattern must eventually break down. And indeed, the value of Euler's polynomial at $n = 41$ is clearly divisible by 41.

There are quite a few questions we might have when confronted with this example. One is whether there is any "explanation" for this long run of primes. There is – surprisingly, one can see this phenomenon as a manifestation of uniqueness of factorization in $\mathbf{Z}[\frac{1+\sqrt{-163}}{2}]$; as this lies rather far afield, we refer the interested reader to [Coh80, Chapter 9, §8]. Another natural question is whether Euler's polynomial represents infinitely many primes. It almost certainly does, but we cannot prove it. Indeed, it is not known whether there is a single polynomial in $\mathbf{Z}[T]$ of degree > 1 which represents infinitely many prime values. Conjectural answers to such questions are considered in §1.8.3.

Humbled, we could ask whether there we can at least establish that there are polynomials of degree > 1 that represent, if not infinitely many primes, at least arbitrarily many. Our next theorem gives an affirmative answer, even when we restrict to a special class of quadratic polynomials:

Theorem 1.6.1 (Sierpiński [Sie64]). *For every N, there exists an integer k for which there are more than N primes represented by $T^2 + k$; i.e.,*

$$\sup_{k \in \mathbf{Z}} |\{p : p = a^2 + k \text{ for some } a \in \mathbf{Z}\}| = \infty.$$

Theorem 1.6.1 is an immediate consequence of the following sharper result, which we derive from Chebyshev's estimates by a simple counting argument:

Lemma 1.6.2. *For every $x \geq 2$, there exists a positive integer $k \leq 2\sqrt{x}$ such that the number of primes $p \leq x$ represented by $T^2 + k$ is $\gg \sqrt{x}/\log x$.*

Proof. For each $p \leq x$, write $p = \lfloor \sqrt{p} \rfloor^2 + k(p)$. Then $k(p) \geq 1$ (since no prime is a square), while

$$k(p) = p - \lfloor \sqrt{p} \rfloor^2 \leq p - (\sqrt{p} - 1)^2 \leq 2\sqrt{p} \leq 2\sqrt{x}.$$

It follows from the Pigeonhole principle that at least $\pi(x)/(2\sqrt{x})$ primes not exceeding x share the same value of $k(p)$. Since $\pi(x) \gg x/\log x$ for $x \geq 2$, the result follows. □

Note that we expect $T^2 + K$ to assume infinitely many prime values for any fixed integer K not of the form $-m^2$. This would follow from Schinzel & Sierpiński's "Hypothesis H," described in §1.8.3.

Our final question is whether it is possible to outdo Euler. That is, can we come up with a polynomial (with integer coefficients) assuming, say, 100 (distinct) prime values at consecutive integers? 1000? arbitrarily many? An affirmative answer is provided by the next theorem, a weakened version of a result proved by Chang & Lih in [CL77]. For the full result, which requires Dirichlet's theorem on primes in progressions, see Exercise 2.1.2:

Theorem 1.6.3. *For every natural number N, there is a polynomial $f(T) \in \mathbf{Z}[T]$ such that $f(n)$ produces $N+1$ distinct primes as n increases from 0 to N.*

Proof. For each $0 \le k \le N$, define

$$g_k(T) = \prod_{\substack{i=0 \\ i \neq k}}^{N} (T - i) \in \mathbf{Z}[T].$$

Then $g_k(k) \neq 0$, while $g_k(n) = 0$ for $1 \le n \neq k \le N$. From Exercise 1.2.16 (applied with $m = g_k(k)$) we may deduce that there are integers $m_0, m_1, \dots, m_N$ for which

$$1 + m_0 g_0(0), 1 + m_1 g_1(1), \dots, 1 + m_N g_N(N)$$

is a list of $N+1$ distinct primes. (Inductively choose the m_i so that each corresponding term is prime and different from its predecessors.) Then if we set

$$f(T) := 1 + \sum_{k=0}^{N} m_k g_k(T) \in \mathbf{Z}[T],$$

f will have the property enunciated in the theorem. □

How small an f (in terms of its degree) can we take in Theorem 1.6.3? Our construction shows such an f exists of degree N. Later (Exercise 1.8.9) we shall see that there is a suitable polynomial of degree 1, provided the *prime k-tuples conjecture* holds. See also Exercise 1.8.10.

Exercise 1.6.1. Give a proof of Theorem 1.6.1 using, instead of Chebyshev's theorem, the result of §1.4 that $\limsup_{x\to\infty} \pi(x)/\sqrt{x} = \infty$.

Exercise 1.6.2 *(Garrison [Gar90]).*

a) Show that for every $d \ge 2$ and every N, there exists a positive integer k for which $T^d + k$ assumes more than N prime values.

b) Show that part a) remains true if "positive" is replaced by "negative."

Exercise 1.6.3 *(Abel & Siebert [AS93]).* Let $f(T) \in \mathbf{Z}[T]$ be a polynomial of degree $d \ge 2$ with positive leading coefficient. Show that for every N, there exists an integer k for which $f(T) + k$ assumes more than N prime values.

Exercise 1.6.4. Show that for every N, there exists an even integer k such that there least N prime pairs $p, p+k$. More generally, for any positive integers r and N, there exist even integers $k_1 < k_2 < \cdots < k_{r-1}$ such that there are more than N prime r-tuples $p, p+k_1, \dots, p+k_{r-1}$.

Exercise 1.6.5. It is conjectured that there is always a prime between any two consecutive squares, moreover that the number between consecutive squares tends to infinity. Using Chebyshev's theorems, show that as $x \to \infty$,

$$\max_{n \le x} \pi((n+1)^2) - \pi(n^2) \gg x/\log x.$$

As one consequence, there is a pair of consecutive squares between which there are more than 1000 primes.

Remark. Sieve methods show that between any two large consecutive squares there are always numbers with "few" prime factors; see Chapter 3, Exercise 3.5.4(b) for a precise statement of this kind.

1.7 Some Estimates of Mertens

By 1737, Euler was not only aware that $\sum 1/p$ diverged, but had assigned the infinite sum the value $\log\log\infty$ [Eul37, Theorema 19], showing he possessed an inkling as to the rate of growth of the partial sums.

Sixty years later, Gauss [Gau73b, pp. 11-16] would make the more precise assertion that

$$\text{“}1+\frac{1}{2}+\frac{1}{3}+\frac{1}{5}+\frac{1}{7}+\frac{1}{11}+\cdots+\frac{1}{x}=(\text{for } x \text{ infinite})\ llx+V,''$$

writing that he suspected V to be a constant near 1.266. It seems reasonable to read this as the conjecture that $\sum_{p\le x} 1/p = \log\log x + V + o(1)$. Gauss also claimed

$$\text{“}\frac{2}{1}\cdot\frac{3}{2}\cdot\frac{5}{4}\cdots\frac{x}{x-1}=(x \text{ inf})\ a.lx''$$

for some constant a about 1.874, which we can read as the conjecture that

$$\prod\nolimits_{p\le x}(1-1/p)^{-1}\sim a\log x.$$

Mertens observed [Mer74] that Chebyshev's results could be used to obtain such asymptotic formulas for the sum $\sum_{p\le x} 1/p$ and the product $\prod_{p\le x}(1-1/p)$. His theorems prove Gauss' claims to be qualitatively right on, although Gauss' constants may be shown to be slightly off.

1.7.1 Mertens' Theorem, sans the Constant

We begin by estimating the quantity

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

From this, estimates for $\sum_{p\le x} 1/p$ follow by partial summation, and then estimates for $\prod_{p\le x}(1-1/p)$ follow by taking logarithms.

Our starting point is the observation that $T(x)$ can be re-expressed by

$$\begin{aligned}T(x) = \sum_{n\le x}\log n &= \sum_{n\le x}\sum_{d\mid n}\Lambda(d)\\ &= \sum_{d\le x}\sum_{\substack{n\le x\\ d\mid n}}\Lambda(d) = \sum_{d\le x}\Lambda(d)\left\lfloor\frac{x}{d}\right\rfloor.\end{aligned}$$

Now drop the greatest integer sign in the final sum; the incurred error is $\ll \sum_{d \leq x} \Lambda(d) = \psi(x) \ll x$ (for $x \geq 4$) by (1.31). Recalling that $T(x) = x \log x + O(x)$, we deduce in turn that

$$x \sum_{d \leq x} \frac{\Lambda(d)}{d} = x \log x + O(x),$$

$$\sum_{d \leq x} \frac{\Lambda(d)}{d} = \log x + O(1). \tag{1.35}$$

Estimate (1.35) was proven for $x \geq 4$, but it clearly remains true for every $x \geq 1$. If we recall the definition of $\Lambda(d)$, we can rewrite this sum in a more revealing form:

$$\sum_{d \leq x} \frac{\Lambda(d)}{d} = \sum_{p^k \leq x} \frac{\log p}{p^k}. \tag{1.36}$$

This is visibly closer to the first sum we are aiming to estimate, the difference being the inclusion of terms corresponding to prime powers p^k with $k \geq 2$. But these make a bounded contribution;

$$\begin{aligned} \sum_{\substack{p^k \leq x \\ k \geq 2}} \frac{\log p}{p^k} &\leq \sum_{p \leq x} \log p \sum_{k=2}^{\infty} p^{-k} \\ &= \sum_{p \leq x} \frac{\log p}{p(p-1)} \leq \sum_{2 \leq n \leq x} \frac{\log n}{n(n-1)} = O(1), \end{aligned} \tag{1.37}$$

since the final sum converges as $x \to \infty$. Combining (1.35), (1.36) and (1.37), we obtain the important result that for $x \geq 1$,

$$\sum_{p \leq x} \frac{\log p}{p} = \log x + O(1). \tag{1.38}$$

Now for $x \geq 2$,

$$\begin{aligned} \sum_{p \leq x} \frac{1}{p} = \int_{3/2}^{x} \frac{dA(t)}{\log t} &= \frac{A(x)}{\log x} - \frac{A(3/2)}{\log(3/2)} + \int_{3/2}^{x} \frac{A(t)}{t \log^2 t}\, dt \\ &= 1 + O\left(\frac{1}{\log x}\right) + \int_{2}^{x} \frac{dt}{t \log t} + \int_{2}^{x} \frac{A(t) - \log t}{t \log^2 t}\, dt \\ &= \log\log x + 1 - \log\log 2 + \int_{2}^{x} \frac{A(t) - \log t}{t \log^2 t}\, dt. \end{aligned}$$

Since $A(t) - \log t = O(1)$ and $\int_2^\infty dt/(t \log^2 t)$ converges, the integral

$$I := \int_{2}^{\infty} \frac{A(t) - \log t}{t \log^2 t}\, dt$$

converges (absolutely) to a finite real number I. Then

$$I - \int_2^x \frac{A(t) - \log t}{t \log^2 t}\, dt \ll \int_x^\infty \frac{dt}{t \log^2 t} = \frac{1}{\log x}.$$

We have thus shown that with

$$B := 1 - \log\log 2 + I, \tag{1.39}$$

we have

$$\sum_{p \le x} \frac{1}{p} = \log\log x + B + O\left(\frac{1}{\log x}\right). \tag{1.40}$$

As a corollary, we now prove a weakened form of the result usually known as Mertens' theorem:

Mertens' Theorem (minus the constant). *For some constant C,*

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

Proof. Let $P_x := \prod_{p \le x} (1 - 1/p)$, so that

$$\begin{aligned} \log P_x &= \sum_{p \le x} \log\left(1 - \frac{1}{p}\right) \\ &= -\sum_{p \le x} \frac{1}{p} - \sum_{p \le x} \sum_{k=2}^\infty \frac{1}{kp^k}. \end{aligned}$$

Since

$$\sum_{k=2}^\infty \frac{1}{kp^k} \le \frac{1}{2} \sum_{k=2}^\infty p^{-k} = \frac{1}{2p(p-1)} \ll p^{-2},$$

the infinite sum $\sum_p \sum_{k=2}^\infty k^{-1} p^{-k}$ converges absolutely to S, say. Then we also have

$$S - \sum_{p \le x} \sum_{k=2}^\infty \frac{1}{kp^k} \ll \sum_{p > x} p^{-2} \ll \sum_{n > x} n^{-2} \ll x^{-1},$$

so that

$$\begin{aligned} \log P_x &= -\log\log x - B + O(1/\log x) - S + O(1/x) \\ &= -\log\log x - B - S + O(1/\log x). \end{aligned}$$

Thus

$$\prod_{p \le x} \left(1 - \frac{1}{p}\right) = \frac{e^{-B-S}}{\log x} e^{O(1/\log x)} = \frac{e^{-B-S}}{\log x}\left(1 + O\left(\frac{1}{\log x}\right)\right)$$

and the result follows with $C = e^{-B-S}$. ∎

Mertens was able to derive the remarkable equality $B + S = \gamma$, where γ is the unique constant with

$$\sum\nolimits_{n\le x} \frac{1}{n} = \log x + \gamma + O(1/x) \quad (x \to \infty). \tag{1.41}$$

The existence of γ is proved in Appendix A, §A.2.2 as an example of Euler's summation formula.

1.7.2 The Constant in Mertens' Theorem

In the previous subsection we proved that

$$\prod\nolimits_{p\le x} (1 - 1/p) \sim C/\log x \quad (x \to \infty),$$

where, keeping our earlier notation, $C = e^{-(S+B)}$. Following Murty ([Mur01, Chapter 9]), we complete the proof of Mertens' theorem by proving that $S+B = \gamma$, the usual Euler-Mascheroni constant.

Let us introduce (or recall, in the case of the zeta function) the notation

$$\zeta(s) := \sum_{n=1}^{\infty} \frac{1}{n^s}, \quad Z(s) = \sum_{p} \frac{1}{p^s} \qquad (s > 1).$$

Both series are uniformly convergent on every half-line $\{s : s \ge 1 + \epsilon\}$; consequently, both $\zeta(s)$ and $Z(s)$ represent continuous functions for $s > 1$. Since both the harmonic series $\sum 1/n$ and the "prime harmonic series" $\sum 1/p$ diverge, we expect $\zeta(s)$ and $Z(s)$ to "blow up" as s tends down to 1. For $\zeta(s)$, this can be seen from the following lemma:

Lemma 1.7.1. *As s tends down to* 1*,*

$$\zeta(s) = \frac{1}{s-1} + O(1).$$

Proof. Fix $s > 1$. Since t^{-s} is a decreasing function of t on $(0, \infty)$,

$$\frac{1}{s-1} = \int_1^{\infty} t^{-s} \le \zeta(s) \le 1 + \int_1^{\infty} t^{-s}\, dt = 1 + \frac{1}{s-1},$$

so that $0 \le \zeta(s) - (s-1)^{-1} \le 1$. □

We can use this lemma to estimate $\log \zeta(s)$ as s tends down to 1:

Lemma 1.7.2. *As s tends down to* 1*,*

$$\log \zeta(s) = (s-1) \int_0^{\infty} H(u) e^{-(s-1)u}\, du + O(s-1), \tag{1.42}$$

where $H(x) = \sum_{n\le x} 1/n$.

Proof. By the preceding lemma, $(s-1)\zeta(s) = 1 + O(s-1)$, so that as s tends down to 1,

$$\log \zeta(s) = \log \frac{1}{s-1} + \log(1 + O(s-1)) = \log \frac{1}{s-1} + O(s-1). \tag{1.43}$$

Since (for $s \downarrow 1$)

$$1 - e^{-(s-1)} = 1 - (1 - (s-1) + O((s-1)^2)) = (s-1)(1 + O(s-1)),$$

we also have

$$\log\left(1 - e^{-(s-1)}\right) = \log(s-1) + O(s-1), \tag{1.44}$$

so that (comparing (1.43) and (1.44))

$$\log \zeta(s) = -\log\left(1 - e^{-(s-1)}\right) + O(s-1) = \sum_{n=1}^{\infty} \frac{e^{-(s-1)n}}{n} + O(s-1). \tag{1.45}$$

Applying partial summation we obtain

$$\begin{aligned}
\log \zeta(s) &= \int_0^{\infty} e^{-(s-1)u}\, dH(u) + O(s-1) \quad (s \downarrow 1) \\
&= H(x)e^{-(s-1)x} + (s-1)\int_0^x H(u)e^{-(s-1)u}\, du \Big]_{x=\infty} + O(s-1) \\
&= (s-1)\int_0^{\infty} H(u)e^{-(s-1)u}\, du + O(s-1).
\end{aligned}$$
□

We now derive an analogous estimate for $P(s)$:

Lemma 1.7.3. *As s tends down to 1, we have*

$$Z(s) = (s-1)\int_0^{\infty} P(e^u)e^{-(s-1)u}\, du, \tag{1.46}$$

where $P(x) = \sum_{p \le x} 1/p$.

Proof. By partial summation,

$$\begin{aligned}
\sum_p \frac{1}{p^s} &= \int_1^{\infty} \frac{dP(t)}{t^{s-1}} \\
&= \frac{P(x)}{x^{s-1}} - \frac{P(1)}{1} + (s-1)\int_1^x \frac{P(t)}{t^s}\, dt \Big]_{x=\infty} = (s-1)\int_1^{\infty} \frac{P(t)}{t^s}\, dt.
\end{aligned}$$

The lemma follows upon making the substitution $t = e^u$. □

Theorem 1.7.4 (Mertens' Theorem). *In the notation of the previous subsection,*

$$S + B = \gamma.$$

Therefore

$$\prod_{p \le x}(1 - 1/p) = \frac{e^{-\gamma}}{\log x}\left(1 + O\left(1/\log x\right)\right).$$

Proof. Define
$$f(s) := -\sum_p \left(\log\left(1 - \frac{1}{p^s}\right) + \frac{1}{p^s}\right).$$
Because of the Euler-factorization of $\zeta(s)$, we have
$$f(s) = \log\zeta(s) - \sum_p \frac{1}{p^s} = \log\zeta(s) - Z(s) \tag{1.47}$$
for $s > 1$. Now for $s \geq 1/2$,
$$\log\left(1 - \frac{1}{p^s}\right) + \frac{1}{p^s} \ll \frac{1}{p^{2s}},$$
so that the series defining $f(s)$ converges absolutely and uniformly on every half-line $\{s : s \geq 1/2 + \epsilon\}$. In particular, f is continuous at $s = 1$, and
$$f(1) = \sum_p \left(-\log\left(1 - \frac{1}{p}\right) + \frac{1}{p}\right) = \sum_p \sum_{k=2}^{\infty} \frac{1}{kp^k} = S.$$
It therefore suffices to establish that as s tends down to 1,
$$f(s) = \gamma - B + o(1).$$

From (1.47), (1.46) and (1.42),
$$f(s) = (s-1)\int_0^\infty (H(t) - P(e^t))e^{-(s-1)t}\,dt + O(s-1) \quad (s \downarrow 1). \tag{1.48}$$
We now employ our existing estimates for $H(x)$ and $P(x)$ given by (1.41) and (1.40) respectively. We find that for large t,
$$\begin{aligned} H(t) - P(e^t) &= (\log t + \gamma + O(1/t)) - (\log t + B + O(1/t)) \\ &= \gamma - B + O(1/(t+1)); \end{aligned}$$
this estimate persists for $t \geq 0$, with perhaps a different implied constant. Substituting into (1.48) shows
$$\begin{aligned} f(s) &= (s-1)\int_0^\infty \left((\gamma - \beta)e^{-(s-1)t} + O\left(\frac{e^{-(s-1)t}}{t+1}\right)\right) dt + O(s-1) \\ &= \gamma - \beta + O\left((s-1)\int_0^\infty \frac{e^{-(s-1)t}}{t+1}\,dt\right) + O(s-1). \end{aligned}$$

We are almost at our goal. The estimate $f(s) = \gamma - \beta + o(1)$ as $s \downarrow 1$ will follow as soon as we show
$$(s-1)\int_0^\infty \frac{e^{-(s-1)t}}{t+1}\,dt = o(1) \quad (s \downarrow 1). \tag{1.49}$$

We do this directly: Let $\epsilon > 0$ be given, and choose a positive integer N with $N^{-1} < \epsilon/2$. Rewrite the left hand side of (1.49) as

$$(s-1)\int_0^N \frac{e^{-(s-1)t}}{t+1}\,dt + (s-1)\int_N^\infty \frac{e^{-(s-1)t}}{t+1}\,dt.$$

Choose $\kappa > 1$ so that for any $s \in (1, \kappa]$, the first summand is smaller than $\epsilon/2$. The latter summand is bounded above by

$$(s-1)\int_0^\infty \frac{e^{-(s-1)t}}{N+1}\,dt = \frac{1}{N+1} < \epsilon/2,$$

so that for any $s \in (1, \kappa]$,

$$0 \le (s-1)\int_0^\infty \frac{e^{-(s-1)t}}{t+1}\,dt < \epsilon/2 + \epsilon/2 = \epsilon.$$

As $\epsilon > 0$ was arbitrary, the estimate (1.49) follows. □

1.7.3 Exercises

Exercise 1.7.1. Fix $\alpha \in (1/2, 1]$ and let $\mathcal{A} = \mathcal{A}_\alpha$ be the set of positive integers n possessing a prime factor $p > n^\alpha$. Here we prove $\mathcal{A}$ has density $\log 1/\alpha$. For example, taking $\alpha = 1/2$, this says that there is about a 70% chance a positive number has a prime divisor exceeding its square root.

a) Because $\alpha > 1/2$, every n has at most one prime divisor exceeding n^α. Use this to prove
$$A(x) = \sum_{p \le x} \sum_{\substack{n \le \max\{x, p^{1/\alpha}\} \\ p \mid n}} 1.$$

b) Split the sum in a) into $S_1 + S_2$, where S_1 is the sum over those primes $p \le x$ for which $x = \max\{x, p^{1/\alpha}\}$, and S_2 is the remainder. Prove that
$$S_1 = \sum_{p \le x^\alpha} \lfloor p^{1/\alpha - 1} \rfloor \ll \pi(x^\alpha) x^{1-\alpha} \ll x/\log x, \text{ and}$$
$$S_2 = \sum_{x^\alpha < p \le x} \lfloor x/p \rfloor = x \log \frac{1}{\alpha} + O(x/\log x).$$
Deduce that $A(x) := x\log(1/\alpha) + O(x/\log x)$.

c) Give a similar argument proving $\psi(x, x^\alpha) \sim x \log \frac{1}{\alpha}$. (For the definition of ψ, see (1.10).)

Exercise 1.7.2. For this exercise, assume the hypothetical estimate

$$\pi(x) = \frac{Ax}{\log x + B + o(1)}. \tag{1.50}$$

For example, if $x/\pi(x) - \log x$ tends to a limit C, then this holds with $A = 1$ and $B = C$. Following Pintz [Pin80], we will show that the only possibility of such an estimate is if $A = 1$ and $B = -1$; this disproves Legendre's conjecture.

a) Using partial summation, deduce from (1.50) that

$$\begin{aligned}\theta(x) &= \pi(x)\log x - \int_2^x \pi(t)\,dt/t \\ &= \frac{Ax}{1 + B/\log x + o(1/\log x)} - \int_2^x \frac{A}{\log t + B + o(1)}\,dt \\ &= Ax - (AB + A)\frac{x}{\log x} + o(x/\log x).\end{aligned}$$

b) Conclude from (1.24) that the estimate of a) holds also for $\psi(x)$.

c) Use this estimate for ψ to show that

$$\begin{aligned}\sum_{p^k \le x} \frac{\log p}{p^k} = \int_{3/2}^x \frac{d\psi(t)}{t} &= \int_2^x \frac{\psi(t)}{t^2}\,dt + O(1) \\ &= A\log x - (AB + A)\log\log x + o(\log\log x).\end{aligned}$$

d) Deduce from (1.35) and (1.36) that

$$A = 1, \quad -(AB + A) = 0,$$

so that $A = 1, B = -1$.

1.8 Motivating some Famous Conjectures about Primes

1.8.1 Primes in Arithmetic Progressions

We have already discussed how Gauss conjectured the prime number theorem based on the (empirical) observation that the density of primes near is x seems to behave like $1/\log x$. That theorem is only the first of many conjectures that Gauss' observation motivates, and we take the opportunity here to single out several others. Many of these rank with the most famous unsolved problems in number theory, and seem immune to the analytic machinery deployed so successfully against the prime number theorem.

To warm-up, let us investigate the distribution of primes in the arithmetic progression $a \pmod q$, where where $\gcd(a, q) = 1$ (and $q > 0$). Assuming the density of primes near x is approximately $1/\log x$, our first guess is that the number of primes in the sequence $a + Mq$ with $M \leq z$ should be well-approximated by

$$\int_2^z \frac{dt}{\log(a+tq)} = \int_2^z \left(\frac{1}{\log t} + O\left(\frac{1}{\log^2 t}\right)\right) dt = \mathrm{li}(z) + O\left(\frac{z}{\log^2 z}\right) \sim \mathrm{li}(z).^1$$

But things aren't quite this simple! We are assuming $a + Mq$ has the same likelihood of being prime as an average integer of its size. But if p divides q, then p *never* divides an integer of the progression $a \pmod q$. That is, the probability that p *doesn't* divide an integer in this progression is 1, while for a random integer it's $1 - 1/p$. On the other hand, if p is a prime not dividing q, then p divides an element of the progression with the same probability (viz. $1/p$) as it divides a random integer (since then $a + mq \equiv 0 \pmod p$ has one solution $(\bmod\ p)$). This suggests we multiply our original guess by the correction factor

$$\prod_{p|q} \frac{1}{1-1/p} = \frac{q}{\phi(q)}.$$

Piecing together the above, we arrive at the following:

Conjecture 1.8.1. *Let a, q be integers with $\gcd(a, q) = 1$ and q positive. Then*

$$\#\{M \leq z : a + Mq \text{ is prime}\} \sim \frac{q}{\phi(q)} \mathrm{li}(z) \quad (z \to \infty).$$

This turns out to be a theorem, usually expressed in the equivalent form

$$\#\{p \leq x : p \equiv a \pmod q\} \sim \frac{1}{\phi(q)} \frac{x}{\log x} \quad (x \to \infty).$$

This can be proved by the same analytic methods used to establish the prime number theorem and is known as the *prime number theorem for arithmetic progressions* (for a sketch of one proof see [Els98]).

1.8.2 The Twin Prime and Goldbach Problems

With this under our belt, we turn our attention to counting *twin primes*, primes p such that $p + 2$ is also prime.

We reason as follows: If $(n, n + 2)$ behaved like a random ordered pair of integers of size near n, we would expect both entries to be simultaneously prime with probability $1/\log^2 n$. But we cannot make such an assumption of randomness: the probability that neither entry of a random pair of integers is divisible by p is $(1-1/p)^2$, but the probability that neither n nor $n+2$ is divisible by p is $(1 - \omega(p)/p)$, where $\omega(p)$ is the number of solutions to the congruence

$$n(n + 2) \equiv 0 \pmod p.$$

[1] Of course, the lower range of integration here is unimportant since it only has the effect of shifting the integral by a constant; we choose 2 in order to ensure that the denominator is well-behaved.

Table 1.3: Comparison of $\pi_2(x)$ and $L_2(x) := 2C_2 \int_2^x dt/\log^2 t$

x	$\pi_2(x)$	$L_2(x) - \pi_2(x)$
10^5	1,224	25
10^6	8,169	79
10^7	58,980	-226
10^8	440,312	56
10^9	3,424,506	802
10^{10}	27,412,679	-1,262
10^{11}	224,376,048	-7,183
10^{12}	1,870,585,220	-25,353
10^{13}	15,834,664,872	-66,567
10^{14}	135,780,321,665	-56,771
10^{15}	1,177,209,242,304	-750,443

Since $\omega(p) = 1$ if $p = 2$ and $\omega(p) = 2$ for $p > 2$, we are led to multiply our former guess by the correction factor

$$\frac{1 - 1/2}{(1 - 1/2)^2} \prod_p \frac{1 - 2/p}{(1 - 1/p)^2}.$$

Simplifying this product, we arrive at:

Twin Prime Conjecture (Quantitative Form). *The set of twin primes is infinite. More precisely, as $x \to \infty$,*

$$\pi_2(x) := \#\{p \le x : p + 2 \text{ is also prime}\} \sim 2C_2 \int_2^x \frac{dt}{\log^2 t},$$

where

$$C_2 := \prod\nolimits_{p>2} \left(1 - (p-1)^{-2}\right)$$

is the twin prime constant*.*

Computational evidence for this conjecture (drawn from [Nic03]) is presented in Table 1.8.2.

This heuristic in support of the twin prime conjecture also motivates the Goldbach conjecture; this should not be surprising, since asking for a representation of an even number N in the form $p + p'$ is equivalent to asking that the polynomials n and $N - n$ simultaneously represent prime values. Naively, one guesses this should happen with probability $\frac{1}{\log n \log (N-n)}$. To determine the correction factor, we notice that $n(N - n) \equiv 0 \pmod p$ has one solution when $p \mid N$ and two solutions otherwise. This leads us to multiply our guess by

$$\prod_{p \mid N} \frac{1 - 1/p}{(1 - 1/p)^2} \prod_{p \nmid N} \frac{1 - 2/p}{(1 - 1/p)^2} = 2C_2 \prod_{p \mid N, p > 2} \frac{p - 1}{p - 2},$$

where C_2 is the twin prime constant. Hence we conjecture

$$\#\{p, p' : p + p' = N\} \sim 2C_2 \left(\prod_{p \mid N, p>2} \frac{p-1}{p-2}\right) \int_2^{N-2} \frac{dt}{\log t \log (N-t)}$$

as $N \to \infty$ through even values. It turns out that (exercise!) the integral on the right behaves asymptotically no differently than $\int_2^N dt/\log^2 t$, leading to:

Goldbach Conjecture (Quantitative Form). *As $N \to \infty$ through even numbers, we have*

$$\#\{p, p' : p + p' = N\} \sim 2C_2 \left(\prod_{p \mid N, p>2} \frac{p-1}{p-2}\right) \int_2^{N} \frac{dt}{\log^2 t}. \tag{1.51}$$

Notice that this asymptotic formula implies not only that every large enough even number N is a sum of two primes, but also that the number of representations tends to infinity with N.

Exercise 1.8.1. Let N be a positive, even integer. Give a heuristic suggesting that the number of primes $p \le x$ for which $p + N$ is also prime is asymptotic to

$$2C_2 \left(\prod_{p \mid N, p>2} \frac{p-1}{p-2}\right) \int_2^{x} \frac{dt}{\log^2 t}.$$

Remark. In Chapter 3 we will discuss methods which can be used to establish *upper bounds* for the quantities of this section of the same order of magnitude as the conjectured asymptotics.

1.8.3 An Extended Hardy-Littlewood Conjecture

Our questions about primes in progressions and about twin primes are instances of the following more general query: Suppose $f_1(T), \dots, f_k(T) \in \mathbf{Z}[T]$ are nonconstant and non-associated over $\mathbf{Q}$; are there infinitely many positive integral n for which $f_1(n), \dots, f_k(n)$ are simultaneously prime?[2] If, for given f_i, the answer is yes, then can we say anything about *how many* such n there are below a given bound?

For the answer to our first question to be affirmative, we surely must require each f_i be irreducible over $\mathbf{Z}$ (cf. Exercise 1.8.2). But this is not enough: the polynomial $T^2 + T + 2$ is irreducible over $\mathbf{Z}$ but assumes only even values. A similar *local obstruction* occurs whenever the product $f_1(T) \dots f_k(T)$ possesses a fixed prime divisor, i.e., whenever there exists a fixed prime p dividing $f_1(n) \dots f_k(n)$ for every integer n. For in this case, choose n_0 large enough that $|f_i(n)| > p$ for each i whenever $n > n_0$; then for every $n > n_0$, some $f_i(n)$ must be composite, since some $f_i(n)$ has p as a proper divisor.

So suppose both

[2] For this discussion, it is convenient to define a prime over $\mathbf{Z}$ as any irreducible element, positive or negative.

i. the f_i are all irreducible over $\mathbf{Z}$,

ii. the product $f_1 \ldots f_k$ has no fixed prime divisor.

The conjecture that in this case, the f_i are simultaneously prime for infinitely many positive integer evaluations is called *Hypothesis H*. Originally formulated by Schinzel, it is known to have a number of interesting number-theoretic consequences; a smattering of such appears in [SS58].

Following Bateman & Horn [BH62], we now derive a quantitative version of this conjecture (for an alternative treatment, see Exercise 3.1.1). Special cases of this had been considered earlier by Hardy & Littlewood. Let d_i denote the degree of f_i. Then $\log |f_i|$ is asymptotically $d_i \log |n|$, so that we expect

$$\pi_{f_1,\ldots,f_k}(x) := |\{n \leq x : f_1(n), \ldots, f_k(n) \text{ simultaneously prime}\}|$$

to be asymptotic to

$$C(f_1, \ldots, f_k) \frac{1}{d_1 \ldots d_k} \int_2^x \frac{dt}{\log^k t}, \tag{1.52}$$

where the correction factor $C(f_1, \ldots, f_k)$ is given by the infinite product

$$C(f_1, \ldots, f_k) := \prod_p \frac{1 - \omega(p)/p}{(1 - 1/p)^k}. \tag{1.53}$$

Here

$$\omega(p) = |\{n \pmod p : f_1(n) \ldots f_k(n) \equiv 0 \pmod p\}|.$$

Bateman & Horn show that the above conditions on the f_i imply that the product defining $C(f_1, \ldots, f_k)$ converges (usually conditionally) to a positive constant, so their conjecture indeed implies Hypothesis H. The proof of this unfortunately requires more algebraic number theory than we are assuming here, so we omit it.

The state of our knowledge about these conjectures is rather pathetic. When $k = 1$ and f_1 is linear, the Bateman-Horn conjecture reduces to the prime number theorem for arithmetic progressions. In all other cases, not even the weaker, qualitative assertion of Hypothesis H is known to hold.

We conclude by mentioning that the Bateman-Horn conjecture has recently been generalized [CCG03] to polynomials over $\mathbf{F}_q[u]$. Rather surprisingly, the natural analog of (1.52) is provably false in this situation. As an extreme (and atypical) example of this phenomenon, we mention that when F is a finite field of characteristic p, the polynomial $g^{4p} + u \in F[u]$ is *never* irreducible when $\deg g > 0$, despite the fact that $T^{4p} + u \in F[u][T]$ is nonconstant, irreducible over $F[u]$ and without a fixed prime divisor. In fact, whenever $\deg g > 0$, the polynomial $g^{4p} + u$ factors into an even number of (distinct) irreducibles.

1.8.4 Exercises: More on the Bateman-Horn Conjecture

Exercise 1.8.2. Suppose $f(T) \in \mathbf{Z}[T]$ is a nonconstant polynomial reducible over $\mathbf{Z}$. Show that $f(n)$ assumes at most $2\deg f$ irreducible values as n ranges over the integers.

Exercise 1.8.3 *(Shanks [Sha60]).* Using Hypothesis H, deduce the existence of infinitely many prime pairs $n, n+1+i \in \mathbf{Z}[i]$ (with $n \in \mathbf{Z}$).

Instead of working through Hypothesis H, one can formulate analogous conjectures directly over $\mathbf{Z}[i]$. For such a conjecture treating the case of several monic linear polynomials over an arbitrary algebraic number field, see [GS00].

Exercise 1.8.4 *(Shanks [Sha64], †).* Let $f(z) = \sum_{n=0}^{\infty} z^{n(n+1)/2}$ and define

$$h(z) := f(z)^2 - 3f(z) + 2 = (f(z)-1)^2 - (f(z)-1).$$

Prove that there are infinitely many primes of the form $\frac{n^2+1}{2}$, as predicted by Hypothesis H, if and only if the power series of h (about 0) has infinitely many negative coefficients.

Exercise 1.8.5. Let $f(T)$ be an irreducible polynomial in $F[u][T]$, where F is an *infinite* field. Show that there are automatically no local obstructions: for every prime $\pi \in F[u]$, there exists $t \in F[u]$ with $\pi \nmid f(t)$. Show that such a t may in fact be chosen from F.

Remark. Suppose $F = \mathbf{Q}$. Assume that $f(t) \in F[u][T]$ is irreducible with $\deg_u f > 0$. Then $f(t)$ will be a polynomial in u of degree $\deg_u f$ for all but finitely many rational t. It is a theorem (a simple version of the Hilbert Irreducibility Theorem) that for infinitely many of these t, the corresponding polynomial in u is irreducible. For a relatively elementary proof and some applications to Galois theory, see [Had78, Chapter 4].

Exercise 1.8.6 *(see [RS98]).* Show that there are polynomials $f(T) \in \mathbf{Z}[T]$ of arbitrarily high degree with $f(T)^2 + 1$ irreducible.

Hint: Eisenstein's criterion.

The remaining exercises concern the special case of the Bateman-Horn conjecture where $f_1, \dots, f_k$ are linear polynomials. This (or the corresponding weaker analog of Hypothesis H) is known as the *prime k-tuples conjecture*.

Exercise 1.8.7. In this special case, prove that $\omega(p) = k$ for all but finitely many primes p. Deduce that the infinite product (1.53) converges to a positive real number.

Exercise 1.8.8 *(cf. [Pol]).* Assume the prime k-tuples conjecture. Show that $\mu(n+1) = \mu(n+2) = \mu(n+3) = 1$ for infinitely many natural numbers n.

Suggestion: Consider triples

$$29(30 \cdot 31t + 1), \quad 30(29 \cdot 31t + 1), \quad 31(29 \cdot 30t + 1).$$

Exercise 1.8.9 *(Schinzel & Sierpiński [SS58]).* Assume the prime k-tuples conjecture. Show that there are arbitrarily long arithmetic progressions consisting only of primes. Show that this remains true if we require the primes be consecutive.

Exercise 1.8.10 *(Granville & Mollin [GM00]).* Assume the prime k-tuples conjecture. Show that for every positive integer N, there exists an integer A for which $n^2 - n + A$ assumes prime values for each $0 \le n \le N$.

Suggestion: Consider the N linear polynomials $T + (n^2 - n)$ for $1 \le n \le N$.

1.9 Elementary Prime Number Theory in $\mathbf{Z}[i]$

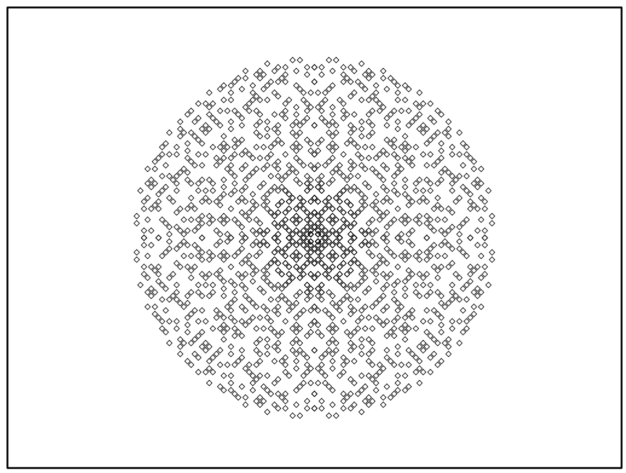


Figure 1.1: The Gaussian Primes with Absolute Value Less than 50.

1.9.1 The Prime Ideal Theorem

In 1903, Landau showed [Lan03] that for any number field K (i.e., any field extension K of $\mathbf{Q}$ with $[K : \mathbf{Q}] < \infty$),

$$\#\{\text{prime ideals } \mathfrak{p} \text{ of } \mathfrak{O}_K : \mathcal{N}\mathfrak{p} \le x\} \sim \frac{x}{\log x},$$

where $\mathfrak{O}_K$ is the "ring of integers" of K (a certain subring of K we won't define here) and the norm $\mathcal{N}I$ of a nonzero ideal I is the size of the (provably finite) quotient ring $\mathfrak{O}_K/I$. In the case when $K = \mathbf{Q}$, this "ring of integers" $\mathfrak{O}_K$ is just the familiar ring $\mathbf{Z}$, and the nonzero prime ideals are just the ideals $p\mathbf{Z}$, different primes p giving rise to different ideals. (This final clause depends on our convention that primes in $\mathbf{Z}$ are always positive.) Since $\mathcal{N}(p\mathbf{Z}) = |\mathbf{Z}/p\mathbf{Z}| = p$, this case of Landau's *prime ideal theorem* amounts to exactly the usual prime number theorem.

In this section we examine the next simplest case, when $K = \mathbf{Q}(i)$. Then $\mathfrak{O}_K$ is the ring of Gaussian integers $\mathbf{Z}[i]$, which bears many similarities to $\mathbf{Z}$. Our goal is to obtain analogs of Chebyshev's theorems (cf. §1.5) by the method of Landau [Lan02], who was the first to carry out this project for arbitrary number rings $\mathfrak{O}_K$.

1.9.2 Chebyshev Analogs

We begin by introducing some notation. We let $\pi_K(x)$ denote the number of Gaussian primes of norm not exceeding x. Analogously to (1.22), we set

$$\theta_K(x) := \sum_{\mathcal{N}\varrho \le x} \log \mathcal{N}\varrho, \quad \psi_K(x) := \theta(x) + \theta(x^{1/2}) + \dots,^3$$

where in this section the letter ϱ always denotes a Gaussian prime. (We would use π, but we will need the constant $3.14159\dots$ later).

Recall that $\mathbf{Z}[i]$ is a principal ideal domain with four units and that the norm of an element is the absolute value of the norm of the ideal it generates. Therefore, we can recast the prime ideal theorem in the form

$$\pi_K(x) \sim 4\frac{x}{\log x}. \tag{1.54}$$

The upper estimate $\pi_K(x) \ll x/\log x$ is easy to obtain from our already-established bounds on $\pi(x)$. Indeed, each Gaussian prime ϱ with norm not exceeding x divides a rational prime p not exceeding x, and each rational prime p has at most 2 prime divisors up to associates, so at most 8 prime divisors total. Hence

$$\pi_K(x) \le 8\pi(x) \ll \frac{x}{\log x} \quad (x \to \infty),$$

by (1.32). By fiddling with the implied constant, this actually holds for all $x \ge 2$. This implies that in the same range both $\theta_K(x)$ and $\psi_K(x)$ are $O(x)$, since

$$\begin{aligned}\theta_K(x) \le \psi_K(x) &= \sum_{k=1}^{\infty} \sum_{\mathcal{N}\varrho \le x^{1/k}} \log \mathcal{N}\varrho = \sum_{\mathcal{N}\varrho^k \le x} \log \mathcal{N}\varrho \\ &= \sum_{\mathcal{N}\varrho \le x} \log \mathcal{N}\varrho \left\lfloor \frac{\log x}{\log \mathcal{N}\varrho} \right\rfloor \le \pi_K(x) \log x \ll x.\end{aligned}$$

The problem of obtaining lower bounds is a bit trickier. For this, we introduce the following analogs of $\lfloor x \rfloor!$ and $T(x)$:

$$F_x := \prod_{\substack{N\alpha \le x \\ \alpha \ne 0}} \alpha, \qquad T_K(x) := \log \mathcal{N} F_x.$$

Let $Z(x)$ denote the number of nonzero elements of norm not exceeding x. This is just the number of lattice points inside or on the circle with radius $\sqrt{x}$ centered at the origin, not counting the origin itself. So by Corollary A.3.3,

$$Z(x) = \pi x + O(\sqrt{x}). \tag{1.55}$$

[3] Beware! these notations are nonstandard; in the literature they denote corresponding counts over prime *ideals*, not elements.

We now estimate $T_K(x) = \sum_{0<N\alpha\le x} \log \mathcal{N}\alpha$ by partial summation:

$$\begin{aligned} T_K(x) = \int_{1/2}^{x} \log t \, dZ(t) &= Z(x)\log x - \int_{1/2}^{x} \frac{Z(t)}{t}\, dt \\ &= \pi x \log x + O(\sqrt{x}\log x) - \int_{1}^{x} \frac{\pi t + O(\sqrt{t})}{t}\, dt \\ &= \pi x \log x - \pi x + O(\sqrt{x}\log x). \end{aligned} \tag{1.56}$$

In order to extract information from this, we need the following lemma, an analog of the usual formula for the highest power of a prime dividing a factorial:

Lemma 1.9.1. *Let ϱ be a Gaussian prime and let x be a positive real number. The exponent on the largest power of ϱ dividing F_x is given by*

$$Z(x/\mathcal{N}\varrho) + Z(x/\mathcal{N}\varrho^2) + Z(x/\mathcal{N}\varrho^3) + \dots .$$

Exercise 1.9.1. Prove this!

Given $x \ge 2$, choose a set of representatives $\mathcal{P}$ of the equivalence classes of Gaussian primes of norm not exceeding x (under the relation of being associates). Then F_x is an associate of

$$\prod_{\substack{\mathcal{N}\varrho\le x\\ \varrho\in\mathcal{P}}} \varrho^{Z(x/\mathcal{N}\varrho)+Z(x/\mathcal{N}\varrho^2)+Z(x/\mathcal{N}\varrho^3)+\dots},$$

so that

$$\begin{aligned} T_K(x) &= \log \mathcal{N}F_x \\ &= \sum_{\substack{\mathcal{N}\varrho\le x\\ \varrho\in\mathcal{P}}} \log\mathcal{N}\varrho \left(Z(x/\mathcal{N}\varrho) + Z(x/\mathcal{N}\varrho^2) + Z(x/\mathcal{N}\varrho^3) + \dots \right) \\ &= \frac{1}{4} \sum_{\mathcal{N}\varrho\le x} \log\mathcal{N}\varrho \left(Z(x/\mathcal{N}\varrho) + Z(x/\mathcal{N}\varrho^2) + Z(x/\mathcal{N}\varrho^3) + \dots \right). \end{aligned} \tag{1.57}$$

Since $Z(x) \ll x$ for all $x > 0$, one has

$$\begin{aligned} \sum_{\mathcal{N}\varrho\le x} \log\mathcal{N}\varrho \sum_{k=2}^{\infty} Z(x/\mathcal{N}\varrho^k) &\ll x \sum_{\mathcal{N}\varrho\le x} \log\mathcal{N}\varrho \sum_{k=2}^{\infty} \frac{1}{\mathcal{N}\varrho^k} \\ &\ll x \sum_{\mathcal{N}\varrho\le x} \frac{\log\mathcal{N}\varrho}{\mathcal{N}\varrho(\mathcal{N}\varrho-1)} \ll x \sum_{2\le n\le x} \frac{\log n}{n(n-1)} \ll x, \end{aligned} \tag{1.58}$$

where in the second line we have used that there are at most $8 = O(1)$ Gaussian primes of any given norm. Combining (1.56), (1.57) and (1.58) yields

$$\frac{1}{4} \sum_{\mathcal{N}\varrho\le x} Z(x/\mathcal{N}\varrho)\log\mathcal{N}\varrho = \pi x \log x + O(x).$$

If we recall that $Z(t) = \pi t + O(\sqrt{t})$ for $t \geq 1$, we have shown that

$$\frac{\pi}{4} x \sum_{\mathcal{N}\varrho \leq x} \frac{\log \mathcal{N}\varrho}{\mathcal{N}\varrho} = \pi x \log x + O(x) + O\left(\sqrt{x} \sum_{\mathcal{N}\varrho \leq x} \frac{\log \mathcal{N}\varrho}{\sqrt{\mathcal{N}\varrho}}\right). \tag{1.59}$$

The final error term is estimated using the upper bound $\pi_K(x) \ll x/\log x$:

$$\begin{aligned}\sum_{\mathcal{N}\varrho \leq x} \frac{\log \mathcal{N}\varrho}{\sqrt{\mathcal{N}\varrho}} = \int_{3/2}^{x} \frac{\log t}{\sqrt{t}}\, d\pi_K(t) &= \frac{\pi_K(x)\log x}{\sqrt{x}} - \int_2^x \pi_K(t) \frac{1-(\log t)/2}{t^{3/2}}\, dt \\ &\ll \sqrt{x} + \int_2^x t^{-1/2}\, dt \ll x^{1/2}.\end{aligned}$$

Substituting into (1.59) and dividing by $\pi x/4$, we obtain

$$\sum_{\mathcal{N}\varrho \leq x} \frac{\log \mathcal{N}\varrho}{\mathcal{N}\varrho} = 4 \log x + O(1). \tag{1.60}$$

It follows by partial summation (exercise!) that

$$\liminf_{x\to\infty} \frac{\pi_K(x)}{x/\log x} \leq 4, \quad \limsup_{x\to\infty} \frac{\pi_K(x)}{x/\log x} \geq 4,$$

so that if the limit exists, it must equal 4. This is the analog of Theorem 1.5.1.

Now choose a constant $C > 0$ with the property that for every $x \geq 2$,

$$\left| \sum_{\mathcal{N}\varrho \leq x} \frac{\log \mathcal{N}\varrho}{\mathcal{N}\varrho} - 4 \log x \right| \leq C.$$

(This is exactly what the estimate (1.60) says we can do!) We set $D = e^C$. Then for $x \geq 2$,

$$\begin{aligned}\sum_{x < \mathcal{N}\varrho \leq Dx} \frac{\log \mathcal{N}\varrho}{\mathcal{N}\varrho} &\geq (4 \log (Dx) - C) - (4 \log (x) + C) \\ &= 4 \log D - 2C = 2C > 0.\end{aligned}$$

Thus there is always a Gaussian prime with norm in the interval $(x, Dx]$; this can be considered an analog of Bertrand's postulate (Theorem 1.5.3). Moreover, since $\log t/t$ is a decreasing function of t for $t \geq 3$, one has for every $x \geq 3$,

$$C \leq \sum_{x < \mathcal{N}\varrho \leq Dx} \frac{\log \mathcal{N}\varrho}{\mathcal{N}\varrho} \leq \frac{\log x}{x} \left(\pi_K(Dx) - \pi_K(x)\right), \tag{1.61}$$

so that there are in fact $\gg x/\log x$ primes with norm falling into this interval. Since a lower bound for $\pi_K(Dx) - \pi_K(x)$ is a lower bound for $\pi_K(Dx)$ also, we see (1.61) implies that for $x \geq 3D$,

$$\pi_K(x) \geq C(x/D)/\log (x/D),$$

so that as $x \to \infty$ we have the estimate

$$\pi_K(x) \gg x/\log x.$$

This final estimate is the analog of the lower bound in Theorem 1.5.2.

1.9.3 Exercises

Exercise 1.9.2. Prove the analog of (1.23) by showing that as $x \to \infty$,

$$\frac{\theta_K(x)}{x} = \frac{\pi_K(x)}{x/\log x} + o(1).$$

Do the same for ψ_K. Deduce that as $x \to \infty$, one has both $\theta_K(x) \asymp x$ and $\psi_K(x) \asymp x$; moreover, the prime ideal theorem for $\mathbf{Z}[i]$ (statement (1.54)) is equivalent to either of $\theta_K(x) \sim 4x$ or $\psi_K(x) \sim 4x$.

Exercise 1.9.3. Owing to the familiar characterization of Gaussian primes, the estimates developed in this section imply certain facts about the distribution of rational primes in progressions $\pmod 4$.

a) Show that as $x \to \infty$, one has

$$\sum_{\substack{p \le x \\ p \equiv 1 \pmod 4}} \frac{\log p}{p} = \frac{1}{2}\log x + O(1).$$

Hint: Note that there are 8 Gaussian primes of norm p for every rational prime $p \equiv 1 \pmod 4$. Show that the primes ϱ dividing some rational prime $p \not\equiv 1 \pmod 4$ contribute a bounded amount to the sum (1.60).

b) Using part a) and estimate (1.38), show that as $x \to \infty$,

$$\sum_{\substack{p \le x \\ p \equiv 1 \pmod 4}} \frac{\log p}{p} = \frac{1}{2}\log x + O(1).$$

c) Show that there is a constant $C > 1$ such that for every $x \ge 2$, there is both a prime $p \equiv 1 \pmod 4$ and a prime $p \equiv 3 \pmod 4$ in the interval $(x, Cx]$.

d) Show that as $x \to \infty$,

$$\pi(x;4,1) \gg x/\log x \quad \text{and} \quad \pi(x;4,3) \gg x/\log x.$$

e) From a) and b), show that if $\frac{\pi(x;4,1)}{x/\log x}$ or $\frac{\pi(x;4,3)}{x/\log x}$ tends to a limit, then that limit must be $1/2$.

When we prove Dirichlet's theorem, we will see that all of these statements generalize to an arbitrary arithmetic progression $a \pmod q$ satisfying $\gcd(a,q) = 1$.

Exercise 1.9.4. Generalize the results of this section to $\mathbf{Z}[\omega]$, where $\omega = e^{2\pi i/3}$.

1.10 The Distribution of Primes in $\mathbf{F}_q[T]$

There is a fruitful analogy between the ring $\mathbf{Z}$ of integers and the ring $F[T]$ of polynomials over a field F. Just as every nonzero integer has a positive associate which factors uniquely as a (possibly empty) product of positive primes, every nonzero polynomial has a monic associate factoring uniquely into monic irreducibles.

The usual prime counting function $\pi(x)$ counts the number of positive irreducible $n \in \mathbf{Z}$ for which $|\mathbf{Z}/n\mathbf{Z}| \leq x$. To introduce an analogous counting function for irreducible polynomials over $F = \mathbf{F}_q$, we first define $\mathcal{N}f := q^{\deg f}$, so that $\mathcal{N}f = |F[T]/(f)|$ for $f \neq 0$. We then define

$$\pi_q(x) := |\{h \in \mathbf{F}[T] : h \text{ monic and irreducible}, \mathcal{N}h \leq x\}|.$$

However, any hope that $\pi_q(x)$ be asymptotic to a smooth, steadily-growing function like $x/\log x$ is shattered by the observation that $\pi_q(x)$ is constant between powers of q and (as one would expect) has rather large jumps at these powers. A more profitable initial object of study is the function $\nu_q(n)$, defined as the number of monic irreducibles of degree n. In fact, we will obtain an exact formula for $\nu_q(n)$, with the consequence that $\nu_q(n) = q^n/n + O(q^{n/2}/n)$. It is this estimate that we call the *prime number theorem*. By partial summation, an asymptotic estimate for $\pi_q(x)$ of order $x/\log x$ as x tends to infinity *along powers of* q follows; this is left to the exercises.

Many questions about the distribution of rational primes can be reformulated in this new context. Somewhat surprisingly, questions which are inaccessible over $\mathbf{Z}$ can sometimes be proven elementarily when stated over $\mathbf{F}_q[T]$.

As an example, we mention one analog of the twin prime problem: Let $F = \mathbf{F}_q$ be a finite field; are there infinitely many "twin primes" $f, f+1 \in F[T]$?

When $q = 2$, there certainly are not. But there are for $q > 3$, as recently demonstrated by C. Hall [Hal03, Corollary 19]. What is actually proved is that in this case, there exists an $\alpha \in \mathbf{F}_q$ and a prime l for which the polynomials

$$T^{l^k} - \alpha - 1, \quad T^{l^k} - \alpha \qquad (k = 1, 2, \dots)$$

are simultaneously irreducible. Our method of proving this irreducibility generalizes to provide a construction of infinitely many twin prime pairs in the single remaining case, that of $q = 3$. We thus have a complete solution to our analog of the twin prime conjecture (but see the more general version of §1.10.4)!

1.10.1 The Prime Number Theorem

It is a remarkable fact that for the ring of polynomials over a finite field $F = \mathbf{F}_q$, the analog of the prime number theorem can be proved in a completely elementary fashion.

Lemma 1.10.1. *Let $h(T) \in F[T]$ be an irreducible polynomial of degree d. Then $h(T) \mid T^{q^n} - T$ if and only if $d \mid n$.*

Proof. Suppose first that $d \mid n$. Since $F[T]/(h(T))$ is a field of size q^d, the q^dth power map is the identity. In particular, $T^{q^d} \equiv T \pmod{h(T)}$, whence $h(T) \mid T^{q^d} - T$. But $d \mid n$ implies $q^d - 1 \mid q^n - 1$, which in turn implies $T^{q^d} - T \mid T^{q^n} - T$.

Now suppose $h(T) \mid T^{q^n} - T$. Choose $g(T) \in F[T]$ with the property that $g(T) \pmod{h(T)}$ is a generator of the multiplicative group $F[T]/(h(T))^*$. Then

$$h(T) \mid T^{q^n} - T \mid g(T^{q^n}) - g(T) = g(T)^{q^n} - g(T),$$

whence $g(T)^{q^n-1} \equiv 1 \pmod{h(T)}$. But the coset of g has order $q^d - 1$, whence $q^d - 1 \mid q^n - 1$ and $d \mid n$. ☐

Lemma 1.10.2. *Over F, we have the factorization*

$$T^{q^n} - T = \prod_{\substack{h \text{ monic, irreducible} \\ \deg h \mid n}} h(T). \tag{1.62}$$

Proof. Write $T^{q^n} - T = p_1(T)p_2(T)\dots p_k(T)$, where the p_i are monic irreducibles. By the Lemma, the p_i occurring are exactly those with degree dividing n. It thus suffices to check that the polynomial on the left is squarefree – but if $A(T)^2 \mid T^{q^n} - T$, then $A(T) \mid (T^{q^n} - T)' = -1$, an absurdity. ☐

Let $\nu_q(n)$ denote the number of monic irreducibles of degree n over F.

Theorem 1.10.3 (Prime Number Theorem for $\mathbf{F}_q[T]$). *The number of monic irreducibles $\nu_q(n)$ of degree $n \geq 1$ over the finite field F with q elements is given by*

$$\nu_q(n) = \frac{1}{n}\sum_{d \mid n} q^d \mu(n/d) = q^n/n + O(q^{n/2}/n).$$

Here the implied constant is absolute.

Proof. We compare the degrees of both sides in the factorization (1.62):

$$q^n = \sum_{d \mid n} d\nu_q(d). \tag{1.63}$$

The stated formula follows by Möbius inversion. To obtain the O-estimate for the error, observe that

$$\begin{aligned} |\nu_q(n) - q^n/n| &= \frac{1}{n}\left|\sum_{d \mid n, d<n} q^d \mu(n/d)\right| \leq \frac{1}{n}\sum_{d=1}^{\lfloor n/2 \rfloor} q^d \\ &\leq \frac{1}{n} q^{\lfloor n/2 \rfloor}\left(1 + \frac{1}{q} + \frac{1}{q^2} + \dots\right) \leq \frac{q^{n/2}}{n}\frac{q}{q-1} \leq 2q^{n/2}/n. \end{aligned}$$ ☐

1.10.2 Exercises: Further Elementary Estimates

For real x, define

$$\pi_q(x) := \sum_{\mathcal{N}h \le x} 1, \quad \theta_q(x) := \sum_{\mathcal{N}h \le x} \log \mathcal{N}h, \quad \psi_q(x) := \sum_{\mathcal{N}h^k \le x} \log \mathcal{N}h,$$

where above (and throughout this series of exercises) h ranges over monic irreducibles of $\mathbf{F}_q[T]$.

Exercise 1.10.1. Use partial summation to deduce the following estimate from Theorem 1.10.3:

$$\pi_q(q^n) = \frac{q^{n+1} - q}{q-1} \frac{1}{n} + O\left(q^n/n^2\right).$$

Thus, setting $x = q^n$, we have

$$\pi_q(x) \sim \frac{q}{q-1} \frac{x}{\log x / \log q} \quad (\text{as } x = q^n \to \infty).$$

Show that this asymptotic estimate does not hold as $x \to \infty$ without restriction.

Exercise 1.10.2 *(Snyder [Sny]).* Show that

$$\psi_q(q^k) = \log q \frac{q^{k+1} - q}{q-1},$$

while the same formula holds for $\theta_q(q^k)$ up to an error of $O(q^{k/2}k)$.

Exercise 1.10.3. Prove that for $x = q^n$, we have

$$\sum_{\mathcal{N}h^k \le x} \frac{\log \mathcal{N}h}{\mathcal{N}h} = \log x.$$

Prove that this estimate remains valid for every $x \ge 1$ up to an error term of $O(1)$. Deduce that

$$\sum_{\mathcal{N}h \le x} \frac{\log \mathcal{N}h}{\mathcal{N}h} = \log x + O(1)$$

for $x \ge 1$, and by applying partial summation that

$$\sum_{\mathcal{N}h \le x} \frac{1}{\mathcal{N}h} = \log\log x + c + O\left(\frac{1}{\log x}\right)$$

for $x \ge 2$ and a certain constant $c = c(q)$.

The next exercise presents an alternate proof of the prime number theorem for $\mathbf{F}_q[T]$. Define

$$\zeta_q(s) := \sum_f \frac{1}{\mathcal{N}f^s}, \tag{1.64}$$

where the sum is over all monic polynomials $f(T) \in F[T]$. We assume at the outset that the terms in the sum are taken with respect to a fixed ordering in which they are arranged by increasing degree.

Exercise 1.10.4.

a) Show that for $\Re(s) > 1$, the sum (1.64) converges absolutely, and

$$\zeta_q(s) = \sum_{n=0}^{\infty} \frac{1}{q^{ns}} q^n = \frac{1}{1 - q^{1-s}}. \tag{1.65}$$

b) Establish the Euler factorization, valid in the same region:

$$\zeta_q(s) = \prod_h \frac{1}{1 - \mathcal{N}h^{-s}} = \prod_{n=1}^{\infty} \left(\frac{1}{1 - q^{-ns}} \right)^{\nu_q(n)}. \tag{1.66}$$

Here the product is taken over all monic irreducibles h.

c) Comparing (1.65) and (1.66), prove that for $|z| < q^{-1}$,

$$\frac{1}{1 - qz} = \prod_{n=1}^{\infty} (1 - z^n)^{-\nu_q(n)}.$$

d) Taking the logarithm of both sides and replacing each term of the form $\log(1 - x)$ by its Taylor series, prove that for $|z| < q^{-1}$,

$$\sum_{N=1}^{\infty} z^N \frac{q^N}{N} = \sum_{N=1}^{\infty} z^N \sum_{nm=N} \frac{n\nu_q(n)}{N}.$$

Rederive (1.63) by comparing coefficients of z^N.

Exercise 1.10.5 *(A Polynomial Analog of Goldbach's Conjecture).* This exercise outlines an elementary argument of Hayes [Hay63] to the effect that if q is large compared to n, every polynomial of degree n in $\mathbf{F}_q[T]$ is a sum of two irreducible polynomials of degree $n + 1$.

a) Let n be a positive integer. Using Theorem 1.10.3, show that if $q > q_0 = q_0(n)$, then

$$\nu_q(n + 1) \geq \frac{1}{2} \frac{q^{n+1}}{n + 1} > q^n.$$

b) Deduce that if $f(T) \in \mathbf{F}_q[T]$ is any polynomial of degree n, where $q > q_0(n)$, then there are two irreducible polynomials h_1, h_2 of degree $n + 1$ which are congruent $\pmod{f(T)}$.

c) Show that $h_1 - h_2 = \beta f$, where $\beta \in \mathbf{F}_q^*$. Thus the identity $f = \beta^{-1} h_1 + (-\beta^{-1}) h_2$ exhibits f as a sum of two irreducibles of degree $n + 1$.

Let us mention another result of Hayes [Hay65] in the same vein: if D is a principal ideal domain with infinitely many nonassociated primes, then every polynomial in $D[T]$ is the sum of two irreducibles. The proof is an elementary but clever application of Eisenstein's irreducibility criterion, and the method extends to any Dedekind domain with infinitely many prime ideals. In particular, in both $\mathbf{Z}[T]$ and $F[T_1, T_2] = F[T_1][T_2]$, every polynomial is a sum of two irreducibles.

1.10.3 New Irreducibles from Old

Let $f(T) \in \mathbf{F}_q[T]$ be an irreducible polynomial not associated to T. The *order of* f is the multiplicative order of any root of f in any extension field containing such a root. This is well-defined, since if $E \supset F$ is an extension containing the root α, then we have an isomorphism

$$\mathbf{F}_q[T]/(f(T)) \cong \mathbf{F}_q(\alpha) \subset E, \qquad T \pmod{f(T)} \mapsto \alpha.$$

Note that the order of f always divides $q^{\deg f} - 1$.

Our goal is the proof of the following result, a special case of [LN97, Theorem 3.35] (for a stronger result, see Exercise 1.10.7):

Theorem 1.10.4. *Let $f(T) \in \mathbf{F}_q[T]$ be an irreducible polynomial of degree d and order e. Suppose l is an odd prime for which $l \mid e, l \nmid (q^d - 1)/e$. Then $f(T^{l^k})$ is irreducible for every $k = 1, 2, 3, \ldots$. The same holds if $l = 2$, provided $q^d \equiv 1 \pmod 4$.*

Remark. It is illuminating to note that the complicated divisibility condition of the theorem, viz. $l \mid e, l \nmid (q^d - 1)/e$, is equivalent to the statement that the roots of f are not lth powers in its splitting field $\mathbf{F}_{q^d}$. I owe this observation to Chris Hall.

We need a series of lemmas:

Lemma 1.10.5. *Let $f(T) \in \mathbf{F}_q[T]$ be an irreducible polynomial of degree d and order e. Then $\mathrm{ord}_e q = d$.*

Proof. We have

$$f(T) \mid T^{q^n} - T \Longleftrightarrow T^{q^n - 1} \equiv 1 \pmod{f(T)} \Longleftrightarrow e \mid q^n - 1 \Longleftrightarrow \mathrm{ord}_e q \mid n.$$

On the other hand, $f(T) \mid T^{q^n} - T$ if and only if $d \mid n$ (Lemma 1.10.1). It follows that d divides $\mathrm{ord}_e q$ and vice-versa. ☐

Our next lemma belongs to elementary number theory:

Lemma 1.10.6. *Let l be an odd prime. Suppose the integer n is such that*

$$n \equiv 1 \pmod{l^a}, \quad \textit{but} \quad n \not\equiv 1 \pmod{l^{a+1}},$$

where $a \geq 1$. If r is a positive integer for which $l^b \| r$, then

$$n^r \equiv 1 \pmod{l^{a+b}}, \quad \textit{but} \quad n^r \not\equiv 1 \pmod{l^{a+b+1}}.$$

The same holds if $l = 2$, provided $n \equiv 1 \pmod 4$, i.e., provided that $a \geq 2$.

Proof. It suffices to establish the lemma for the two cases when r is coprime to l and when $r = l$, the general case following from successive application.

When r is coprime to l, write

$$n^r - 1 = (n - 1)(n^{r-1} + n^{r-2} + \cdots + 1).$$

The last factor is $\equiv r \not\equiv 0 \pmod{l}$, so that the same power of l exactly divides both $n-1$ and n^r-1, and the result follows.

When $r = l$, write $n - 1 = l^a q$, where $l \nmid q$. Then

$$n^r - 1 = (l^a q + 1)^l - 1 = \sum_{j=1}^{l} \binom{l}{j} (l^a q)^j \equiv l^{a+1} q \pmod{l^E},$$

where we may take $E = 2a$ in any case and $E = 2a + 1$ if l is odd. When $l = 2$, we have assumed $a \geq 2$, so that in any case $E > a + 1$. Since $l \nmid q$, it follows that $l^{a+1} \parallel n^r - 1$. □

Proof of Theorem 1.10.4. Let α be a root of $f(T^{l^k})$ in a suitable extension E of $\mathbf{F}_q$. Then α^{l^k} has (multiplicative) order e. Since $l \mid e$, it follows that α has order el^k (see Exercise 1.10.6).

Since we are aiming to prove $f(T^{l^k})$ is irreducible, we would like to establish $f(T^{l^k})$ is in fact the minimal polynomial for α over $\mathbf{F}_q$. Since $f(T^{l^k})$ possesses α as a root, it suffices to show the degree D of the minimal polynomial for α is the same as the degree of $f(T^{l^k})$, i.e., that $D = dl^k$.

Define a by the relation $l^a \parallel e$. Since $l \nmid (q^d - 1)/e$, the same power of l exactly divides both e and $q^d - 1$. The hypotheses of the theorem imply $a \geq 1$ in any case and $a \geq 2$ if $l = 2$.

By Lemma 1.10.5, $d = \mathrm{ord}_e q$ and $D = \mathrm{ord}_{el^k} q$. Thus $d \mid D$. Now $q^{dQ} \equiv 1 \pmod{el^k}$ if and only if both

$$q^{dQ} \equiv 1 \pmod{l^{a+k}} \quad \text{and} \quad q^{dQ} \equiv 1 \pmod{e}.$$

The latter relation imposes no restriction on Q while the former holds if and only if $l^k \mid Q$ (by Lemma 1.10.6). The desired relation $D = dl^k$ follows. □

As an application of Theorem 1.10.4, we prove:

Corollary 1.10.7. *Let $F = \mathbf{F}_q$ be a finite field. Suppose $l \mid q - 1$ for the odd prime l and that $\alpha \in F$ is not an lth power. Then $T^{l^k} - \alpha$ is irreducible over F for every $k = 1, 2, 3, \ldots$. The same holds if $l = 2$, provided $q \equiv 1 \pmod{4}$.*

Proof. Apply Theorem 1.10.4 to the irreducible polynomial $T - \alpha$ of degree $d = 1$. The order e of f is the order of α. Let g generate F^* and write $\alpha = g^j$, so that $l \nmid j$. Then $e = (q-1)/(q-1, j)$. Thus $l \mid e, l \nmid (q-1)/e$. □

Exercise 1.10.6. Let G be a group and suppose $g \in G$ has order e. Suppose that $h \in G$ is a Lth root of G, i.e., satisfies $h^L = g$, and that every prime dividing L also divides e. Then h has order eL.

Exercise 1.10.7. Let $f(T) \in \mathbf{F}_q[T]$ be an irreducible polynomial of degree d and order e. Suppose L is chosen so that every prime $l \mid L$ also satisfies $l \mid e, l \nmid (q^d - 1)/e$. Moreover, assume $q^d \equiv 1 \pmod{4}$ if $L \equiv 0 \pmod{4}$. Then $f(t^L)$ is irreducible.

1.10.4 The Twin Prime Problem

Let $F = \mathbf{F}_q$ be a finite field and fix $d \in F[T]$. In analogy with the twin prime conjecture over $\mathbf{Z}$, it is natural to ask if there are infinitely many prime pairs $f, f + d \in F[T]$.

To see when this is a sensible question, let us first rule out local obstructions. When is there is a prime $h \in F[T]$ which divides $f(f + d)$ for every f? The only way this can happen is if $0, -d$ exhaust the residue classes $\pmod h$, so that $\mathcal{N}h = 2$. This can only happen if $q = 2$ and $\deg h = 1$, so that $h = T$ or $T + 1$.

So let us assume now that either $q \neq 2$ or that $T(T + 1) \mid d$. Then one conjectures (cf. [CCG03, Conjecture 10.10])

$$\begin{aligned}&|\{g \in F[T] : \deg f = n, f \text{ and } f + d \text{ both irreducible}\}| \\ &\qquad \sim \prod_{h \nmid d} \frac{1 - 2/\mathcal{N}h}{(1 - 1/\mathcal{N}h)^2} \prod_{h \mid d} (1 - 1/\mathcal{N}h)^{-1} \frac{(q-1)q^n}{n^2} \quad (n \to \infty), \quad (1.67)\end{aligned}$$

where the products are over monic irreducibles h. The infinite product here converges (exercise), so we expect infinitely many twin primes. Note that the sampled f are *not* required to be monic; this explains the presence of the factor $q - 1$.

As with the analogous conjecture over $\mathbf{Z}$ (cf. Exercise 1.8.1), a proof of (1.67) seems hopelessly out of reach. But whereas over $\mathbf{Z}$, there are no d for which the set of prime pairs $h, h + d$ is provably infinite, the construction of §1.10.3 implies this (actually, a little more) is true for $\mathbf{F}_q[T]$ when $d = 1$:

Theorem 1.10.8. *Let $F = \mathbf{F}_q$, where $q \neq 2$. Then there exist infinitely many monic twin prime pairs $h, h + 1 \in F[T]$.*

The idea of the proof is to show that for almost all q, one can find consecutive lth power nonresidues $\alpha, \alpha + 1 \in \mathbf{F}_q$ for some prime $l \mid q - 1$ (where additionally $4 \mid q - 1$ if $l = 2$). Then by Corollary 1.10.7,

$$T^{l^k} - \alpha - 1, \quad T^{l^k} - \alpha \qquad (k = 1, 2, \dots)$$

are an infinite family of twin prime pairs. In the single case where this strategy does not go through, we will use the construction of the last section to exhibit another family of twin prime pairs.

Executing this plan requires a series of lemmas. The first of these has a combinatorial flavor; it shows there are two consecutive non-lth powers provided there are not too many lth powers.

Lemma 1.10.9. *Let $S \subset F = \mathbf{F}_q$, and let $b \in F^*$. Suppose that for each $a \in F$, either a or $a + b$ is an element of S. Then $|S| \geq q/2$.*

Proof. We have

$$\begin{aligned} q = \sum_{a\in F} 1 &\le \sum_{a\in F} |S\cap\{a,a+b\}| \\ &= \sum_{a\in F}\sum_{c\in S\cap\{a,a+b\}} 1 = \sum_{c\in S}\sum_{\substack{a\in F\\ c\in\{a,a+b\}}} 1 = \sum_{s\in S} 2 = 2|S|. \qquad \square \end{aligned}$$

When $l \mid q-1$, the number of lth powers in $\mathbf{F}_q$ is $(q-1)/l+1$, and this exceeds $q/2$ when $l=2$. Consequently, if 2 is the only prime dividing $q-1$, then the preceding lemma cannot be used to guarantee consecutive lth power nonresidues for some prime $l \mid q-1$. Our next lemma tells us precisely when we are in this unfortunate situation:

Lemma 1.10.10. *Suppose n is positive integer for which $2^n+1=p^k$ is a prime power. Then either $k=1$ or $n=3, p=3, k=2$.*

Proof. Suppose n,k is a solution. If k is odd, write

$$2^n = p^k - 1 = (p-1)(p^{k-1}+\cdots+p+1).$$

Since p is clearly odd, the final factor is congruent to $k \pmod 2$, so is also odd. As an odd divisor of 2^n, it must equal unity. It follows that $k=1$.

If k is even, then

$$2^n = (p^{k/2}-1)(p^{k/2}+1),$$

forcing the two factors on the right to be powers of 2 differing by 2. It follows that $p^{k/2}-1=2$, whence $p=3, k=2$, and $n=3$. $\square$

Lemma 1.10.11. *Let $F=\mathbf{F}_q$, where $q\neq 2,3$. Then for some prime $l \mid q-1$, there exist consecutive lth power nonresidues $\alpha, \alpha+1\in F$. Moreover, we can choose $l\neq 2$ except possibly when $q\equiv 1 \pmod 4$.*

Proof. We consider two cases, depending on whether or not $q-1$ possesses an odd prime factor l. Suppose it does and that l is *any* such prime factor; we claim that there are two consecutive lth power nonresidues. Otherwise, with S as the set of lth powers,

$$1+\frac{q-1}{3} \ge 1+\frac{q-1}{l} = |S| \ge q/2,$$

so that $q\le 4$. However, when $q=4$, we necessarily have $l=3$, and $\mathbf{F}_q^l = \mathbf{F}_q^3 = \{0,1\}\subset \mathbf{F}_2$. In particular, if we choose $\alpha\in\mathbf{F}_4\setminus\mathbf{F}_2$, then $\alpha,\alpha+1$ are consecutive cubic nonresidues.

Suppose now that $q-1$ is a power of 2. Then necessarily $q\equiv 1 \pmod 4$. In fact, by Lemma 1.10.10, either $q=9$ or q is a Fermat prime. In the former case, we make the identification

$$\mathbf{F}_9 = \mathbf{F}_3(\beta), \quad \beta^2+1=0.$$

There are five squares in $\mathbf{F}_9$, namely $0, 1, \beta^2 = -1, (1+\beta)^2 = 2\beta, \beta^2(1+\beta)^2 = \beta$. Thus $\beta + 1, \beta + 2$ are consecutive quadratic nonresidues.

Finally, suppose $q = 2^{2^k} + 1$ is a Fermat prime, where k is a positive integer. If $k = 1$, then $q = 5$ and $2, 3$ are consecutive nonsquares. Otherwise, we shall prove

$$\left(\frac{2}{q}\right) = 1, \quad \left(\frac{3}{q}\right) = \left(\frac{5}{q}\right) = -1.$$

It follows that $5, 6$ are a pair of consecutive quadratic nonresidues. To verify the status of these Legendre symbols, note first that $q \equiv 1 \pmod 8$, whence 2 is a square. The other claims follow from quadratic reciprocity and the congruences $q \equiv 2 \pmod 3, q \equiv 2 \pmod 5$, which are easily verified by induction on k. □

Proof of Theorem 1.10.8. For $q \neq 3$, the theorem follows from Corollary 1.10.7 and Lemma 1.10.11. When $q = 3$, we apply Theorem 1.10.4 directly to the irreducible polynomials $T^3 - T + 1$ and $T^3 - T + 2$ of degree $d = 3$.

With f denoting either of these polynomials, the order e of f is a divisor of $3^3 - 1 = 26$ exceeding 2 (since the roots do not live in F). It follows that we may take $l = 13$ in Theorem 1.10.4, so that

$$T^{3\cdot 13^k} - T^{13^k} + 1, \quad T^{3\cdot 13^k} - T^{13^k} + 2$$

are a twin prime pair for each $k = 1, 2, 3, \ldots$. □

Since we can scale the twin prime pair $h, h + 1$ by any unit α, Theorem 1.10.8 has the following easy consequence, mentioned in the introduction to this section:

Corollary 1.10.12. *Let $F = \mathbf{F}_q$, where $q \neq 2$, and let $\alpha \in F^*$. Then there are infinitely many twin prime pairs $h, h + \alpha \in F[T]$.*

This is a qualitative improvement on an earlier sieve result of Cherly [Che78, Theorem 1.2] that there are infinitely many pairs $f, f+\alpha$, each member of which has at most four (non-associated) prime divisors.

The preceding corollary does not guarantee there are infinitely many such *monic* twin prime pairs. Nevertheless, this is true:

Theorem 1.10.13. *Let $F = \mathbf{F}_q$, where $q \neq 2$, and let $\alpha \in F^*$. Then there are infinitely many monic twin prime pairs $h, h + \alpha \in F[T]$.*

The proof of this is outlined in the exercises.

1.10.5 Exercises: Proof of Theorem 1.10.13

Exercise 1.10.8. Let q and α be as in Theorem 1.10.13. Suppose $q - 1$ is not a power of 2, so that it is divisible by some odd prime l. By imitating the proof of Lemma 1.10.11, show that if $q > 4$, there is a pair of lth power nonresidues differing by α. Now invoke Corollary 1.10.7 to prove that the conclusion of Theorem 1.10.13 holds for q.

Exercise 1.10.9. Suppose $q = 4$ and make the identification

$$\mathbf{F}_4 = \mathbf{F}_2(\beta), \quad \beta^2 + \beta + 1 = 0.$$

Thus the nonzero elements of $\mathbf{F}_4$ are $1, \beta, \beta + 1$. The case of Theorem 1.10.13 when $q = 4$, $\alpha = 1$ is covered by Theorem 1.10.8, so we may assume $\alpha = \beta$ or $\alpha = \beta + 1$. Use Theorem 1.10.4 to prove that

$$T^{2\cdot 5^k} + (\beta+1)T^{5^k} + 1, \quad T^{2\cdot 5^k} + (\beta+1)T^{5^k} + (\beta+1) \qquad (k = 1, 2, 3, \dots)$$

describes an infinite family of twin prime pairs differing by β.

To handle the case of pairs differing by $\beta + 1$, consider the automorphism of $\mathbf{F}_4[T]$ induced by the nontrivial automorphism of $\mathbf{F}_4$.

By Lemma 1.10.10, the only remaining cases of Theorem 1.10.13 are when $q = 9$ or q is a Fermat prime.

Exercise 1.10.10. By direct computation, establish that every element of $\mathbf{F}_9$ is a difference of two nonsquares. Now deduce the case $q = 9$ of Theorem 1.10.13 from Corollary 1.10.7.

Exercise 1.10.11. Let p be an odd prime and c an integer coprime to p. Define

$$S := \sum_{a \pmod p} \left(\frac{a}{p}\right)\left(\frac{a+c}{p}\right).$$

Using Euler's criterion $\left(\frac{n}{p}\right) \equiv n^{(p-1)/2} \pmod p$, show that $S \equiv -1 \pmod p$. Show also that $|S| \leq p - 2$. Conclude $S = -1$.

Exercise 1.10.12. Let p be an odd prime and let $\mathcal{N}$ be the set of nonsquares in $\mathbf{F}_p$. Show that $\mathcal{N} - \mathcal{N}$ is all of $\mathbf{F}_p$ whenever $p \geq 7$. Suggestion: For fixed c, relate the number of pairs of quadratic nonresidues with difference c to the sum

$$\sum_{a \pmod p} \left(1 - \left(\frac{a}{p}\right)\right)\left(1 - \left(\frac{a+c}{p}\right)\right);$$

then invoke the result of the previous exercise. For an alternate proof, see Chapter 5, Exercise 5.2.2.

Deduce from Corollary 1.10.7 that Theorem 1.10.13 holds if q is a Fermat prime and $q \geq 7$.

Exercise 1.10.13. It remains only to treat the cases $q = 3$ and $q = 5$. Suppose $q = 3$. Then by replacing α with $-\alpha$ we can assume $\alpha = 1$, so that the result follows from Theorem 1.10.8. When $q = 5$, we can similarly assume $\alpha = 1$ or $\alpha = 2$. The former case follows from Theorem 1.10.8; to handle the latter, use Theorem 1.10.4 to prove that

$$T^{3\cdot 31^k} + T^{31^k} + 4, \quad T^{3\cdot 31^k} + T^{31^k} + 1 \qquad (k = 1, 2, 3, \dots)$$

describes an infinite family of twin prime pairs differing by 2.

References

[AB03] J. M. Aldaz and A. Bravo, *Euclid's argument on the infinitude of primes*, Amer. Math. Monthly **110** (2003), no. 2, 141–142. MR 2003k:11009

[Apo76] Tom M. Apostol, *Introduction to analytic number theory*, Springer-Verlag, New York, 1976, Undergraduate Texts in Mathematics. MR 55 #7892

[AS93] Ulrich Abel and Hartmut Siebert, *Sequences with large numbers of prime values*, Amer. Math. Monthly **100** (1993), no. 2, 167–169. MR 94b:11092

[Bel43] Richard Bellman, *A note on the divergence of a series*, Amer. Math. Monthly **50** (1943), 318–319.

[BH62] Paul T. Bateman and Roger A. Horn, *A heuristic asymptotic formula concerning the distribution of prime numbers*, Math. Comp. **16** (1962), 363–367. MR 26 #6139

[Bro76] J. L. Brown, Jr., *Generalization of Richert's theorem*, Amer. Math. Monthly **83** (1976), no. 8, 631–634. MR 55 #2741

[CCG03] Brian Conrad, Keith Conrad, and Robert Gross, *Irreducible specializations in genus 0*, Preliminary Draft, May 2003.

[Cha02] Robin Chapman, *Evaluating* $\zeta(2)$, available online from `http://www.maths.ex.ac.uk/~rjc/rjc.html`, 2002.

[Che51] P.L. Chebyshev, *Sur la fonction qui detérmine la totalité des nombres premiers inférieurs à une limite donnée*, Mémoires présentés à l'Académie Impériale des Sciences de St. Pétersbourg par divers Savants **6** (1851), 141–157.

[Che52] ———, *Mémoire sur les nombres premiers*, Journal de Mathématique pures et appliquées **17** (1852), 366–390.

[Che78] J. Cherly, *A lower bound theorem in* $F_q[x]$, J. Reine Angew. Math. **303/304** (1978), 253–264. MR 80e:12022

[CL77] Gerard J. Chang and Ko Wei Lih, *Polynomial representation of primes*, Tamkang J. Math. **8** (1977), no. 2, 197–198. MR 58 #27718

[Coh80] Harvey Cohn, *Advanced number theory*, Dover Publications Inc., New York, 1980, Reprint of *A second course in number theory*, 1962, Dover Books on Advanced Mathematics. MR 82b:12001

[Dav00] Harold Davenport, *Multiplicative number theory*, third ed., Graduate Texts in Mathematics, vol. 74, Springer-Verlag, New York, 2000, Revised and with a preface by Hugh L. Montgomery. MR 2001f:11001

[Dia82] Harold G. Diamond, *Elementary methods in the study of the distribution of prime numbers*, Bull. Amer. Math. Soc. (N.S.) **7** (1982), no. 3, 553–589. MR 83m:10002

[Dic66] Leonard Eugene Dickson, *History of the theory of numbers. Vol. I: Divisibility and primality.*, Chelsea Publishing Co., New York, 1966. MR 39 #6807a

[Dre75] Robert E. Dressler, *A lower bound for $\pi(n)$*, Amer. Math. Monthly **82** (1975), 151–152. MR 50 #9768

[Els98] Jürgen Elstrodt, *A quick proof of the prime number theorem for arithmetic progressions*, Charlemagne and his heritage. 1200 years of civilization and science in Europe, Vol. 2 (Aachen, 1995), Brepols, Turnhout, 1998, pp. 521–530. MR 99m:11102

[ET48] P. Erdös and P. Turán, *On some new questions on the distribution of prime numbers*, Bull. Amer. Math. Soc. **54** (1948), 371–378. MR 9,498k

[Eul37] L. Euler, *Variae observationes circa series infinites*, Comm. Acad. Petropolitanae **9** (1737), 160–188.

[Fur55] Harry Furstenberg, *On the infinitude of primes*, Amer. Math. Monthly **62** (1955), 353. MR 16,904e

[Gar90] Betty Garrison, *Polynomials with large numbers of prime values*, Amer. Math. Monthly **97** (1990), no. 4, 316–317. MR 91i:11124

[Gau73a] Carl Friedrich Gauß, *Werke. Band II*, Georg Olms Verlag, Hildesheim, 1973, Reprint of the 1863 original. MR 82e:01122b

[Gau73b] ———, *Werke. Band X. Abt. I, II*, Georg Olms Verlag, Hildesheim, 1973, Reprint of the 1917 and the 1922–1933 originals. MR 82e:01122j

[GM00] Andrew Granville and Richard A. Mollin, *Rabinowitsch revisited*, Acta Arith. **96** (2000), no. 2, 139–153. MR 2002a:11109

[Gol59] Solomon W. Golomb, *A connected topology for the integers*, Amer. Math. Monthly **66** (1959), 663–665. MR 21 #6347

[Gol73] L. J. Goldstein, *A history of the prime number theorem*, Amer. Math. Monthly **80** (1973), 599–615. MR 47 #1726

[Gol98] Robert Goldblatt, *Lectures on the hyperreals*, Graduate Texts in Mathematics, vol. 188, Springer-Verlag, New York, 1998, An introduction to nonstandard analysis. MR 2000a:03113

[GS00] Robert Gross and John H. Smith, *A generalization of a conjecture of Hardy and Littlewood to algebraic number fields*, Rocky Mountain J. Math. **30** (2000), no. 1, 195–215. MR 2001g:11175

[Guy94] Richard K. Guy, *Unsolved problems in number theory*, second ed., Problem Books in Mathematics, Springer-Verlag, New York, 1994, Unsolved Problems in Intuitive Mathematics, I. MR 96e:11002

[Had78] Charles Robert Hadlock, *Field theory and its classical problems*, Carus Mathematical Monographs, vol. 19, Mathematical Association of America, Washington, D.C., 1978. MR 82c:12001

[Hal03] Christopher J. Hall, *L-functions of twisted Legendre curves*, Ph.D. thesis, Princeton University, 2003.

[Har56] V. C. Harris, *Another proof of the infinitude of primes*, Amer. Math. Monthly **63** (1956), 711. MR 18,379g

[Has50] Helmut Hasse, *Vorlesungen über Zahlentheorie*, Die Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen mit besonderer Berücksightigung der Anwendungsgebiete. Band LIX, Springer-Verlag, Berlin, 1950. MR 14,534c

[Hay63] David Hayes, *A polynomial analog of the Goldbach conjecture*, Bull. Amer. Math. Soc. **69** (1963), 115–116. MR 26 #109

[Hay65] D. R. Hayes, *A Goldbach theorem for polynomials with integral coefficients*, Amer. Math. Monthly **72** (1965), 45–46. MR 30 #2001

[Hem66] R.L. Hemminger, *More on the infinite primes theorem*, Amer. Math. Monthly **73** (1966), 1001–1002.

[Hir02] Michael D. Hirschhorn, *There are infinitely many prime numbers*, Austral. Math. Soc. Gaz. **29** (2002), no. 2, 103. MR 1 910 511

[Hua82] Loo Keng Hua, *Introduction to number theory*, Springer-Verlag, Berlin, 1982, Translated from the Chinese by Peter Shiu. MR 83f:10001

[Kno75] John Knopfmacher, *Abstract analytic number theory*, North-Holland Publishing Co., Amsterdam, 1975, North-Holland Mathematical Library, Vol. 12. MR 54 #7404

[KP97] John Knopfmacher and Stefan Porubsky, *Topologies related to arithmetical properties of integral domains*, Exposition. Math. **15** (1997), no. 2, 131–148. MR 98g:13029

[Lan02] Edmund Landau, *Über die zu einem algebraischen Zahlkörper gehörige Zetafunction und die Ausdehnung der Tschebyschefschen Primzahlentheorie auf das Problem der Verteilung der Primideale.*, J. Reine Angew. Math. **128** (1902), 64–188.

[Lan03] ———, *Neuer Beweis des Primzahlsatzes und Beweis des Primidealsatzes*, Math. Ann. **56** (1903), 645–670.

[Lan53] ______, *Handbuch der Lehre von der Verteilung der Primzahlen. 2 Bände*, Chelsea Publishing Co., New York, 1953, 2d ed, With an appendix by Paul T. Bateman. MR 16,904d

[LN97] Rudolf Lidl and Harald Niederreiter, *Finite fields*, second ed., Encyclopedia of Mathematics and its Applications, vol. 20, Cambridge University Press, Cambridge, 1997, With a foreword by P. M. Cohn. MR 97i:11115

[Lor96] Dino Lorenzini, *An invitation to arithmetic geometry*, Graduate Studies in Mathematics, vol. 9, American Mathematical Society, Providence, RI, 1996. MR 97e:14035

[Mer74] F. Mertens, *Ein Beitrag zur analytischen Zahlentheorie*, J. Reine Angew. Math. **78** (1874), 46–62.

[Moh78] S. P. Mohanty, *The number of primes is infinite*, Fibonacci Quart. **16** (1978), no. 4, 381–384. MR 80c:10009

[Moh79] ______, *The number of primes is infinite*, Bull. Math. Assoc. India **11** (1979), no. 1-2, 62–68. MR 82f:10001

[Mos58] Leo Moser, *On the series,* $\sum 1/p$, Amer. Math. Monthly **65** (1958), 104–105. MR 21 #232

[Mur01] M. Ram Murty, *Problems in analytic number theory*, Graduate Texts in Mathematics, vol. 206, Springer-Verlag, New York, 2001, Readings in Mathematics. MR 2001k:11002

[Nag64] Trygve Nagell, *Introduction to number theory*, Second edition, Chelsea Publishing Co., New York, 1964. MR 30 #4714

[Nai82] M. Nair, *On Chebyshev-type inequalities for primes*, Amer. Math. Monthly **89** (1982), no. 2, 126–129. MR 83f:10043

[Nar00] Władysław Narkiewicz, *The development of prime number theory*, Springer Monographs in Mathematics, Springer-Verlag, Berlin, 2000, From Euclid to Hardy and Littlewood. MR 2001c:11098

[Nic03] Thomas Nicely, *Counts of twin-prime pairs and Brun's constant to* $3 \cdot 10^{15}$, online at `http://www.trnicely.net/twins/tabpi2.html`, 2003.

[Pin80] János Pintz, *On Legendre's prime number formula*, Amer. Math. Monthly **87** (1980), no. 9, 733–735. MR 82e:01055

[Pol] P. Pollack, *Some consequences of the prime k-tuples conjecture*, in preparation.

[Ram19] S. Ramanujan, *A proof of Bertrand's postulate*, J. Indian Math. Soc. **11** (1919), 181–182.

[Rib96] Paulo Ribenboim, *The new book of prime number records*, Springer-Verlag, New York, 1996. MR 96k:11112

[Ric49] Hans-Egon Richert, *Über Zerfällungen in ungleiche Primzahlen*, Math. Z. **52** (1949), 342–343. MR 11,502b

[Rie59] B. Riemann, *Ueber die Anzahl der Primzahlen unter einer gegebenen Grösse*, Monatsberichte der Berliner Akademie (1859), 671–680.

[RS98] A. Rattan and C. Stewart, *Goldbach's conjecture for* $\mathbf{Z}[x]$, C. R. Math. Acad. Sci. Soc. R. Can. **20** (1998), no. 3, 83–85. MR 99f:11132

[Rub93] M. Rubinstein, *A formula and a proof of the infinitude of the primes*, Amer. Math. Monthly **100** (1993), 388–392.

[Sha60] Daniel Shanks, *A note on Gaussian twin primes*, Math. Comput. **14** (1960), 201–203. MR 22 #2586

[Sha64] ———, *An analytic criterion for the existence of infinitely many primes of the form* $\frac{1}{2}(n^2+1)$, Illinois J. Math. **8** (1964), 377–379. MR 29 #3423

[Sha83] Harold N. Shapiro, *Introduction to the theory of numbers*, Pure and Applied Mathematics, John Wiley & Sons Inc., New York, 1983, A Wiley-Interscience Publication. MR 84f:10001

[Sie64] W. Sierpiński, *Les binômes* x^2+n *et les nombres premiers*, Bull. Soc. Roy. Sci. Liège **33** (1964), 259–260. MR 29 #4746

[Sny] Noah Snyder, *On the distribution of primes in* $\mathbf{F}_q[x]$, Harvard tutorial paper. Unpublished.

[SS58] A. Schinzel and W. Sierpiński, *Sur certaines hypothèses concernant les nombres premiers*, Acta Arith. 4 (1958), 185–208; erratum **5** (1958), 259. MR 21 #4936

[VE80] Charles Vanden Eynden, *Proofs that* $\sum 1/p$ *diverges*, Amer. Math. Monthly **87** (1980), no. 5, 394–397. MR 81g:10013

[Wun65] M. Wunderlich, *Another proof of the infinite primes theorem*, Amer. Math. Monthly **72** (1965), 305.

[Zag97] D. Zagier, *Newman's short proof of the prime number theorem*, Amer. Math. Monthly **104** (1997), no. 8, 705–708. MR 98j:11069

Chapter 2

Dirichlet's Theorem

2.1 Introduction and a Special Case

In this chapter we prove that whenever $\gcd(a, q) = 1$, there exist infinitely many primes $p \equiv a \pmod q$. This is the famous theorem of Dirichlet [Dir37] on primes in an arithmetic progression. Actually, we shall prove more, namely that for $x \geq 4$,

$$\sum_{\substack{p \leq x \\ p \equiv a \pmod q}} \frac{\log p}{p} = \frac{1}{\phi(q)} \log x + O(1). \tag{2.1}$$

The infinitude of the primes $p \equiv a \pmod q$ is of course an easy consequence of this, but (2.1) says much more. In view of (1.38), it says that in a certain sense the fraction of primes falling into the given residue class is exactly $1/\phi(q)$. The prime number theorem for arithmetic progressions – i.e., the assertion that for fixed a and q,

$$\pi(x; q, a) \sim \frac{1}{\phi(q)} \frac{x}{\log x} \qquad (x \to \infty) \tag{2.2}$$

– implies this is actually true in the sense of relative asymptotic density.

Both the prime number theorem and the corresponding equidistribution statement are difficult, and we shall not prove either here. It is worth noting that estimate (2.1) does allow us to prove a weaker statement in the direction of (2.2), namely

$$\pi(x; q, a) \gg_{a,q} \frac{x}{\log x}, \tag{2.3}$$

which can be considered an analog of Chebyshev's lower bounds from the preceding chapter. For the deduction of (2.3) from (2.1), see the exercises of §2.6; we will concentrate our efforts in the text on the proof of (2.1).

In order to illustrate the usefulness of Dirichlet's theorem, we conclude the chapter with a proof of Legendre's characterization of the integers expressible as a sum of three squares.

The proof of Dirichlet's theorem given in this chapter is due to H.N. Shapiro ([Sha50], [Sha83, Chapter 9]). Our exposition is closely based on the proof as described in [GL66, §3.2], but our treatment of the nonvanishing of $L(1,\chi)$ for real χ incorporates simplifications due to Monsky [Mon93]. For other presentations of this proof, see also [Apo76, Chapter 7], [Nat00, Chapter 10]. The usual complex-analytic proof is well-described in [IR90, Chapter 16]. Our argument for the three squares theorem is that of Ankeny [Ank57], incorporating the insights of Mordell [Mor58].

2.1.1 The Case of Progressions (mod 4)

Developing the tools necessary to handle the general case will require a few sections of preparation, but we can already give the proof of (2.1) in the case of progressions mod 4. Note that if q is any positive integer and a is any integer, one has

$$\begin{aligned}\sum_{\substack{n\le x\\ n\equiv a \pmod q}} \frac{\Lambda(n)}{n} &= \sum_{\substack{p^k\le x\\ p^k\equiv a \pmod q}} \frac{\log p}{p^k}\\ &= \sum_{\substack{p\le x\\ p\equiv a \pmod q}} \frac{\log p}{p} + \sum_{k\ge 2}\sum_{\substack{p\le x^{1/k}\\ p^k\equiv a \pmod q}} \frac{\log p}{p^k}.\end{aligned}$$

Moreover, the double sum is bounded above by the finite (absolute) constant

$$\sum_{k\ge 2}\sum_{n\ge 2}\frac{\log n}{n^k} = \sum_{n\ge 2}\frac{\log n}{n(n-1)}.$$

Consequently,

$$\sum_{\substack{p\le x\\ p\equiv a \pmod q}} \frac{\log p}{p} = \sum_{\substack{n\le x\\ n\equiv a \pmod q}} \frac{\Lambda(n)}{n} + O(1). \tag{2.4}$$

Thus estimates for $\sum \log p/p$ over any residue class mod q will follow from corresponding estimates for $\sum \Lambda(n)/n$, which are easier to obtain.

Now specialize by taking $q=4$. Define the functions χ and χ_0 by

$$\chi(n) = \begin{cases}(-1)^{(n-1)/2} & \text{if } 2\nmid n,\\ 0 & \text{otherwise.}\end{cases} \quad \text{and} \quad \chi_0(n) = \begin{cases}1 & \text{if } 2\nmid n,\\ 0 & \text{otherwise.}\end{cases}$$

It is straightforward to check that with this definition, one has $\chi(ab)=\chi(a)\chi(b)$ for every pair of integers a,b.

These functions are useful to us because $\chi_0+\chi$ is twice the characteristic function of the arithmetic progression 1 (mod 4), and $\chi_0-\chi$ is twice the characteristic function of the arithmetic progression 3 (mod 4). This suggests studying the summatory functions

$$\sum_{n\leq x}\frac{\chi_0(n)\Lambda(n)}{n}, \qquad \sum_{n\leq x}\frac{\chi(n)\Lambda(n)}{n}. \tag{2.5}$$

The first of these behaves very much like the sum $\sum_{n\leq x}\Lambda(n)/n$ investigated in the last chapter:

$$\begin{aligned}\sum_{n\leq x}\frac{\chi_0(n)\Lambda(n)}{n} &= \sum_{n\leq x}\frac{\Lambda(n)}{n} - \sum_{2^k\leq x}\frac{\log 2}{2^k}\\ &= \sum_{n\leq x}\frac{\Lambda(n)}{n} + O(1) = \log x + O(1),\end{aligned}\tag{2.6}$$

the final equality coming from (1.35).

To investigate the second sum, we introduce the series

$$L := 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \cdots = \sum_{n=1}^{\infty}\frac{\chi(n)}{n}.$$

Grouping terms, we notice $L = (1-1/3) + (1/5-1/7) + \cdots > 0$.

Since L is an alternating series with terms decreasing in absolute value, if we use N to denote the smallest odd integer exceeding x, then for every $x \geq 1$,

$$\left|\sum_{n>x}\frac{\chi(n)}{n}\right| \leq \left|\frac{\chi(N)}{N}\right| = \frac{1}{N} \leq \frac{1}{x}. \tag{2.7}$$

We can now estimate the second function appearing in (2.5) by an ingenious device of Mertens. We note that

$$\begin{aligned}\sum_{n\leq x}\frac{\chi(n)\log(n)}{n} &= \sum_{n\leq x}\frac{\chi(n)}{n}\sum_{d\mid n}\Lambda(d)\\ &= \sum_{d\leq x}\Lambda(d)\sum_{\substack{n\leq x\\ d\mid n}}\frac{\chi(n)}{n}\\ &= \sum_{de\leq x}\frac{\chi(de)\Lambda(d)}{de} = \sum_{d\leq x}\frac{\chi(d)\Lambda(d)}{d}\sum_{e\leq x/d}\frac{\chi(e)}{e}.\end{aligned}$$

The inner sum here is equal to $L - \sum_{e>x/d}\chi(e)/e = L + O(d/x)$. Substituting this estimate above yields

$$\begin{aligned}\sum_{n\leq x}\frac{\chi(n)\log(n)}{n} &= L\sum_{d\leq x}\frac{\chi(d)\Lambda(d)}{d} + O\left(\frac{1}{x}\sum_{d\leq x}\Lambda(d)\right)\\ &= L\sum_{d\leq x}\frac{\chi(d)\Lambda(d)}{d} + O(1),\end{aligned}$$

since $\sum_{d \le x} \Lambda(d) = \psi(x) \ll x$. But we also have $\sum \chi(n) \log n/n = O(1)$, since

$$\frac{\log 1}{1} - \frac{\log 3}{3} + \frac{\log 5}{5} - \dots$$

is, with the first few terms omitted, an alternating series with decreasing terms. Thus

$$L \sum_{d \le x} \frac{\chi(d)\Lambda(d)}{d} = O(1),$$

and since $L \neq 0$, it follows that

$$\sum_{d \le x} \frac{\chi(d)\Lambda(d)}{d} = O(1). \tag{2.8}$$

From (2.6) and (2.8), we deduce that

$$\sum_{\substack{n \le x \\ n \equiv 1 \mod 4}} \frac{\Lambda(n)}{n} + \sum_{\substack{n \le x \\ n \equiv 3 \mod 4}} \frac{\Lambda(n)}{n} = \log x + O(1),$$

$$\sum_{\substack{n \le x \\ n \equiv 1 \mod 4}} \frac{\Lambda(n)}{n} - \sum_{\substack{n \le x \\ n \equiv 3 \mod 4}} \frac{\Lambda(n)}{n} = O(1).$$

If we add these estimates, we see

$$\sum_{\substack{n \le x \\ n \equiv 1 \pmod 4}} \frac{\Lambda(n)}{n} = \frac{1}{2} \log x + O(1),$$

and if we subtract we get the same estimate for $n \equiv 3 \pmod 4$. Referring to equation (2.4) shows that the same estimates hold for the sums $\sum \log p/p$, finishing the proof in this case.

For a general progression $a \pmod q$, we have to consider $\phi(q) - 1$ series analogous to our L, corresponding to the $\phi(q) - 1$ nontrivial Dirichlet characters (whose role above was played by the single character χ). The difficult point turns out to be showing that $L \neq 0$, which in the case considered above was trivial.

Remark. For the remainder of this chapter, we adopt the convention that **all implied constants (unless otherwise stated) may depend on** q. Further dependence will be mentioned explicitly.

2.1.2 Exercises

Exercise 2.1.1 *(Sierpiński [Sie48]).* Using Dirichlet's theorem, prove that for each positive integer K, there exists a prime $p > K$ for which all the numbers $p \pm i$, $i = 1, 2, \dots, K$, are composite.

For another proof (of a somewhat stronger result) utilizing sieve methods instead of Dirichlet's theorem, see Exercise 3.4.2. Erdős & Rényi ([ER50]) showed that the smallest such prime does not exceed e^{cK} for a certain absolute constant c.

Exercise 2.1.2 *(Chang & Lih [CL77]).* Using Dirichlet's theorem on primes in arithmetic progressions, prove the following strengthening of Theorem 1.6.3: For each prime p, there is a polynomial f_p with integer coefficients producing distinct prime values for $0 \le n \le p-1$, with prescribed initial value $f_p(0) = p$.

Exercise 2.1.3 *(D.J. Newman).* Dov Jarden, in the book *Recurring Sequences* (1973), made the observation that $\phi(30n+1) > \phi(30n)$ for all $n \le 10{,}000$. D.J. Newman later verified this continues to hold for all $n \le 20{,}000{,}000$.

Using Dirichlet's theorem on primes in progressions, prove that contrary to what one might expect from the computational evidence, the reverse inequality

$$\phi(30n+1) < \phi(30n) \tag{2.9}$$

holds for infinitely many n. The smallest such n, which has over 1000 decimal digits, has been given explicitly by Martin [Mar99]. The result of this exercise is a special case of a theorem of Newman [New97]; a second proof of his general result, which avoids Dirichlet's theorem, can be found in [ABGU01].

Hint: Let $P_x := \prod_{5<p\le x} p$. For each x, choose a positive prime $n = n(x)$ such that $P_x \mid 30n+1$. Show (2.9) then holds for large enough x.

Exercise 2.1.4. This exercise illustrates the utility of (2.1) as an equidistribution statement. Define $n^\diamond$ as that portion of $n!$ composed of primes 3 (mod 4), i.e.,

$$n^\diamond := \prod_{\substack{p^k \| n! \\ p \equiv 3 \pmod 4}} p^k.$$

a) Using the results of this section on primes in progressions mod 4, show that
$$\log n^\diamond = \frac{1}{2} n \log n + O(n) \sim \frac{1}{2} \log n!.$$

b) Suppose that n and y are positive integers with $n! + 1 = y^8$. Using the factorization
$$n! = y^8 - 1 = \left((y^4+1)(y^2+1)\right)(y^2-1),$$
prove that $n^\diamond \le y^2 - 1 \le (n!)^{1/4}$. Deduce from part a) that the equation $n! + 1 = y^8$ has only finitely many solutions.

c) Assuming (2.1) in full generality, show by similar methods that the equation $n! + 1 = x^p$ has at most finitely many solutions for each fixed odd prime p. Note that in combination with part b), this shows $n! + 1 = x^m$ has only finitely many solutions for each positive integer $m > 1$ except possibly for $m = 2$ and $m = 4$.

It has been shown that $n! + 1 = y^m$ has *no* solutions for any $m > 2$ (see [EO37] for $m \ne 4$ and [PS73] for the case $m = 4$). When $m = 2$, there is an 1885 conjecture of Brocard that the only solutions correspond to $n = 4, 5$ and 7, but even proving the finiteness of their number remains open.

2.2 The Characters of a Finite Abelian Group

2.2.1 The Classification of Characters

Let G be a finite abelian group. The *characters* of G are the homomorphisms

$$\chi\colon G \to \mathbf{C}^*,$$

i.e., the functions from G to the nonzero complex numbers satisfying

$$\chi(ab) = \chi(a)\chi(b) \tag{2.10}$$

for every $a, b \in G$. The set of characters is denoted $\hat{G}$.

Let us attempt to describe the elements of $\hat{G}$. We always have the *trivial* or *principal* character that takes everything to 1, i.e.,

$$\chi_0(g) = 1 \quad \text{for every } g \in G.$$

The existence of nonprincipal characters is not at all clear this point.

One thing we see immediately is that if χ is a character of G, then every value χ assumes is a root of unity. Indeed, if the order of $g \in G$ is n, then

$$\chi(g)^n = \chi(g^n) = \chi(1) = 1.$$

Suppose that G is cyclic, and fix a generator g_0. Then knowing $\chi(g_0)$ determines $\chi(g)$ for every $g \in G$; indeed, if $g = g_0^k$ for some positive integer k, then

$$\chi(g) = \chi(g_0^k) = \chi(g_0)^k.$$

Since the values of $\chi(g_0)$ must all be $|G|$th roots of unity, and since the value of $\chi(g_0)$ determines the character, we see there are at most $|G|$ characters. There are exactly $|G|$ if and only if for every $|G|$th root of unity η there is a character with $\chi(g) = \eta$. And that is actually true: just define

$$\chi(g_0^k) = \eta^k.$$

We have to check that this definition makes sense. It certainly defines χ on all of G, since g_0 generates G, so that the only thing that needs checking now is that if $g = g_0^{k_1} = g_0^{k_2}$, then also $\eta^{k_1} = \eta^{k_2}$. But since g_0 has order $|G|$, the first condition forces $k_1 \equiv k_2 \pmod{|G|}$, which in turn forces $\eta^{k_1} = \eta^{k_2}$. So we've found all the characters of G.

A similar approach works if G is a direct sum of two cyclic groups, say $G \cong \mathbf{Z}/m\mathbf{Z} \oplus \mathbf{Z}/n\mathbf{Z}$. It suffices to classify the characters of $\mathbf{Z}/m\mathbf{Z} \oplus \mathbf{Z}/n\mathbf{Z}$, and for this we can proceed much as before: if we know where $(1,0)$ and $(0,1)$ are sent by a character χ, we know where everything is sent. Since $(1,0)$ has order m and $(0,1)$ has order n, we must have $\chi((1,0))^m = 1$ and $\chi((0,1))^n = 1$. If we now arbitrarily prescribe the values of $\eta_1 = \chi((1,0))$ and $\eta_2 = \chi((0,1))$ subject to these conditions, then

$$\chi((c,d)) = \eta_1^c \eta_2^d$$

is seen to be a well-defined character by essentially the same argument as we gave before. So we've found all the characters in this case as well. Since there are m choices for η_1 and n choices for η_2, we see G has exactly mn characters.

The same argument generalizes, of course, to give us a complete description of the characters of G when

$$G \cong \mathbf{Z}/m_1\mathbf{Z} \oplus \mathbf{Z}/m_2\mathbf{Z} \oplus \cdots \oplus \mathbf{Z}/m_k\mathbf{Z}$$

is any finite direct sum of cyclic groups. In particular, we see there are

$$m_1 m_2 \dots m_k = |G|$$

characters. And now we must confess that we have actually reached the end of the story because of the following classification theorem from algebra:

Theorem. *Every finite abelian group is a direct sum of cyclic groups.*

This theorem (and more) is proved as Theorem B.1.1.

Remark. For the purposes of this chapter, it is not necessary to invoke this difficult result. We only need to understand the characters of $G = \mathbf{Z}/m\mathbf{Z}^*$, the group of units $\pmod{m}$. In this case the existence of a decomposition into cyclic groups is entirely elementary. Indeed, the Chinese remainder theorem guarantees that if $m = \prod_{i=1}^{k} p_i^{e_i}$, then

$$\mathbf{Z}/m\mathbf{Z}^* \cong \mathbf{Z}/p_1^{e_1}\mathbf{Z}^* \oplus \cdots \oplus \mathbf{Z}/p_k^{e_k}\mathbf{Z}^*.$$

Noting that (see, e.g., [IR90, Theorems 2, 2'])

$$\mathbf{Z}/p^e\mathbf{Z}^* \cong \begin{cases} \mathbf{Z}/2\mathbf{Z} \oplus \mathbf{Z}/2^{e-2}\mathbf{Z} & \text{if } p = 2, e > 2, \\ \mathbf{Z}/(p-1)p^{e-1}\mathbf{Z} & \text{otherwise,} \end{cases}$$

we obtain the desired decomposition.

Unfortunately, this simple argument does not work in other cases which are of interest. For example, if $\pi(x)$ is a prime in $\mathbf{F}_q[x]$, then one can show that typically the units group of the quotient ring $\mathbf{F}_q[x]/(\pi(x)^k)$ is not cyclic; in fact, for fixed π the minimal number of elements required to generate this group tends to infinity with k (cf. [Ros02, Proposition 1.6]).

2.2.2 The Orthogonality Relations

The characters of an abelian group satisfy special relations known as *orthogonality relations*, which play a key role in the proof of Dirichlet's theorem. In the situation that concerns us, when $G = \mathbf{Z}/m\mathbf{Z}^*$, these relations allow us to express the characteristic function of a reduced residue class $\pmod{m}$ as a linear combination of characters.

Before stating these relations, we note that $\hat{G}$ can be made into a group (the so-called *dual group of* G) by defining, for characters $\chi, \psi \in \hat{G}$,

$$\chi\psi(g) = \chi(g)\psi(g),$$

i.e., by defining the multiplication pointwise. An identity for this operation is the trivial homomorphism χ_0. Associativity and commutativity follow from the corresponding properties of $\mathbf{C}^*$. Inverses are easy; just define the character χ^{-1} of G by

$$\chi^{-1}(g) = \chi(g)^{-1}.$$

This makes sense since χ takes values in the *nonzero* complex numbers, and the desired multiplicativity follows from just taking the reciprocal of (2.10). Notice that because the values χ assumes are always roots of unity,

$$\chi^{-1} = \overline{\chi},$$

where $\overline{\chi}$ is defined by

$$\overline{\chi}(g) := \overline{\chi(g)}.$$

Now suppose $\chi \in \hat{G}$ is a nonprincipal character, i.e., $\chi \neq \chi_0$. Then there exists $h \in G$ with $\chi(h) \neq 1$. Since G is a group, hg runs over the elements of G as g does. Thus, setting $S_\chi = \sum_{g \in G} \chi(g)$, one has

$$\chi(h) S_\chi = \chi(h) \sum_{g \in G} \chi(g) = \sum_{g \in G} \chi(hg) = \sum_{g \in G} \chi(g) = S_\chi.$$

Since $\chi(h) \neq 1$, this forces

$$S_\chi = \sum_{g \in G} \chi(g) = 0.$$

We have thus shown

$$\sum_{g \in G} \chi(g) = \begin{cases} |G| & \text{if } \chi = \chi_0, \\ 0 & \text{otherwise.} \end{cases} \tag{2.11}$$

Since $\overline{\chi} = \chi^{-1}$ for any character χ, this can be recast in the following form: if χ and ψ are two characters of G, then

$$\sum_{g \in G} \overline{\chi}(g) \psi(g) = \begin{cases} |G| & \text{if } \chi = \psi, \\ 0 & \text{otherwise.} \end{cases} \tag{2.12}$$

For the second orthogonality relation, instead of fixing $\chi \in \hat{G}$ and summing over all $g \in G$, we fix $g \in G$ and sum over all $\chi \in \hat{G}$. We need the following lemma:

Lemma 2.2.1. *Let G be a finite abelian group and let $g \neq 1$ be an element of G. Then there exists a character $\chi \in \hat{G}$ with $\chi(g) \neq 1$.*

Proof. Let

$$\theta \colon G \to \mathbf{Z}/m_1\mathbf{Z} \oplus \mathbf{Z}/m_2\mathbf{Z} \oplus \cdots \oplus \mathbf{Z}/m_k\mathbf{Z}$$

be an isomorphism of G with a direct sum of cyclic groups. It suffices to find a character of this direct sum not vanishing at the image of g. Since $g \neq 1$, not all the components of $\theta(g)$ can vanish; suppose the rth component $(1 \leq r \leq k)$ is nonzero. By our classification of characters of direct sums of cyclic groups, there exists a character χ of the direct sum with

$$\chi(0, 0, \ldots, 0, \underbrace{1}_{j\text{th entry}}, 0, \ldots, 0) = \begin{cases} e^{2\pi i/m_r} & \text{if } j = r, \\ 1 & \text{otherwise.} \end{cases}$$

Then $\chi(\theta(g)) \neq 1$. $\square$

Now we can proceed as before. Let $g \neq 1$ be an element of G and choose $\psi \in \hat{G}$ with $\psi(g) \neq 1$. Set $S_g = \sum_{\chi \in \hat{G}} \chi(g)$; then since $\hat{G}$ forms a group, $\psi\chi$ runs over all elements of $\hat{G}$ as χ does. Consequently,

$$\psi(g)S_g = \psi(g) \sum_{\chi \in \hat{G}} \chi(g) = \sum_{\chi \in \hat{G}} \psi\chi(g) = \sum_{\chi \in \hat{G}} \chi(g) = S_g.$$

Hence

$$\sum_{\chi \in \hat{G}} \chi(g) = \begin{cases} |G| & \text{if } g = 1 \\ 0 & \text{otherwise.} \end{cases}$$

If we note that for any $g \in G$,

$$\chi(g^{-1}) = \chi(g)^{-1} = \overline{\chi(g)} = \overline{\chi}(g),$$

we can recast this in the following form:

$$\sum_{\chi \in \hat{G}} \overline{\chi}(g)\chi(h) = \begin{cases} 1 & \text{if } g = h, \\ 0 & \text{otherwise.} \end{cases} \tag{2.13}$$

2.2.3 Dirichlet Characters

Let q be a positive integer and set $G = \mathbf{Z}/q\mathbf{Z}^*$, the group of units $\pmod q$. If $\chi \in \hat{G}$, there is an associated function $\tilde{\chi}$ defined on defined on the integers coprime to q given by setting

$$\tilde{\chi}(a) = \chi(a \pmod q).$$

In order to obtain a function defined on all of the integers, we define $\tilde{\chi}(a) = 0$ when $\gcd(a, q) > 1$.

The functions $\tilde{\chi}$ are called *Dirichlet characters* $\pmod q$. Instead of $\tilde{\chi}$, in what follows we abuse notation and use the same symbol χ to denote both the function on G and the associated function on the integers.

The following properties can now be readily verified (exercise!):

i. χ is periodic $\pmod q$, i.e., $\chi(a + q) = \chi(a)$ for every $a \in Z$.

ii. χ is completely multiplicative, i.e., or every $a, b \in \mathbf{Z}$,

$$\chi(ab) = \chi(a)\chi(b)$$

The orthogonality relations translate into the following two theorems:

Theorem 2.2.2. *Let q be a positive integer and let χ, ψ be two Dirichlet characters* $\pmod q$.

$$\sum_{a \pmod q} \overline{\chi}(a)\psi(a) = \begin{cases} \phi(q) & \text{if } \chi = \psi^{-1}, \\ 0 & \text{otherwise.} \end{cases} . \tag{2.14}$$

Theorem 2.2.3. *Let q be a positive integer. If a, b are integers with* $\gcd(a, q) = 1$, *then*

$$\sum_{\chi} \overline{\chi}(a)\chi(b) = \begin{cases} \phi(q) & \text{if } a \equiv b \pmod q, \\ 0 & \text{otherwise.} \end{cases} \tag{2.15}$$

These theorems are proved by applying (2.12) and (2.13), respectively, to the group $\mathbf{Z}/q\mathbf{Z}^*$ of size $\phi(q)$. Indeed, the first theorem is immediate from (2.12), since the contribution to the sum from those a with $\gcd(a, q) > 1$ is 0.

To prove the second theorem, we note the stated equality follows from (2.13) if $\gcd(a, q) = \gcd(b, q) = 1$. If, however, $\gcd(b, q) > 1$ then the sum vanishes (since each term is 0 owing to the presence of $\chi(b)$); since $\gcd(b, q) > 1$ implies $a \not\equiv b \pmod q$, the theorem holds in this case as well. (This is where we need the condition that $\gcd(a, q) = 1$.)

2.2.4 Exercises

Exercise 2.2.1. Fill in the following tables of group characters for $\mathbf{Z}/5\mathbf{Z}^*$ and $\mathbf{Z}/8\mathbf{Z}^*$:

	1	2	3	4
χ_0	1	1	1	1
χ_1	1			
χ_2	1			
χ_3	1			

	1	3	5	7
χ_0	1	1	1	1
χ_1	1			
χ_2	1			
χ_3	1			

Exercise 2.2.2 *(A. Granville).* Define the Carmichael λ-function by setting $\lambda(p^k) = \phi(p^k)$ for p an odd prime, setting

$$\lambda(2^k) = \begin{cases} \phi(2^k) & \text{if } k \le 2, \\ \frac{1}{2}\phi(2^k) & \text{otherwise,} \end{cases}$$

and in general if $n = \prod_{i=1}^{k} p_i^{e_i}$, by setting

$$\lambda(n) = \operatorname{lcm}[\lambda(p_1^{e_1}), \lambda(p_2^{e_2}), \ldots, \lambda(p_k^{e_k})].$$

a) Show that every value assumed by a character χ of $\mathbf{Z}/n\mathbf{Z}^*$ is a $\lambda(n)$th root of unity.

b) Prove that for some character χ, the image $\chi(\mathbf{Z}/n\mathbf{Z}^*)$ is the complete set of $\lambda(n)$th roots of unity.

Exercise 2.2.3. Use the explicit description of characters described in this section to show that $\hat{G} \cong G$. You may find it helpful to handle the case when G is cyclic first.

Exercise 2.2.4. Say that a Dirichlet character χ is *real* if χ assumes only real values. Show that χ is real if and only if $\chi^2 = \chi_0$. Using the result of the previous exercise, determine all moduli m for which all Dirichlet characters are real.

Exercise 2.2.5. Let χ be a function defined on the integers possessing the following three properties:

i. χ is periodic $\pmod q$.

ii. χ is completely multiplicative.

iii. $\chi(n) = 0$ if and only if $\gcd(n, q) > 1$.

Show that χ is a Dirichlet character $\pmod q$.

The next two exercises require some familiarity with linear algebra.

Exercise 2.2.6. Let G be a finite abelian group and let $\mathbf{C}[G]$ denote the space of functions $f\colon G \to \mathbf{C}$. For $\phi, \psi \in \mathbf{C}[G]$, define

$$(\phi, \psi) = \frac{1}{|G|} \sum_{g \in G} \phi(g)\overline{\psi(g)}.$$

Show that this is a scalar product on $\mathbf{C}[G]$. Using (2.12) show that the characters of G form an orthonormal basis for $\mathbf{C}[G]$. This explains the name "orthogonality relation."

Exercise 2.2.7. Suppose G is a finite abelian group of order n with elements $g_1, g_2, \ldots, g_n$ and characters $\chi_1, \chi_2, \ldots, \chi_n$. Define the matrix

$$M := \begin{pmatrix} \chi_1(g_1) & \chi_1(g_2) & \cdots & \chi_1(g_n) \\ \chi_2(g_1) & \chi_2(g_2) & \cdots & \chi_2(g_n) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_n(g_1) & \chi_n(g_2) & \cdots & \chi_n(g_n) \end{pmatrix}.$$

Let M^* denote the conjugate-transpose of M. Using (2.12), show that $MM^* = nI$, where I is the $n \times n$ identity matrix. Deduce (from linear algebra) that $M^*M = nI$ also, and show this implies (2.13). That is, the first orthogonality relation implies the second.

2.3 The L-series at $s = 1$

If χ is a Dirichlet character mod q, it is usual to associate with it the L-series

$$L(s, \chi) := \sum_{n=1}^{\infty} \frac{\chi(n)}{n^s}. \tag{2.16}$$

For our purposes, only the series corresponding to the nontrivial characters are of interest and these are only of interest at $s = 1$. Nevertheless, because there is no extra difficulty involved, we begin by treating the series corresponding to nontrivial Dirichlet characters whenever $s > 0$.

Lemma 2.3.1. *Let χ be a nontrivial Dirichlet character mod q. Then the series* (2.16) *converges for $s > 0$. Moreover, for every $x \geq 1$,*

$$\left| \sum_{n>x} \frac{\chi(n)}{n^s} \right| \leq 2qx^{-s}.$$

In particular,

$$\sum_{n>x} \frac{\chi(n)}{n} \ll \frac{1}{x}.$$

Proof. Define

$$S(x) := \sum_{n \leq x} \chi(n). \tag{2.17}$$

Theorem 2.2.2 implies that $\sum \chi(n) = 0$ when taken over any block of q consecutive integers. This implies $|S(x)| \leq q$ for every x. Now we apply partial summation:

$$\sum_{n \leq x} \frac{\chi(n)}{n^s} = \int_{1/2}^{x} \frac{dS(t)}{t^s} = \frac{S(x)}{x^s} + \int_1^x s \frac{S(t)}{t^{s+1}}\, dt.$$

As $x \to \infty$, the first factor on the right tends to 0 since $S(x)$ remains bounded while x^s tends to infinity. The last factor converges as $x \to \infty$, by comparison with the absolutely convergent integral

$$\int_1^{\infty} s \frac{q}{t^{s+1}}\, dt = q.$$

This proves the first claim.

Bounding the tail end is another similar application of partial summation:

$$\begin{aligned}\sum_{n>x} \frac{\chi(n)}{n^s} &= \int_x^{\infty} \frac{dS(t)}{t^s} \\ &= \frac{S(y)}{y^s} - \frac{S(x)}{x^s} + \int_x^y s \frac{S(t)}{t^{s+1}}\, dt \Bigg]_{y=\infty} = -\frac{S(x)}{x^s} + \int_x^{\infty} s \frac{S(t)}{t^{s+1}}\, dt.\end{aligned}$$

The first term is bounded in absolute value by qx^{-s} and the second by

$$\int_x^{\infty} s \frac{q}{t^{s+1}}\, dt = qx^{-s},$$

so that the stated estimate follows by the triangle inequality. $\square$

2.4 The Nonvanishing of $L(1,\chi)$ for complex χ

We say that a Dirichlet character χ is *real* if it assumes only real values; otherwise, we call it a *complex* character.

We first prove a lemma showing how the behavior of the sum $\sum \chi(n)\Lambda(n)/n$ is dependent on the vanishing or nonvanishing of $L(1,\chi)$.

Lemma 2.4.1. *Let χ be any nontrivial Dirichlet character mod q. Then for $x \geq 4$,*

$$\sum_{n\leq x} \frac{\chi(n)\Lambda(n)}{n} = \begin{cases} O(1) & \text{if } L(1,\chi)\neq 0 \\ -\log x + O(1) & \text{otherwise.} \end{cases}$$

It is convenient to prove this lemma in two parts:

Proof when $L(1,\chi)\neq 0$. The strategy of the proof is exactly the same as for the special case given in the introduction; only the justifications are somewhat different. We start with reexpressing

$$\begin{aligned} \sum_{n\leq x} \frac{\chi(n)\log(n)}{n} &= \sum_{n\leq x} \frac{\chi(n)}{n} \sum_{d\mid n} \Lambda(d) \\ &= \sum_{de\leq x} \frac{\chi(de)\Lambda(d)}{de} \\ &= \sum_{d\leq x} \frac{\chi(d)\Lambda(d)}{d} \sum_{e\leq x/d} \frac{\chi(e)}{e}. \end{aligned}$$

The inner sum here is equal to $L(1,\chi) - \sum_{e>x/d} \chi(e)/e = L(1,\chi) + O(d/x)$, using the result of Lemma 2.3.1. Inserting this estimate in the above shows

$$\begin{aligned} \sum_{n\leq x} \frac{\chi(n)\log(n)}{n} &= L(1,\chi)\sum_{d\leq x} \frac{\chi(d)\Lambda(d)}{d} + O\left(\frac{1}{x}\sum_{d\leq x}\Lambda(d)\right) \\ &= L(1,\chi)\sum_{d\leq x} \frac{\chi(d)\Lambda(d)}{d} + O(1), \end{aligned} \tag{2.18}$$

since $\sum_{d\leq x}\Lambda(d) = \psi(x) \ll x$ by (1.31).

But we also have

$$\sum_{n\leq x} \frac{\chi(n)\log n}{n} = O(1). \tag{2.19}$$

Indeed, with S defined as in (2.17),

$$\begin{aligned} \sum_{n\leq x} \frac{\chi(n)\log n}{n} &= \int_{1/2}^{x} \frac{dS(t)\log t}{t} \\ &= \frac{S(x)\log x}{x} - \int_{1}^{x} S(t)\frac{1-\log t}{t^2}\,dt, \end{aligned}$$

so that (noting that $\log t/t$ is decreasing for $t \geq e$)

$$\left|\sum_{n\leq x} \frac{\chi(n)\log(n)}{n}\right| \leq q\frac{\log 4}{4} + q\int_1^\infty \frac{dt}{t^2} + q\int_1^\infty \frac{\log t}{t^2}\,dt \ll 1.$$

Together, (2.18) and (2.19) imply

$$L(1,\chi)\sum_{d\leq x} \frac{\chi(d)\Lambda(d)}{d} = O(1).$$

Since $L(1,\chi) \neq 0$, it follows that

$$\sum_{d\leq x} \frac{\chi(d)\Lambda(d)}{d} = O(1),$$

which is the statement of the lemma in this case. (The implied constant here depends on the value of $L(1,\chi)$, but as there are only finitely many Dirichlet characters mod q, it may be chosen to depend only on q.) $\square$

Proof when $L(1,\chi) = 0$. Applying Möbius inversion to the relation

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

we obtain

$$\begin{aligned}\Lambda(n) = \sum_{d|n} \mu(d)\log\frac{n}{d} &= \sum_{d|n}\mu(d)\log n - \sum_{d|n}\mu(d)\log d\\ &= \log n\sum_{d|n}\mu(d) - \sum_{d|n}\mu(d)\log d\\ &= -\sum_{d|n}\mu(d)\log d,\end{aligned}$$

since for every positive integer n, either $\log n$ or $\sum_{d|n}\mu(d)$ vanishes. Thus for positive x, one has

$$\begin{aligned}\sum_{d|n}\mu(d)\log\frac{x}{d} &= \log x\sum_{d|n}\mu(d) + \Lambda(n)\\ &= \begin{cases}\log x + \Lambda(n) & \text{if } n = 1,\\ \Lambda(n) & \text{otherwise.}\end{cases}\end{aligned}$$

It follows that

$$\begin{aligned}
\log x + \sum_{n\le x}\frac{\chi(n)\Lambda(n)}{n} &= \sum_{n\le x}\frac{\chi(n)}{n}\sum_{d|n}\mu(d)\log\frac{x}{d}\\
&= \sum_{d\le x}\mu(d)\log\frac{x}{d}\sum_{\substack{n\le x\\ d|n}}\frac{\chi(n)}{n}\\
&= \sum_{de\le x}\mu(d)\log\frac{x}{d}\frac{\chi(de)}{de}\\
&= \sum_{d\le x}\mu(d)\log\frac{x}{d}\frac{\chi(d)}{d}\sum_{e\le x/d}\frac{\chi(e)}{e}\\
&= L(1,\chi)\sum_{d\le x}\mu(d)\log\frac{x}{d}\frac{\chi(d)}{d} + R(x), \qquad (2.20)
\end{aligned}$$

where (using the estimate of Lemma 2.3.1)

$$\begin{aligned}
R(x) &\ll \sum_{d\le x}\left(\log\frac{x}{d}\right)\frac{1}{d}\frac{d}{x}\\
&\ll \frac{1}{x}\sum_{d\le x}(\log x - \log d)\\
&= \frac{1}{x}\left([x]\log x - \log [x]!\right)\\
&= \frac{1}{x}\left(x\log x + O(\log x) - (x\log x - x + O(\log x))\right) \ll 1.
\end{aligned}$$

Since also $L(1,\chi)=0$ in this case, (2.20) implies

$$\log x + \sum_{n\le x}\frac{\chi(n)\Lambda(n)}{n} = O(1),$$

which proves the claim. □

We also require an estimate for the sum $\sum_{n\le x}\chi(n)\Lambda(n)/n$ corresponding to the principal character χ.

Lemma 2.4.2. *Let χ_0 be the principal character mod q. Then for $x\ge 4$,*

$$\sum_{n\le x}\frac{\chi_0(n)\Lambda(n)}{n} = \log x + O(1).$$

Proof. Observe that

$$\sum_{n\le x}\frac{\Lambda(n)}{n} - \sum_{n\le x}\frac{\chi_0(n)\Lambda(n)}{n} = \sum_{p|q}\sum_{p^k\le x}\frac{\log p}{p^k} \ll \sum_{p|q}\frac{\log p}{p-1} \ll 1.$$

The result now follows from (1.35). □

We can now prove the main result of this section.

Theorem 2.4.3. *Let χ be a complex character mod q. Then $L(1,\chi) \neq 0$.*

Proof. The results of Lemmas 2.4.1 and 2.4.2 together imply that for $x \geq 4$,

$$\sum_{\chi} \sum_{n \leq x} \frac{\chi(n)\Lambda(n)}{n} = (1-V)\log x + O(1), \tag{2.21}$$

where V denotes the number of nonprincipal χ with $L(1,\chi) = 0$ and the sum is taken over all Dirichlet characters mod q. On the other hand, taking $a = 1$ in the orthogonality relation (2.15) shows that

$$\sum_{\chi} \sum_{n \leq x} \frac{\chi(n)\Lambda(n)}{n} = \phi(q) \sum_{\substack{n \leq x \\ n \equiv 1 \pmod q}} \frac{\Lambda(n)}{n} \geq 0. \tag{2.22}$$

If $V > 1$, then (2.21) and (2.22) contradict each other for large enough x. Thus $V \leq 1$, i.e., $L(1,\chi)$ is nonzero for at most one nonprincipal character χ.

But if $L(1,\chi) = 0$ for some complex character χ, then

$$0 = L(1,\overline{\chi}) = \overline{\sum_{n=1}^{\infty} \frac{\chi(n)}{n}} = \sum_{n=1}^{\infty} \frac{\overline{\chi}(n)}{n} = L(1,\overline{\chi})$$

also. Since χ is complex, $\chi \neq \overline{\chi}$, so that $V \geq 2$, a contradiction. □

2.5 The Nonvanishing of $L(1,\chi)$ for real, non-principal χ

The most difficult step in the proof of Dirichlet's theorem is the nonvanishing of $L(1,\chi)$ for real nonprincipal χ. The proof we present in this section is due to Monsky [Mon93] and is a simplification of an earlier argument due to Gelfond & Linnik (e.g., see [GL66, Chapter 3]). For other elementary proofs of the non-vanishing of the L functions at $s = 1$, see [Nar00, §2.4].

Define, for $0 < x < 1$,

$$f(x) := \sum_{d=1}^{\infty} \sum_{k=1}^{\infty} \chi(d) x^{kd}. \tag{2.23}$$

We begin with the observation that for each $0 < x < 1$,

$$\sum_{d=1}^{\infty} \sum_{k=1}^{\infty} |\chi(d)| x^{kd}$$

converges. To see this, it suffices to show that

$$\sum_{d=1}^{\infty} \sum_{k=1}^{\infty} x^{kd} = \sum_{d=1}^{\infty} \frac{x^d}{1-x^d}$$

converges for $0 < x < 1$. But for every $0 < x < 1$, we have $x^d/(1-x^d) \leq x^d/(1-x)$ for each $d \geq 1$, so the result follows by comparison with the convergent geometric series $\sum_{d=1}^{\infty} x^d$.

These remarks imply [Tit39, §1.6] that the series defining f converges for $0 < x < 1$, and that the terms may be rearranged in any order without affecting the convergence or the value of the sum.

The inner sum in the original definition of f is a geometric series and

$$f(x) = \sum_{d=1}^{\infty} \chi(d) \frac{x^d}{1-x^d}. \tag{2.24}$$

But we may also write

$$f(x) = \sum_{n=1}^{\infty} \sum_{dk=n} \chi(d) x^n = \sum_{n=1}^{\infty} c_n x^n,$$

where

$$c_n = \sum_{d \mid n} \chi(d).$$

We claim that each c_n is nonnegative. To see this, note that since χ is multiplicative, c_n (considered as a function of n) is also multiplicative, whence

$$c_n = \prod_{p^e \| n} c_{p^e} = \prod_{p^e \| n} \left(1 + \chi(p) + \cdots + \chi(p^e)\right).$$

But since χ is real, one has either $\chi(p) = 0, 1$, or -1, so that

$$1 + \chi(p) + \cdots + \chi(p^e) = \begin{cases} 1 & \text{if } \chi(p) = 0, \\ e+1 & \text{if } \chi(p) = 1, \\ 0 & \text{if } \chi(p) = -1 \text{ and } 2 \nmid e, \\ 1 & \text{if } \chi(p) = -1 \text{ and } 2 \mid e. \end{cases}$$

Since the sum is nonnegative in each case, our claim follows.

Moreover, we see that $c_{p^e} \geq 1$ whenever e is even; it follows that $c_n \geq 1$ whenever n is a square. This implies that

$$f(x) = \sum c_n x^n \to \infty \quad (x \uparrow 1).$$

Indeed, let N be a positive integer, choose $M > 4N^2$ and define the polynomial $h(x)$ as the Mth partial sum of f, i.e., $h(x) := \sum_{n=1}^{M} c_n x^n$. Then

$$\lim_{x \uparrow 1} h(x) = h(1) \geq \sum_{\substack{n \leq M \\ n = \blacksquare}} 1 = \lfloor \sqrt{M} \rfloor \geq 2N.$$

By continuity, there is a half-open interval $[1-\epsilon, 1)$ (with $0 < \epsilon < 1$) such that $h(x) \geq N$ whenever $x \in [1-\epsilon, 1)$. Since $f(x) \geq h(x)$ for every $x \in (0,1)$, it follows that $f(x) \geq N$ whenever $x \in [1-\epsilon, 1)$. Since N can be chosen arbitrarily large, $\lim_{x\uparrow 1} f(x) = \infty$, as claimed.

Now suppose that $L(1,\chi) = 0$. Then for $0 < x < 1$,

$$\begin{aligned} -f(x) &= \frac{L(1,\chi)}{1-x} - f(x) \\ &= \sum_{n=1}^{\infty} \chi(n) \left(\frac{1}{n(1-x)} - \frac{x^n}{1-x^n} \right) = \sum_{n=1}^{\infty} \chi(n) b_n(x), \end{aligned} \tag{2.25}$$

say. We claim that

$$b_1(x) \geq b_2(x) \geq b_3(x) \geq \cdots \geq 0. \tag{2.26}$$

To prove this, note that

$$\begin{aligned} &(1-x)(b_n(x) - b_{n+1}(x)) \\ &\qquad = \frac{1}{n} - \frac{1}{n+1} - \frac{x^n}{1+x+\cdots+x^{n-1}} + \frac{x^{n+1}}{1+x+\cdots+x^n} \\ &\qquad = \frac{1}{n(n+1)} - \frac{x^n}{(1+x+\cdots+x^{n-1})(1+x+\cdots+x^n)}. \end{aligned} \tag{2.27}$$

The arithmetic-geometric mean inequality now implies both

$$\begin{aligned} 1+x+\cdots+x^{n-1} &\geq n x^{(n-1)/2} \geq n x^{n/2}, \\ 1+x+\cdots+x^n &\geq (n+1)x^{n/2}. \end{aligned}$$

Substituting into (2.27) shows that $(1-x)(b_n(x) - b_{n+1}(x)) \geq 0$; this proves the $b_n(x)$ are nonincreasing in n for fixed $x \in (0,1)$. Since $b_n(x) \to 0$ as $n \to \infty$, and $b_n(x)$ is nonincreasing in n, it follows that each $b_n(x)$ is nonnegative.

Define $S(x) = \sum_{n \leq x} \chi(n)$. Then (cf. (A.1), Appendix A)

$$\begin{aligned} \sum_{n=1}^{M} \chi(n) b_n(x) &= \sum_{n=1}^{M} (S(n) - S(n-1)) b_n(x) \\ &= S(M) b_M(x) + \sum_{n=1}^{M-1} S(n)(b_n(x) - b_{n+1}(x)). \end{aligned}$$

Using the inequality $|S(x)| \leq q$ and the nonnegativity of both $b_M(x)$ and $b_n(x) - b_{n+1}(x)$ (by (2.26)), we have by the triangle inequality

$$\begin{aligned} \left| \sum_{n=1}^{M} \chi(n) b_n(x) \right| &\leq q b_M(x) + q \sum_{n=1}^{M-1} (b_n(x) - b_{n+1}(x)) \\ &\leq q b_M(x) + q(b_1(x) - b_M(x)) = q b_1(x). \end{aligned}$$

But

$$b_1(x) = \frac{1}{1-x} - \frac{x}{1-x} = 1.$$

Consequently, $|S_M(x)| \leq q$ for every M; letting M tend to infinity, it follows from (2.25) that $|f(x)| \leq q$. This holds for for every $x \in (0,1)$, contradicting that $f(x) \to \infty$ as to $x \uparrow 1$.

2.6 Completion of the Proof

Let q be a positive integer and let a be any integer coprime to q.

The results of the last two sections imply that $L(1,\chi) \neq 0$ for every nontrivial character χ mod q. It follows from Lemma 2.4.1 that for every such χ, one has

$$\sum_{n \leq x} \frac{\chi(n)\Lambda(n)}{n} = O(1). \tag{2.28}$$

We record here also the result of Lemma 2.4.2 that

$$\sum_{n \leq x} \frac{\chi_0(n)\Lambda(n)}{n} = \log x + O(1). \tag{2.29}$$

Then from the orthogonality relation (2.15), we see

$$\begin{aligned} \sum_{\substack{n \leq x \\ n \equiv a \pmod q}} \frac{\Lambda(n)}{n} &= \frac{1}{\phi(q)} \sum_{\chi} \overline{\chi}(a) \sum_{n \leq x} \frac{\chi(n)\Lambda(n)}{n} \\ &= \frac{1}{\phi(q)} \overline{\chi_0}(a) \log x + O(1) = \frac{1}{\phi(q)} \log x + O(1), \end{aligned} \tag{2.30}$$

since $\chi_0(a) = 1$ (because $\gcd(a,q) = 1$).

But already in the introduction we showed (2.4)

$$\sum_{\substack{p \leq x \\ p \equiv a \pmod q}} \frac{\log p}{p} = \sum_{\substack{n \leq x \\ n \equiv a \pmod q}} \frac{\Lambda(n)}{n} + O(1), \tag{2.31}$$

with an absolute implied constant. Hence (2.30) implies

$$\sum_{\substack{p \leq x \\ p \equiv a \pmod q}} \frac{\log p}{p} = \frac{1}{\phi(q)} \log x + O(1).$$

2.6.1 Exercises

Exercise 2.6.1. Let $\mathcal{P}$ be a set of primes for which the estimate

$$\sum_{p \leq x, p \in \mathcal{P}} \frac{\log p}{p} = \kappa \log x + O(1)$$

holds for some constant $\kappa > 0$ and every $x \geq 2$.

a) Show that for some constant $D > 1$, there are $\gg x/\log x$ primes in the interval $(x, Dx]$ for every $x \geq 2$.

   Hint: Review the material of Chapter 1, §1.9.

b) Define

$$\pi_{\mathcal{P}}(x) := \#\{p \leq x : p \in \mathcal{P}\}.$$

   Using the result of a), show that

$$\pi_{\mathcal{P}}(x) \gg_{\mathcal{P}} \frac{x}{\log x}$$

   as $x \to \infty$.

c) Show that if

$$\lim_{x\to\infty} \frac{\pi_{\mathcal{P}}(x)}{x/\log x}$$

   exists, then it equals κ.

Exercise 2.6.2 *(Mertens [Mer97]).* Let a and q be integers with q positive and a coprime to q. Define

$$M(q) := \sum_{\chi\neq\chi_0} \frac{1}{|L(1,\chi)|},$$

the sum being over all nonprincipal characters $\pmod q$.

a) By making explicit the dependence on q in our proofs, show that for $x \geq 4$,

$$\phi(q) \sum_{\substack{p\leq x \\ p\equiv a \pmod q}} \frac{\log p}{p} =$$

$$\log x + O(1) + O(\phi(q)) + O\left(\sum_{p\mid q} \frac{\log p}{p-1}\right) + O(qM(q)),$$

   where the implied constants are absolute.

b) Let χ be a nontrivial character $\pmod q$. By splitting the sum defining $L(1,\chi)$ as $\sum_{n\leq q} + \sum_{n>q}$, prove that $L(1,\chi) \ll \log q$, with an absolute implied constant.

c) Prove that there is an absolute constant C with the property that for every $x \geq 4$, there is a prime $p \equiv a \pmod q$ in the interval $[x, x\exp(CqM(q))]$.

Exercise 2.6.3. Show that under the hypotheses of Exercise 2.6.1, there is a positive constant $C = C(\mathcal{P})$ such that

$$\prod_{p\leq x, p\in\mathcal{P}} \left(1 - \frac{1}{p}\right) = \frac{C}{\log^{\kappa} x}\left(1 + O\left(\frac{1}{\log x}\right)\right).$$

for $x \geq 2$. Here the implied constant may depend on $\mathcal{P}$.

Remark (on Exercise 2.6.3). When $\mathcal{P}$ is the set of primes $p \equiv a \pmod q$ (so that $\kappa = 1/\phi(q)$), K.S. Williams [Wil74] has given an explicit determination of the constant C, which we now describe. For each character $\chi \pmod q$, let k_χ be the completely multiplicative function satisfying

$$k_\chi(p) := p\left(1 - \left(1 - \frac{\chi(p)}{p}\right)\left(1 - \frac{1}{p}\right)^{-\chi(p)}\right).$$

It may be shown that $K(s,\chi) := \sum_{n=1}^{\infty} k_\chi(n)/n^s$ converges absolutely for every real $s > 0$, in particular at $s = 1$. The constant C above is given by

$$\left(e^{-\gamma}\frac{q}{\phi(q)}\prod_{\chi\neq\chi_0}\left(\frac{K(1,\chi)}{L(1,\chi)}\right)^{\overline{\chi}(a)}\right)^{1/\phi(q)}.$$

2.7 Sums of Three Squares

It is usual for elementary texts to characterize the set of integers expressible as a sum of two squares and to prove that every positive integer admits a representation as a sum of four squares.

However, representability by three squares usually receives short shrift in these texts. Most end their discussion with the following necessary condition for representability:

Theorem 2.7.1. *Suppose the positive integer n is a sum of three integer squares. Then n is not a power of 4 multiplied by a number $7 \pmod 8$.*

Proof. Introduce the function

$$r_3(m) := |\{(x_1,x_2,x_3) \in \mathbf{Z}^3 : x_1^2 + x_2^2 + x_3^2 = m\}|, \tag{2.32}$$

which counts the number of representations of m as a sum of three squares. Because every square is congruent to $0, 1$ or $4 \pmod 8$, we have $r_3(m) = 0$ whenever $m \equiv 7 \pmod 8$.

We next claim that $r_3(4m) = r_3(m)$ for every positive integer m. Indeed, given any representation of m as a sum of three integral squares, we can multiply the values of x_1, x_2 and x_3 by 2 to obtain a representation of $4m$. Different representations of m yield different representations of $4m$. To see every representation of $4m$ arises in this way, suppose $X_1^2 + X_2^2 + X_3^2 = 4m$. Looking mod 4, we see X_1, X_2 and X_3 are all even, so $X_1 = 2x_1, X_2 = 2x_2$ and $X_3 = 2x_3$ for some x_1, x_2 and x_3 whose squares sum to m.

Putting these two pieces together, we see $r_3(m) = 0$ whenever m has the shape $4^k(8n + 7)$. □

The goal of this section is to prove that the necessary condition of the previous theorem is also sufficient:

Theorem 2.7.2 (Legendre, 1798). *Every positive integer n not of the form $4^k(8m+7)$ is expressible as a sum of three squares.*

Before describing the proof, we mention two important corollaries. These are the first two cases of Cauchy's polygonal number theorem stating that every nonnegative integer is the sum of k k-gonal numbers (where $k \geq 3$):

Theorem 2.7.3 (Gauss). *Every nonnegative is a sum of three triangular numbers, i.e., three numbers of the form $(k^2+k)/2$, with k nonnegative.*

Proof. The equation

$$m = \frac{k_1^2+k_1}{2} + \frac{k_2^2+k_2}{2} + \frac{k_3^2+k_3}{2}$$

is equivalent to

$$8m+3 = (2k_1+1)^2 + (2k_2+1)^2 + (2k_3+1)^2.$$

Therefore, the assertion of the theorem is equivalent to the claim that every positive integer $\equiv 3 \pmod 8$ is a sum of three odd squares. Legendre's theorem implies all positive integers $\equiv 3 \pmod 8$ are sums of three squares; looking $\pmod 8$ we find all the squares are odd. The result follows. ☐

Theorem 2.7.4 (Lagrange). *Every nonnegative integer is a sum of four squares.*

Proof. By Legendre's theorem it suffices to establish this for m of the form $4^k(8n+7)$. By the same theorem, $m - 4^k = 4^k(8n+6)$ is a sum of three squares, say $x_1^2+x_2^2+x_3^2$. Therefore $m = x_1^2+x_2^2+x_3^2+(2^k)^2$ is a sum of four squares. ☐

This proof of Theorem 2.7.2 is long. The next four subsections are devoted to outlining that portion of the general theory of quadratic forms needed for the proof, which is given in the succeeding two subsections. The last subsection discusses the *number* of representations of an integer as a sum of three squares.

2.7.1 Quadratic Forms

For the purposes of this section, an (integral) *n-ary quadratic form* is a polynomial f in $x_1,\dots,x_n$ which can be written (necessarily uniquely) in the shape

$$f(x_1,\dots,x_n) = \sum_{1\leq i,j\leq n} a_{ij}x_ix_j,^{1} \tag{2.33}$$

where the a_{ij} are integers and $a_{ij}=a_{ji}$ for each pair i,j. If we associate f with the symmetric matrix (a_{ij}), we obtain a bijection between the n-ary forms and the symmetric $n\times n$ integer matrices (a_{ij}).

[1]For reasons that will be clear later, we will sometimes write the arguments of f as a column vector of indeterminates.

Observe that the binary quadratic forms are the integer polynomials of the form $ax^2 + 2bxy + cy^2$, where a, b and c are integers. In general, an n-ary quadratic form is a degree 2 homogeneous polynomial in $\mathbf{Z}[x_1, \ldots, x_n]$ for which the coefficients of all the cross terms $x_i x_j$ are even.

2.7.2 Equivalent Forms

For each n, we define an action of $\mathrm{GL}(n, \mathbf{Z})$ (the space of invertible $n \times n$ matrices with integral entries) on $\mathbf{Z}[x_1, \ldots, x_n]$ by

$$A \cdot f \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = f\left(A \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}\right) \qquad (A \in \mathrm{GL}(n, \mathbf{Z}), f \in \mathbf{Z}[x_1, \ldots, x_n]).$$

This is a bona fide action: clearly $I_n \cdot f = f$ for each f, where I_n is the $n \times n$ identity matrix, and the reader can easily check that $A \cdot (B \cdot f) = (AB) \cdot f$ for each $A, B \in \mathrm{GL}(n, \mathbf{Z})$ and each polynomial f. We then get an equivalence relation on $\mathbf{Z}[x_1, \ldots, x_n]$ by saying f and g are equivalent (written $f \sim g$) if $g = A \cdot f$ for some invertible $n \times n$ matrix A.

The reason for introducing this notion is that equivalent polynomials represent the same integers. More precisely, if f and g are equivalent, with say $f = A \cdot g$, then the map $\mathbf{x} \to A\mathbf{x}$ is a bijection between $\{\mathbf{x} \in \mathbf{Z}^n : f(\mathbf{x}) = m\}$ and $\{\mathbf{x} \in \mathbf{Z}^n : g(\mathbf{x}) = m\}$, for each integer m.

Let us now check that our equivalence relation on polynomials descends to an equivalence relation on n-ary quadratic forms. It suffices to verify that if f is an n-ary quadratic form, then so is $A \cdot f$, for any $A \in \mathrm{GL}(n, \mathbf{Z})$. So suppose f is given by (2.33), and let M be the corresponding symmetrix matrix $M := (a_{ij})$. Observe that with

$$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix},$$

we have

$$f(\mathbf{x}) = \mathbf{x}^T M \mathbf{x},$$

where ^{T} denotes the transpose. Consequently, for $A \in \mathrm{GL}(n, \mathbf{Z})$,

$$(A \cdot f)(\mathbf{x}) = f(A\mathbf{x}) = \mathbf{x}^T A^T M A \mathbf{x}. \qquad (2.34)$$

The matrix $A^T M A$ is symmetric, so we conclude that $A \cdot f$ is again a quadratic form (with associated matrix $A^T M A$). We have thus shown that our equivalence relation does descend to an equivalence relation on forms; moreover, the forms corresponding to the symmetric matrices M_1 and M_2 are equivalent precisely when there is an invertible integer matrix A with

$$M_2 = A^T M A.$$

In this case the symmetric matrices M_1 and M_2 are said to be *congruent* (written $M_1 \sim M_2$).

Example. Let $n = 2$, and let $f(x_1, x_2) = x_1^2 + x_2^2$ and $g(x_1, x_2) = x_1^2 + 4x_1x_2 + 5x_2^2$. Then $f \sim g$, since $g(x_1, x_2) = f(x_1 + 2x_2, x_2)$ while $\left(\begin{smallmatrix} 1 & 2 \\ 0 & 1 \end{smallmatrix}\right)$ is invertible. Note that

$$\begin{pmatrix} 1 & 2 \\ 2 & 5 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}.$$

Now define the *determinant of a form* f as the determinant of the associated matrix $M = (a_{ij})$. Since, for $A \in \mathrm{GL}(n, \mathbf{Z})$,

$$\begin{aligned} \det(A^T MA) &= \det(A^T) \det(M) \det(A) \\ &= \det(A)^2 \det(M) = (\pm 1)^2 \det(M) = \det(M), \end{aligned}$$

the determinant depends only on the equivalence class of f.

The next two subsections are occupied with the proof of the following important result concerning this invariant:

Theorem 2.7.5. *Let f be a ternary quadratic form of determinant 1, and suppose $f(x_1, x_2, x_3) \geq 0$ for all integral x_1, x_2, x_3 with equality only when $x_1 = x_2 = x_3 = 0$. Moreover, suppose f represents 1 for integral values of x_1, x_2 and x_3. Then f is equivalent to $x^2 + y^2 + z^2$.*

Remark. The hypothesis that f represent 1 is redundant. We will verify this hypothesis directly in the relevant special case later; for the general fact, see the result of Exercise 2.7.2.

2.7.3 Bilinear Forms on $\mathbf{Z}^n$

A map $l \colon \mathbf{Z}^n \to \mathbf{Z}$ is said to be *linear* if

$$l(m\mathbf{v} + n\mathbf{w}) = ml(\mathbf{v}) + nl(\mathbf{w})$$

for every $\mathbf{v}, \mathbf{w} \in \mathbf{Z}^n$ and every $m, n \in \mathbf{Z}$. A map $b \colon \mathbf{Z}^n \times \mathbf{Z}^n \to \mathbf{Z}$ is said to be *bilinear* if $b(\mathbf{v}, \mathbf{w})$ is linear as a function of $\mathbf{w}$ for each fixed $\mathbf{v}$ and linear as a function of $\mathbf{v}$ for each fixed $\mathbf{w}$.

Now suppose we are given a basis $\mathcal{B}$ of $\mathbf{Z}^n$, say $\mathbf{e_1}, \ldots, \mathbf{e_n}$, and a bilinear form $b : \mathbf{Z}^n \times \mathbf{Z}^n \to \mathbf{Z}$. (By Lemma B.2.1, every basis of $\mathbf{Z}^n$ has n elements.) A generic $u \in \mathbf{Z}^n$ may be written $\mathbf{u} = \sum c_i \mathbf{e_i}$, where the coefficients c_i are integers; we set

$$\mathbf{u}_{\mathcal{B}} := \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix} \in \mathbf{Z}^n.$$

Then if $\mathbf{v}, \mathbf{w} \in \mathbf{Z}^n$, with

$$\mathbf{v}_{\mathcal{B}} = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}, \quad \mathbf{w}_{\mathcal{B}} = \begin{pmatrix} d_1 \\ \vdots \\ d_n \end{pmatrix},$$

we have

$$b(\mathbf{v}, \mathbf{w}) = b(\sum c_i \mathbf{e_i}, \sum c_j \mathbf{e_j}) = \sum_{1 \leq i,j \leq n} c_i c_j b(\mathbf{e_i}, \mathbf{e_j}).$$

If we introduce the matrix

$$M := \begin{pmatrix} b(\mathbf{e_1}, \mathbf{e_1}) & b(\mathbf{e_1}, \mathbf{e_2}) & \dots & b(\mathbf{e_1}, \mathbf{e_n}) \\ b(\mathbf{e_2}, \mathbf{e_1}) & b(\mathbf{e_2}, \mathbf{e_2}) & \dots & b(\mathbf{e_2}, \mathbf{e_n}) \\ \vdots & \vdots & \ddots & \vdots \\ b(\mathbf{e_n}, \mathbf{e_1}) & b(\mathbf{e_n}, \mathbf{e_2}) & \dots & b(\mathbf{e_n}, \mathbf{e_n}) \end{pmatrix} \tag{2.35}$$

we therefore have

$$b(\mathbf{v}, \mathbf{w}) = \mathbf{v}_{\mathcal{B}}^T M \mathbf{w}_{\mathcal{B}}. \tag{2.36}$$

In fact, M is the *unique* matrix for which (2.36) holds. For suppose (2.36) holds for M. Let $\mathbf{v} = \mathbf{e_i}$ and $\mathbf{w} = \mathbf{e_j}$, so that $\mathbf{v}_{\mathcal{B}}$ and $\mathbf{w}_{\mathcal{B}}$ are the ith and jth standard basis vectors, respectively. Then (2.36) shows the ith row, jth column entry of M is be $b(\mathbf{e_i}, \mathbf{e_j})$. We call M *the matrix of b with respect to the basis $\mathcal{B}$.*

Now fix a bilinear form $b\colon \mathbf{Z}^n \to \mathbf{Z}$. Let us examine the question of how the matrix of b changes upon a change of basis. So suppose $\mathbf{e_1}, \dots, \mathbf{e_n}$ and $\mathbf{e_1}', \dots, \mathbf{e_n}'$ are two bases $\mathcal{B}$ and $\mathcal{B}'$, respectively, and let $M_{\mathcal{B}}$ and $M_{\mathcal{B}'}$ denote respectively the matrices of b with respect to $\mathcal{B}$ and $\mathcal{B}'$. Write

$$\mathbf{e_j}' = \sum_{i=1}^{n} a_{ij} \mathbf{e_i},$$

where the a_{ij} are integers. Then $A := (a_{ij}) \in \mathrm{GL}(n, \mathbf{Z})$. Also, if if $\mathbf{v}, \mathbf{w} \in \mathbf{Z}^n$ with $\mathbf{v}_{\mathcal{B}'} = (c_1, \dots, c_n)^T$ and $\mathbf{w}_{\mathcal{B}'} = (d_1, \dots, d_n)^T$, then $\mathbf{v}_{\mathcal{B}} = A\mathbf{v}_{\mathcal{B}'}$ and similarly for $\mathbf{w}_{\mathcal{B}}$. Therefore,

$$b(\mathbf{v}, \mathbf{w}) = \mathbf{v}_{\mathcal{B}}^T M_{\mathcal{B}} \mathbf{w} = \mathbf{v}_{\mathcal{B}'}^T A^T M_{\mathcal{B}} A \mathbf{w}_{\mathcal{B}'}.$$

Since also

$$b(\mathbf{v}, \mathbf{w}) = \mathbf{v}_{\mathcal{B}'}^T M_{\mathcal{B}'} \mathbf{w}_{\mathcal{B}'},$$

the aforementioned uniqueness implies

$$M_{\mathcal{B}'} = A^T M_{\mathcal{B}} A.$$

Therefore, changing bases replaces the matrix of b with a congruent matrix.

Now suppose we begin with a symmetric integer matrix M. Then we may define a bilinear form on $\mathbf{Z}^n$ by setting

$$b(\mathbf{v}, \mathbf{w}) = \mathbf{v}^T M \mathbf{w};$$

this bilinear form has matrix M with respect to the standard basis for $\mathbf{Z}^n$. The form b defined in this way is not only bilinear, but is also *symmetric*: for each pair $\mathbf{v}, \mathbf{w} \in \mathbf{Z}^n$,

$$b(\mathbf{v}, \mathbf{w}) = b(\mathbf{w}, \mathbf{v}).$$

The simplest way of seeing this is to observe that as $b(\mathbf{v}, \mathbf{w})$ is a scalar,

$$\begin{aligned} b(\mathbf{v}, \mathbf{w}) = b(\mathbf{v}, \mathbf{w})^T &= (\mathbf{v}^T M \mathbf{w})^T \\ &= \mathbf{w}^T M^T \mathbf{v} = \mathbf{w}^T M \mathbf{v} = b(\mathbf{w}, \mathbf{v}). \end{aligned}$$

Conversely, if b is a symmetric bilinear form, then its matrix with respect to any basis is clearly symmetric, as is evident from (2.35).

Let us briefly summarize how the results of this discussion will be used in the sequel. Let f be a quadratic form associated to the symmetric matrix M. Let b be the bilinear form defined above. Then

$$b(\mathbf{v}, \mathbf{v}) = \mathbf{v}^T M \mathbf{v} = f(\mathbf{v})$$

for every $\mathbf{v} \in \mathbf{Z}^n$, where f is the quadratic form associated to the matrix A. The results of this section imply that to understand the equivalence class of f, we can we can study the symmetric bilinear form b; if M' is any matrix of this bilinear form with respect to a change of basis, then M' is congruent to M, so our earlier results imply the form corresponding to M' is equivalent to f.

We end with a useful technical lemma:

Lemma 2.7.6. *Let $b : \mathbf{Z}^n \times \mathbf{Z}^n \to \mathbf{Z}$ be a bilinear form on $\mathbf{Z}^n$, and suppose $b(\mathbf{e_1}, \mathbf{e_1}) = 1$ for a certain $e_1 \in \mathbf{Z}^n$. Then we can extend $\mathbf{e_1}$ to a basis $\mathbf{e_1}, \mathbf{e_2}, \ldots, \mathbf{e_n}$ for $\mathbf{Z}^n$ with the additional property that*

$$b(\mathbf{e_1}, \mathbf{e_j}) = 0 \quad (j = 2, 3, \ldots, n).$$

Proof. Let $H := \{\mathbf{v} \in \mathbf{Z}^n : b(\mathbf{e_1}, \mathbf{v}) = 0\}$. We claim

$$\mathbf{Z}^n = \mathbf{Z}\mathbf{e_1} \oplus H.$$

Given $\mathbf{y} \in \mathbf{Z}^n$, write

$$\mathbf{y} = b(\mathbf{e_1}, \mathbf{y})\mathbf{e_1} + (\mathbf{y} - b(\mathbf{e_1}, \mathbf{y})\mathbf{e_1}).$$

The first term lies in $\mathbf{Z}\mathbf{e_1}$, and one checks easily that the second lies in H. This proves $\mathbf{Z}^n = \mathbf{Z}\mathbf{e_1} + H$; to see the sum is direct, suppose $\mathbf{y} \in \mathbf{Z}\mathbf{e_1} \cap H$, say $\mathbf{y} = n\mathbf{e_1}$; then

$$0 = b(\mathbf{e_1}, \mathbf{y}) = b(\mathbf{e_1}, n\mathbf{e_1}) = nb(\mathbf{e_1}, \mathbf{e_1}) = n,$$

so $n = 0$ and $\mathbf{y} = 0$.

To complete the proof, choose a basis $e_2, \ldots, e_n$ for H; then $\mathbf{e_1}, \ldots, \mathbf{e_n}$ is a basis for $\mathbf{Z}^n$, and the additional property is guaranteed by the definition of H. (The existence of a basis for H comes from Theorem B.2.2; that any such basis has $n-1$ elements follows from the fact that $\mathbf{Z}^n$ has rank n; cf. Lemma B.2.1.) □

2.7.4 Forms of Determinant 1

We now complete the proof of Theorem 2.7.5. It is helpful to begin by establishing the analogous result for *binary* quadratic forms. We make use of the following fundamental result from the geometry of numbers:

Theorem 2.7.7 (Minkowski). *Let X be region in $\mathbf{R}^n$ which is convex and symmetric about the origin. Let $\mathcal{L}$ be a complete lattice in $\mathbf{R}^n$; that is, let there be n linearly independent vectors $\mathbf{v_1}, \ldots, \mathbf{v_n} \in \mathbf{R}^n$ with*

$$\mathcal{L} = \mathbf{Z}\mathbf{v_1} + \cdots + \mathbf{Z}\mathbf{v_n}.$$

Let V be the volume of the "fundamental parallelepiped"

$$P := \left\{ \sum_{i=1}^{n} a_i \mathbf{v_i} : 0 \le a_i \le 1 \text{ for each } i \right\}. \tag{2.37}$$

Then if $\mathrm{vol}(X) > 2^n V$, the set X contains a point of the lattice $\mathcal{L}$ other than the origin. (Here $\mathrm{vol}(X)$ denotes the Lebesgue measure of X.)

Proof. Let $\mathbf{e_1}, \ldots, \mathbf{e_n}$ be the standard basis vectors, and let T be the (invertible) linear transformation with $T\mathbf{v_i} = \mathbf{e_i}$ for each i. Then T restricts to a bijection between $\mathcal{L}$ and $\mathbf{Z}^n$, and takes the parallelepiped (2.37) of volume V to the unit box

$$\{(x_1, \ldots, x_n) \in \mathbf{R}^n : 0 \le x_i \le 1 \text{ for each } i\}$$

of volume 1. A linear transformation T always shrinks volumes by a fixed ratio, so it must shrink all volumes by the factor $1/V$. In particular, X goes to a set $T(X)$ of volume $\mathrm{vol}(X)/V > 2^n$. One checks without difficulty that $T(X)$ is also convex and centrally symmetric.

We can therefore assume from the start that $\mathcal{L} = \mathbf{Z}^n$ is the standard lattice and that X is a convex, centrally symmetric set of volume exceeding 2^n. Under this hypothesis, consider the system of $2 \times 2 \times \cdots \times 2$ boxes of volume 2^n centered at the points $(a_1, \ldots, a_n)$ where all the a_i are even integers. Imagine shifting the portion of X from each box to the single box containing the origin. Since $\mathrm{vol}(X) > 2^n$ and since volume is both countably additive and translation-invariant, two of the translates must overlap. That is, there must exist $\mathbf{x_1}, \mathbf{x_2} \in X$ with

$$\mathbf{x_1} = \mathbf{x_2} + (2b_1, \ldots, 2b_n)$$

for some integers $b_1, \ldots b_n$, not all 0. Since X is centrally symmetric, $-\mathbf{x_2} \in X$; finally, since X is convex, it now follows that

$$\frac{1}{2}(\mathbf{x_1} + -\mathbf{x_2}) = \frac{1}{2}((\mathbf{x_2} + (2b_1, \ldots, b_n)) - \mathbf{x_2}) = (b_1, \ldots, b_n) \in X.$$

The proof is complete, as $(b_1, \ldots, b_n)$ is a nonzero point of $\mathbf{Z}^n$. □

Say that an integral n-ary quadratic form f is *positive definite* if $f(\mathbf{v}) \ge 0$ for every $\mathbf{v} \in \mathbf{Z}^n$, with equality only when $\mathbf{v} = 0$. Note that if f is a positive definite quadratic form, and $g \sim f$, then g is also positive definite: $g(\mathbf{v}) \le 0$ implies $f(A\mathbf{v}) \le 0$, so $A\mathbf{v} = 0$, and $\mathbf{v} = 0$.

Corollary 2.7.8. *Let $f(x, y) = ax_1^2 + 2bx_1y_1 + cy_1^2$ be a positive definite binary quadratic form of determinant $ac - b^2 = 1$. Then there are integers x, y with $f(x, y) = 1$.*

Proof. Observe that for (real) x, y, we have

$$af(x,y) = (ax+by)^2 + (ac-b^2)y^2 = (ax+by)^2 + y^2.$$

Since f is positive definite with $f(1,0) = a$, we must have $a > 0$; therefore, we can write

$$f(x,y) = \left(\sqrt{a}x + \frac{b}{\sqrt{a}}y\right)^2 + \left(\frac{1}{\sqrt{a}}y\right)^2. \tag{2.38}$$

It will therefore suffice to show the circle $u^2+v^2 < 2$ contains a nonzero point of the lattice $\mathcal{L}$ generated by $\mathbf{v_1} := (\sqrt{a}, b/\sqrt{a})$ and $\mathbf{v_2} := (0, 1/\sqrt{a})$. But the fundamental parallelepiped (2.37) has volume

$$\begin{vmatrix} \sqrt{a} & 0 \\ b/\sqrt{a} & 1/\sqrt{a} \end{vmatrix} = 1,$$

and the circle $u^2+v^2 < 2$ has area $2\pi > 4 \cdot 1$, so this is immediate from Minkowski's theorem. □

Theorem 2.7.9. *Let f be as in the above corollary. Then $f \sim x_1^2 + y_1^2$.*

Proof. Let b be the symmetric bilinear form on $\mathbf{Z}^2$ associated to f (see the last paragraph of the preceding subsection). By the corollary, there is an $\mathbf{e_1} \in \mathbf{Z}^2$ with $b(\mathbf{e_1}, \mathbf{e_1}) = 1$. By Lemma 2.7.6, we can extend this to a basis $\mathbf{e_1}, \mathbf{e_2}$ of $\mathbf{Z}^2$ with $b(\mathbf{e_1}, \mathbf{e_2}) = 0$. Then (as b is symmetric) the matrix of b with respect to $\mathbf{e_1}, \mathbf{e_2}$ looks like

$$\begin{pmatrix} b(\mathbf{e_1},\mathbf{e_1}) & b(\mathbf{e_1},\mathbf{e_2}) \\ b(\mathbf{e_2},\mathbf{e_1}) & b(\mathbf{e_2},\mathbf{e_2}) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & b(\mathbf{e_2},\mathbf{e_2}) \end{pmatrix}.$$

As a matrix congruent to the matrix of f, it must have the same determinant of f, whence $b(\mathbf{e_2}, \mathbf{e_2}) = 1$. Therefore, the identity matrix is congruent to the matrix of f, so the form associated to the identity matrix, namely $x_1^2 + y_1^2$, is equivalent to f. □

We can similarly prove Theorem 2.7.5, that a positive definite ternary form of discriminant 1 representing 1 is equivalent to $x_1^2 + x_2^2 + x_3^2$:

Proof of Theorem 2.7.5. Let b be the symmetric bilinear form on $\mathbf{Z}^3$ associated to f. By hypothesis, there is an $\mathbf{e_1} \in \mathbf{Z}^3$ with $b(\mathbf{e_1}, \mathbf{e_1}) = 1$. Extend $\mathbf{e_1}$ to a basis $\mathbf{e_1}, \mathbf{e_2}, \mathbf{e_3}$, as in Lemma 2.7.6. Then the matrix of b with respect to this basis has the shape

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & b(\mathbf{e_2},\mathbf{e_2}) & b(\mathbf{e_2},\mathbf{e_3}) \\ 0 & b(\mathbf{e_3},\mathbf{e_2}) & b(\mathbf{e_3},\mathbf{e_3}) \end{pmatrix}.$$

Because f has determinant 1, the bottom-right 2×2 matrix, say M, also has determinant 1.

Now M is a symmetric matrix, so we can associate with it the binary quadratic form

$$g(x_2, x_3) = \sum_{2 \le i,j \le 3} b(\mathbf{e_i}, \mathbf{e_j}) x_i x_j$$

of determinant 1. If g were not positive definite, then the ternary form

$$x_1^2 + g(x_2, x_3) = x_1^2 + \sum_{2 \le i,j \le 3} b(\mathbf{e_i}, \mathbf{e_j}) x_i x_j$$

would also fail to be positive definite. But as this latter form is equivalent to f, it must be positive definite. So g is positive-definite. Theorem 2.7.9 now implies M is congruent to the identity. Suppose $A^T M A$ is the 2×2 identity matrix, with $A \in \mathrm{GL}(2, \mathbf{Z})$; then

$$\begin{pmatrix} 1 & 0 \\ 0 & A \end{pmatrix}^T \begin{pmatrix} 1 & 0 \\ 0 & M \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & A \end{pmatrix} = I_3,$$

the 3×3 identity matrix. Since the matrix of f was congruent to $\left(\begin{smallmatrix} 1 & 0 \\ 0 & M \end{smallmatrix}\right)$, we now find the matrix of f is congruent to the 3×3 identity matrix. The result follows. $\square$

2.7.5 Proof of the Three Squares Theorem

We are now ready to prove Theorem 2.7.2.

Theorem 2.7.2 follows once it is known that every squarefree $m \not\equiv 7 \pmod 8$ is a sum of three squares. For let m be given, where m is not a power of 4 multiplied by a number 7 (mod 8). Write $m = m'r^2$, where m' is squarefree, and $r = 2^k s$ with s odd. Then were $m' \equiv 7 \pmod 8$, we would have

$$m = m'(2^k s)^2 = 4^k (m's^2),$$

with

$$m's^2 \equiv 7s^2 \equiv 7 \pmod 8,$$

and this would contradict our initial hypothesis on m. So there is a representation

$$m' = a_1^2 + a_2^2 + a_3^2 \quad (a_1, a_2, a_3 \in \mathbf{Z}),$$

which implies the representation

$$m = (ra_1)^2 + (ra_2)^2 + (ra_3)^2.$$

Our plan is as follows: Given a squarefree $m \not\equiv 7 \pmod 8$, we will produce a ternary quadratic form f of determinant 1 representing both 1 and m. By Theorem 2.7.5, f is equivalent to $x_1^2 + x_2^2 + x_3^2$. By the remarks following the definition of equivalence, $x_1^2 + x_2^2 + x_3^2$ must also represent m. It remains to construct such a form and verify it has the needed properties.

We suppose for the moment that it is possible to choose integers A, B, a, h and b so that

$$a > 0, \quad ab - h^2 = m, \tag{2.39}$$

and so that the polynomial f determined by

$$mf(x_1, x_2, x_3) = (Ax_1 + Bx_2 + mx_3)^2 + \varphi(x_2, x_3), \tag{2.40}$$

where

$$\varphi(x_1, x_2) = ax_1^2 + 2hx_1x_2 + bx_2^2,$$

is an integral ternary quadratic form. (Since any choice of A, B, a, h and b determines f as a homogeneous polynomial of degree 2, what needs to be verified is that f has integer coefficients and even cross-term coefficients.) We defer the proof of this, which will be accomplished by quadratic reciprocity and Dirichlet's theorem, to the next subsection.

We claim that the f determined this way has the properties needed to complete the proof. We need check four things:

i. The form f is positive-definite.

ii. The form f has determinant 1.

iii. The form f represents 1.

iv. The form f represents m.

We proceed to establish each of these in succession.

We can write (using (2.39))

$$\begin{aligned} mf(x_1, x_2, x_3) &= (Ax_1 + Bx_2 + mx_3)^2 + ax_2^2 + 2hx_2x_3 + bx_3^2 \\ &= (Ax_1 + Bx_2 + mx_3)^2 + \frac{1}{a}((ax_1 + hx_2)^2 + (ab - h^2)x_2^2) \\ &= (Ax_1 + Bx_2 + mx_3)^2 + \frac{1}{a}((ax_1 + hx_2)^2 + mx_2^2), \end{aligned}$$

so that

$$f(x_1, x_2, x_3) = \\ \left(\frac{A}{\sqrt{m}}x_1 + \frac{B}{\sqrt{m}}x_2 + \sqrt{m}x_3\right)^2 + \left(\sqrt{\frac{a}{m}}x_1 + \frac{h}{\sqrt{am}}x_2\right)^2 + (\frac{1}{\sqrt{a}}x_2)^2. \tag{2.41}$$

Since f is a sum of squares, we have $f(\mathbf{x}) \geq 0$ for every $\mathbf{x} \in \mathbf{Z}^3$. Suppose equality holds; then all three squares on the right hand side of (2.41) must vanish. Starting with the last term we find $x_2 = 0$. This together with the vanishing of the second term implies $x_1 = 0$, and these together with the vanishing of the first term imply $x_3 = 0$ also. So f is positive definite, and we have verified the first of the requisite properties.

To verify the second, introduce the ternary form $g = x_1^2 + x_2^2 + x_3^2$. Then we can rewrite (2.41) in the form

$$f = g\left(A\left(\begin{smallmatrix} x_1 \\ x_2 \\ x_3 \end{smallmatrix}\right)\right),$$

where

$$A = \begin{pmatrix} \frac{A}{\sqrt{m}} & \frac{B}{\sqrt{m}} & \sqrt{m} \\ \sqrt{\frac{a}{m}} & \frac{h}{\sqrt{am}} & 0 \\ 0 & \frac{1}{\sqrt{a}} & 0 \end{pmatrix}.$$

Therefore, since g corresponds to the identity matrix, we have (by the proof of (2.34), which nowhere requires that A have integer entries)

$$f\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} x_1 & x_2 & x_3 \end{pmatrix} A^T I_2 A \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix};$$

it follows that $A^T I_2 A = A^T A$ is the matrix of f, so that the determinant of f is

$$\det(A^T A) = \det(A)^2 = \left(\sqrt{m}\sqrt{a/m}\frac{1}{\sqrt{a}}\right)^2 = 1;$$

here we have computed the determinant of A by expanding by minors along the third column. We have therefore verified the second required property.

For the third property we again invoke Minkowski's theorem (Theorem 2.7.7). By (2.41), it suffices to show that the $\mathbf{Z}$-span $\mathcal{L}$ of the three vectors

$$\mathbf{v_1} := \begin{pmatrix} \frac{A}{\sqrt{m}} & \frac{B}{\sqrt{m}} & \sqrt{m} \end{pmatrix}, \quad \mathbf{v_2} := \begin{pmatrix} \sqrt{\frac{a}{m}} & \frac{h}{\sqrt{am}} & 0 \end{pmatrix}, \quad \mathbf{v_3} := \begin{pmatrix} 0 & \frac{1}{\sqrt{a}} & 0 \end{pmatrix}$$

intersects the sphere $u^2 + v^2 + w^2 < 2$ in a point other than the origin. Because $\det(A) = 1$, the vectors $\mathbf{v_1}, \mathbf{v_2}$ and $\mathbf{v_3}$ are linearly independent, so $\mathcal{L}$ is a complete lattice, and the fundamental parallelepiped (2.37) has volume 1. Since the sphere has volume

$$\frac{4}{3}\pi(\sqrt{2})^3 \approx 11.8 > 8 \cdot 1,$$

the third claim follows from Minkowski's theorem.

The fourth property is easiest of all to verify; since the coefficient of x_3^2 in f is m, we have $f\left(\begin{smallmatrix} 0 \\ 0 \\ 1 \end{smallmatrix}\right) = m$.

2.7.6 Completion of The Proof

We now show that for any squarefree $m \not\equiv 7 \pmod 8$, it is possible to choose the integers A, B, a, h and b in accordance with the previous conditions.

Expanding out the right hand side of (2.40) we find

$$\begin{aligned} mf(x_1, x_2, x_3) &= (Ax_1 + Bx_2 + mx_3)^2 + (ax_1^2 + 2hx_1x_2 + bx_2^2) \\ &= A^2x_1^2 + B^2x_2^2 + m^2x_3^2 + 2ABx_1x_2 + 2Bmx_2x_3 + \\ &\qquad 2Amx_1x_3 + ax_1^2 + 2hx_1x_2 + mx_2^2. \end{aligned}$$

Collecting like terms and dividing by m shows

$$f(x_1, x_2, x_3) = \left(\frac{A^2 + a}{m}\right) x_1^2 + \left(\frac{B^2 + b}{m}\right) x_2^2 + m x_3^2 \\ + 2\left(\frac{AB + h}{m}\right) x_1 x_2 + 2A x_1 x_3 + 2B x_2 x_3. \quad (2.42)$$

Now f will have integer coefficients and even cross-term coefficients provided the following congruences on the integers A, B, a, h and b hold:

$$A^2 \equiv -a \pmod{m}, \quad B^2 \equiv -b \pmod{m}, \quad AB \equiv -h \pmod{m}. \quad (2.43)$$

So we need to prove these congruences can be simultaneously satisfied with our other condition (2.39), namely

$$a > 0, \quad ab - h^2 = m. \quad (2.44)$$

To satisfy (2.44) it is sufficient to choose a to be a positive integer coprime to m for which $-m$ is a square. We can then choose h so $h^2 \equiv -m \pmod{a}$ and determine b to satisfy $ab - h^2 = m$. Note that because h is only determined $\pmod{a}$ and $\gcd(a, m) = 1$, we can choose h so that $h \equiv 0 \pmod{m}$; then $ab - h^2 = m$ implies $b \equiv 0 \pmod{m}$. We choose B as any integer with $B \equiv 0 \pmod{m}$. With all these choices, the final two congruences of (2.43) hold.

We have thus reduced the problem to showing we can choose a positive integer a coprime to m with $-a$ a square $\pmod{m}$ and $-m$ a square $\pmod{a}$. It is no surprise that quadratic reciprocity (used below for the Jacobi symbol) enters the picture at this point.

We take several cases depending on the residue class of $m \pmod{8}$.

Suppose first that $m \equiv 1 \pmod{4}$ (i.e., $m \equiv 1$ or $5 \pmod{8}$). Choose a to be an odd prime satisfying the congruences

$$a \equiv -1 \pmod{m}, \quad a \equiv 1 \pmod{4};$$

that this is possible is a consequence of Dirichlet's theorem. Then $-a$ is trivially a square $\pmod{m}$; to see $-m$ is a square $\pmod{a}$, observe

$$\left(\frac{-m}{a}\right) = \left(\frac{-1}{a}\right)\left(\frac{m}{a}\right) = \left(\frac{m}{a}\right) = \left(\frac{a}{m}\right) = \left(\frac{-1}{m}\right)\left(\frac{-a}{m}\right) = \left(\frac{-a}{m}\right) = \left(\frac{1}{m}\right) = 1.$$

Suppose next that $m \equiv 3 \pmod{8}$. In this case we take $a = 2p$, where p is an odd prime satisfying

$$2p \equiv -1 \pmod{m}, \quad p \equiv 1 \pmod{4}.$$

As before, $-a$ is trivially a square $\pmod{m}$. To see $-m$ is a square $\pmod{a}$, it suffices to check $-m$ is a square $\pmod{p}$. But

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

Finally, suppose m is even (so that $m \equiv 2$ or $6 \pmod 8$) and write $m = 2m_1$, where m_1 is odd. If $m_1 \equiv 1 \pmod 4$, choose a an odd prime with

$$a \equiv -1 \pmod m, \quad a \equiv 1 \pmod 8.$$

Then $-a$ is trivially a square $\pmod m$. To check $-m$ is a square $\pmod a$, observe

$$\left(\frac{-m}{a}\right) = \left(\frac{-2}{a}\right)\left(\frac{m_1}{a}\right) = \left(\frac{m_1}{a}\right) = \left(\frac{a}{m_1}\right) = \left(\frac{-1}{m_1}\right) = 1.$$

If $m_1 \equiv 3 \pmod 4$, choose a an odd prime with

$$a \equiv -1 \pmod m, \quad a \equiv 3 \pmod 8.$$

As before $-a$ is a square $\pmod m$. To check $-m$ is a square $\pmod a$ in this case, note

$$\left(\frac{-m}{a}\right) = \left(\frac{-2}{a}\right)\left(\frac{m_1}{a}\right) = \left(\frac{m_1}{a}\right) = -\left(\frac{a}{m_1}\right) = -\left(\frac{-1}{m_1}\right)\left(\frac{-a}{m_1}\right) = -(-1)(1) = 1.$$

2.7.7 The Number of Representations

There are elementary formulas in terms of divisor sums for the number of representations of an integer n as a sum of 2 and 4 squares. Namely, with $r_2(n)$ and $r_4(n)$ defined in analogy with (2.32), we have (see [IR90, Proposition 17.6.1, Corollary to Proposition 17.7.2])

$$r_2(n) = 4\left(\sum_{\substack{d\mid n\\ 2\nmid d}} (-1)^{(d-1)/2}\right), \quad \text{and} \quad r_4(n) = \begin{cases} 8\sum_{d\mid n} d & \text{if } n \text{ is odd,}\\ 24\sum_{d\mid n} d & \text{if } n \text{ is even.}\end{cases}$$

However, for sums of three squares the situation is much more complicated. In his *Disquisitiones Arithmeticae* (1801), Gauss expressed the number of representations in terms of certain (complicated) quantities arising in his theory of binary quadratic forms.

We conclude this chapter by stating the following theorem (see [Bat51]) which expresses $r_3(n)$ in terms of $L(1,\chi)$ for a certain character $\chi \pmod{4n}$:

Theorem. *Let n be a positive integer, with $n = 4^a n_1$ and $4 \nmid n_1$. Then*

$$r_3(n) = \frac{16}{\pi}\sqrt{n}L(1,\chi)q(n)P(n),$$

where

$$q(n) = \begin{cases} 0 & \text{if } n_1 \equiv 7 \pmod 8,\\ 2^{-a} & \text{if } n_1 \equiv 3 \pmod 8,\\ 3\cdot 2^{a-1} & \text{if } n_1 \equiv 1, 2, 5 \text{ or } 6 \pmod 8,\end{cases}$$

$$P(n) = \prod_{\substack{p^{2b} \| n \\ p \text{ odd}, b \geq 1}} \left(1 + \sum_{j=1}^{b-1} p^{-j} + p^{-b}\left(1 - \left(\frac{(-n/p^{2b})}{p}\right)\frac{1}{p}\right)^{-1}\right)$$

(understood so that $P(n) = 1$ for squarefree n), and χ is the character $\pmod{4n}$ *given by*

$$\chi(m) := \begin{cases} 0 & \text{if } m \text{ is even,} \\ \left(\frac{-n}{m}\right) & \text{if } m \text{ is odd.} \end{cases}$$

The standard reference on sums of squares is [Gro85]. Chapter 4 of that text contains a wealth of additional information on representability by three squares, including a detailed discussion of alternate expressions for $r_3(n)$.

2.7.8 Exercises

Exercise 2.7.1. Let $\mathcal{L} \subset \mathbf{R}^n$ be a complete lattice, i.e., the $\mathbf{Z}$-span of n linearly independent vectors.

a) Prove that any bounded subset of $\mathbf{R}^n$ contains only finitely many elements of $\mathcal{L}$.

b) Prove that the conclusion of Minkowski's Theorem continues to hold whenever the *nonstrict* inequality $\mathrm{vol}(X) \geq 2^n V$ is satisfied, provided the given convex, centrally symmetric set X is closed.

Exercise 2.7.2 *(Minkowski).* It is known that for any symmetric matrix M of rational numbers, one can find a matrix A of rational numbers with $A^T MA$ a diagonal matrix.

a) Suppose f is an n-ary integral quadratic form and let M be the $n \times n$ symmetric matrix associated to f. Let A be as above. Prove that f is positive definite if and only if the diagonal entries of $A^T MA$ are all positive. Deduce that the determinant D of a positive definite form always satisfies $D > 0$.

b) Prove that if f is a positive definite (integral) n-ary quadratic form of determinant D, then there exists $\mathbf{v} \in \mathbf{Z}^n$ with

$$0 < f(\mathbf{v}) \leq 4\omega_n^{-2/n} \sqrt[n]{D},$$

where ω_n is the volume of the n-dimensional unit ball.

Exercise 2.7.3. Use Theorem 2.7.9 to prove that every prime $p \equiv 1 \pmod 4$ is expressible as a sum of two squares. Suggestion: let $a^2 \equiv -1 \pmod p$ and consider the form

$$px_1^2 + 2ax_1x_2 + \left(\frac{a^2+1}{p}\right)x_2^2.$$

Exercise 2.7.4. Let m be an odd positive integer.

a) Prove that m can be written as a sum of four squares with two of them equal.

b) Prove that m can be written as a sum of four squares with two of them consecutive.

Suggestions: For a), write $2m = x^2 + y^2 + z^2$, as is possible by Legendre's theorem; by reordering, show we can assume x, y are odd while z is even. Let $a = (x+y)/2, b = (x-y)/2$ and use the identity

$$a^2 + b^2 = \frac{1}{2}\left(x^2 + y^2\right). \tag{2.45}$$

Complete the proof by taking $c = z/2$. For b), write $2m - 1 = x^2 + y^2 + z^2$. Show that after a rearrangement we can assume $x = 2a, y = 2b$ are even while $z = 2c + 1$ is odd. Then use the identities (2.45) and

$$c^2 + (c+1)^2 = \frac{1}{2}\left((2c+1)^2 + 1\right).$$

Exercise 2.7.5 *(Turski [Tur33]).* Prove that every positive integer is the sum of at most 10 odd squares and that infinitely many require 10.

Exercise 2.7.6. Prove that the set of positive integers expressible as a sum of three squares has density 5/6.

Exercise 2.7.7 *(Carlitz [Car75]).* Let F be a finite field of odd order q. Prove that every polynomial in $F[T]$ is a sum of three squares. Show that "three" can be replaced by "two" if $q \equiv 1 \pmod 4$, but not if $q \equiv 3 \pmod 4$.

Hint (for the first part): First consider the problem of expressing

$$T = (rT + r')^2 + (sT + s')^2 + (uT + u')^2 \tag{2.46}$$

with $r, s, u, r', s', u' \in F$. This is equivalent to the system

$$r^2 + s^2 + u^2 = r'^2 + s'^2 + u'^2 = 0, \quad 2rr' + 2ss' + 2uu' = 1.$$

To solve this system, choose r and s with $r^2 + s^2 + 1 = 0$ (arguing that this is always possible by the Pigeonhole principle), and let $u = 1$. Then let $r'' = -s, s'' = r$ and $u'' = 1$, and scale these appropriately to obtain r', s' and u' respectively. Finally, replace T in (2.46) with a generic $f(T) \in F[T]$.

References

[ABGU01] J. M. Aldaz, A. Bravo, S. Gutiérrez, and A. Ubis, *A theorem of D. J. Newman on Euler's ϕ function and arithmetic progressions*, Amer. Math. Monthly **108** (2001), no. 4, 364–367. MR 2002i:11099

[Ank57] N. C. Ankeny, *Sums of three squares*, Proc. Amer. Math. Soc. **8** (1957), 316–319. MR 19,15d

[Apo76] Tom M. Apostol, *Introduction to analytic number theory*, Springer-Verlag, New York, 1976, Undergraduate Texts in Mathematics. MR 55 #7892

[Bat51] Paul T. Bateman, *On the representations of a number as the sum of three squares*, Trans. Amer. Math. Soc. **71** (1951), 70–101. MR 13,111i

[Car75] L. Carlitz, *A note on sums of three squares in* GF[q, x], Math. Mag. **48** (1975), 109–110. MR 50 #12987

[CL77] Gerard J. Chang and Ko Wei Lih, *Polynomial representation of primes*, Tamkang J. Math. **8** (1977), no. 2, 197–198. MR 58 #27718

[Dir37] P.G.L. Dirichlet, *Beweis der Satz, dass jede unbegrentze arithmetische progression, deren erstes Glied und Differenz ganze Zahlen ohne gemeinschaftliche Factor sind, unendliche viele Primzahlen enthält.*, Abh. der Königlichen Preuss. Akad. der Wiss. (1837), 45–81.

[EO37] P. Erdős and R. Obláth, *Über diophantische Gleichungen der Form* $n! = x^p \pm y^p$ *und* $n! \pm m! = x^p$, Acta Litt. Sci. Univ., Szeged **8** (1937), 241–255.

[ER50] P. Erdös and A. Rényi, *Some problems and results on consecutive primes*, Simon Stevin **27** (1950), 115–125. MR 11,644d

[GL66] A. O. Gel′fond and Yu. V. Linnik, *Elementary methods in the analytic theory of numbers*, Translated from the Russian by D. E. Brown. Translation edited by I. N. Sneddon. International Series of Monographs in Pure and Applied Mathematics, Vol. 92, Pergamon Press, Oxford, 1966. MR 34 #1252

[Gro85] Emil Grosswald, *Representations of integers as sums of squares*, Springer-Verlag, New York, 1985. MR 87g:11002

[IR90] Kenneth Ireland and Michael Rosen, *A classical introduction to modern number theory*, second ed., Graduate Texts in Mathematics, vol. 84, Springer-Verlag, New York, 1990. MR 92e:11001

[Mar99] Greg Martin, *The smallest solution of* $\phi(30n + 1) < \phi(30n)$ *is* ..., Amer. Math. Monthly **106** (1999), no. 5, 449–451. MR 2000d:11011

[Mer97] F. Mertens, *Über Dirichlet's Beweis des Satzes, daß jede unbegrenzte ganzzahlige arithmetische Progression, deren Differenz zu ihren Gliedern teilerfremd ist, unendlich viele Primzahlen enthält.*, Sber. Kais. Akad. Wissensch. Wien **106** (1897), 254–286.

[Mon93] Paul Monsky, *Simplifying the proof of Dirichlet's theorem*, Amer. Math. Monthly **100** (1993), no. 9, 861–862. MR 95b:11085

[Mor58] L. J. Mordell, *On the representation of a number as a sum of three squares*, Rev. Math. Pures Appl. **3** (1958), 25–27. MR 23 #A117

[Nar00] Władysław Narkiewicz, *The development of prime number theory*, Springer Monographs in Mathematics, Springer-Verlag, Berlin, 2000, From Euclid to Hardy and Littlewood. MR 2001c:11098

[Nat00] Melvyn B. Nathanson, *Elementary methods in number theory*, Graduate Texts in Mathematics, vol. 195, Springer-Verlag, New York, 2000. MR 2001j:11001

[New97] D. J. Newman, *Euler's ϕ function on arithmetic progressions*, Amer. Math. Monthly **104** (1997), no. 3, 256–257. MR 97m:11010

[PS73] Richard M. Pollack and Harold N. Shapiro, *The next to last case of a factorial diophantine equation*, Comm. Pure Appl. Math. **26** (1973), 313–325. MR 50 #12915

[Ros02] Michael Rosen, *Number theory in function fields*, Graduate Texts in Mathematics, vol. 210, Springer-Verlag, New York, 2002. MR 1 876 657

[Sha50] Harold N. Shapiro, *On primes in arithmetic progression. II*, Ann. of Math. (2) **52** (1950), 231–243. MR 12,81b

[Sha83] ———, *Introduction to the theory of numbers*, Pure and Applied Mathematics, John Wiley & Sons Inc., New York, 1983, A Wiley-Interscience Publication. MR 84f:10001

[Sie48] W. Sierpiński, *Remarque sur la répartition des nombres premiers*, Colloquium Math. **1** (1948), 193–194. MR 10,431b

[Tit39] E. C. Titchmarsh, *The theory of functions*, second ed., The Clarendon Press Oxford University Press, New York, 1939.

[Tur33] S. Turski, *Décomposition de nombres entiers en sommes de carrés de nombres impairs*, Bull. Soc. Roy. Sci. Liége **2** (1933), 70–71.

[Wil74] Kenneth S. Williams, *Mertens' theorem for arithmetic progressions*, J. Number Theory **6** (1974), 353–359. MR 51 #392

Chapter 3

Sieve Methods

> Brun's [sieve] method ... is perhaps our most powerful elementary tool in number theory. – P. Erdős, *Some Recent Advances and Current Problems in Number Theory* [Erd65]

3.1 Introduction

3.1.1 The Sieve of Eratosthenes

Granville has pointed out [Gra95] that ancient Greek mathematics produced two results in prime number theory that have proved of first importance in subsequent thought. The first is Euclid's proof of the infinitude of the primes, which we have discussed in the opening to Chapter 1. The second is the sieve of Eratosthenes.

Eratosthenes' method allows one to determine the primes not exceeding x based on knowledge only of the primes not exceeding $\sqrt{x}$. His procedure is roughly as follows: begin with a list of all positive integers at least 2 but not exceeding x and for each prime $p \leq \sqrt{x}$ cross out all the multiples of p on the list; the numbers remaining are exactly the primes in the interval $(\sqrt{x}, x]$. We can illustrate this process with $x = 30$, sieving by the primes $2, 3, 5$.

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

This is remarkable not only insofar as it gives a fast algorithm for listing primes, but also in that it suggests the useful viewpoint of the primes as the integers surviving a "sieving process."

3.1.2 Legendre's Formula

Let us attempt to count how many integers remain after Eratosthenes' sieving procedure is carried out. More generally, let us count the number of positive integers up to x remaining after the deletion (or "sifting out") of the multiples of all primes not exceeding z, where z is a parameter at our disposal; in Eratosthenes' sieve, $z = \sqrt{x}$. We use $\pi(x,z)$ to denote this quantity, i.e.,

$$\pi(x,z) := |\{n \le x : p \mid n \Rightarrow p > z\}.$$

Then for any z,

$$\pi(x) \le z + \pi(x,z),$$

and

$$\pi(x, x^{1/2}) = \pi(x) - \pi(\sqrt{x}) + 1.$$

Our estimate of $\pi(x,z)$ proceeds by successive approximation. We begin with the total number $\lfloor x \rfloor$ of positive integers not exceeding x, and then for each prime $p \le z$ we subtract off the number of multiples of p:

$$\lfloor x \rfloor - \sum_{p_1 \le z} \left\lfloor \frac{x}{p_1} \right\rfloor.$$

This counts correctly those n with at most one prime factor, but those n with two or more prime factors $p \le z$ have been subtracted off twice. Hence, we add these back in to obtain our next approximation,

$$\lfloor x \rfloor - \sum_{p_1 \le z} \left\lfloor \frac{x}{p_1} \right\rfloor + \sum_{p_1 < p_2 \le z} \left\lfloor \frac{x}{p_1 p_2} \right\rfloor.$$

But now those integers divisible by three primes $p \le z$ have been added back in too many times; for instance, if n has exactly three prime divisors not exceeding z, it is counted with weight $1 - 3 + 3 > 0$. Thus we should subtract off a term corresponding to the integers divisible by three primes $p \le z$; we would then find ourselves needing to add a term corresponding to integers divisible by four such p, etc. Continuing in this manner, we are led to the formula

$$\pi(x,z) = \lfloor x \rfloor - \sum_{p_1 \le z} \left\lfloor \frac{x}{p_1} \right\rfloor + \cdots + (-1)^r \sum_{p_1 < \cdots < p_r \le z} \left\lfloor \frac{x}{p_1 \ldots p_r} \right\rfloor. \tag{3.1}$$

Note that if we set

$$P := \prod_{p \le z} p,$$

we can put this in the alternate form

$$\pi(x,z) = \sum_{d \mid P} \mu(d) \left\lfloor \frac{x}{d} \right\rfloor. \tag{3.2}$$

This line of reasoning, attributed to Legendre, can be tightened into a proof of (3.1). For the time being, we assume (3.1), postponing a rigorous justification to §3.3, where we shall establish a more general result.

3.1.3 Consequences

We now have an exact formula for $\pi(x,z)$. Unfortunately this exact formula is a bit unsatisfying, because it seems to leave the most natural question unanswered: how large is $\pi(x,z)$? Can we transition from our formula to an estimate?

Sums involving the greatest-integer function are generally hard to work out directly, so we drop the greatest integer signs in (3.2) and transfer the incurred error to a separate sum. This is advantageous, as the "main term" can now be reexpressed as a product:

$$\begin{aligned}\pi(x,z) &= x\sum_{d\mid P}\frac{\mu(d)}{d} + \sum_{d\mid P}\mu(d)\left(\left\lfloor\frac{x}{d}\right\rfloor - \frac{x}{d}\right)\\ &= x\prod_{p\le z}\left(1-\frac{1}{p}\right) + \sum_{d\mid P}\mu(d)\left(\left\lfloor\frac{x}{d}\right\rfloor - \frac{x}{d}\right) = x\prod_{p\le z}\left(1-\frac{1}{p}\right) + O\left(2^{\pi(z)}\right);\end{aligned} \tag{3.3}$$

here the final estimate, understood to hold with an absolute implied constant, comes from noting that each of the $\tau(P) = 2^{\pi(z)}$ terms in the sum has magnitude bounded by 1.

How useful is estimate (3.3)? Suppose first that z is fixed while x is tending to infinity; then the error term in (3.3) is $O_z(1)$ and we obtain the asymptotic formula $\pi(x,z) \sim x\prod_{p\le z}(1-1/p)$. The same asymptotic estimate holds if z is not fixed, but instead is tending to infinity with x sufficiently slowly. Whenever $z = z(x) \to \infty$, Mertens' theorem implies

$$x\prod_{p\le z}(1-1/p) \sim e^{-\gamma}x/\log z \quad (x\to\infty). \tag{3.4}$$

If this $z = z(x)$ satisfies $z \le \log x$ whenever x is sufficiently large, then the O-term in (3.3) is $\ll 2^z \ll x^{\log 2}$, which is of smaller order than the main term $x/\log z$. Consequently, the asymptotic formula $\pi(x,z) \sim e^{-\gamma}x/\log z$ holds in this case as well. Fixing the choice $z = \log x$, we obtain the corollary

$$\pi(x) \le \pi(x,\log x) + \log x \sim e^{-\gamma}\frac{x}{\log\log x},$$

which provides another proof that the primes have density 0.

We have yet to treat the case corresponding to Eratosthenes' sieve, that of $z = \sqrt{x}$. In this case, the bound $2^{\pi(\sqrt{x})}$ for the "error term" dwarfs the value of the "main term,"[1]

$$x\prod_{p\le x^{1/2}}(1-1/p) \sim 2e^{-\gamma}x/\log x = (1.229\ldots)x/\log x, \tag{3.5}$$

so that we cannot prove the asymptotic formula by this method. In fact, the prime number theorem implies

$$\pi(x,\sqrt{x}) = \pi(x) - \pi(\sqrt{x}) + 1 \sim x/\log x, \tag{3.6}$$

so that *it is not even true* that $\pi(x,x^{1/2}) \sim x\prod_{p\le x^{1/2}}(1-1/p)$.

We will discuss to what extent this difficulty can be overcome in §3.1.5.

[1] For example, Chebyshev's results imply $2^{\pi(\sqrt{x})} \ge 2^{\sqrt[3]{x}}$ for all large x.

3.1.4 General Sieving Situations

The problem treated in the last section is of the following form: Given a finite sequence of integers $\mathcal{A}$ and a finite set of primes $\mathcal{P}$, estimate the number $S(\mathcal{A}, \mathcal{P})$ of terms of $\mathcal{A}$ divisible by no prime $p \in \mathcal{P}$. For example, if

$$\mathcal{A} := \{n \leq x\}, \quad \mathcal{P} = \{p \leq z\}, \tag{3.7}$$

then $S(\mathcal{A}, \mathcal{P})$ is what we have been calling $\pi(x, z)$.

Many problems in number theory fit into this framework. For example, let x be a positive real number and z a parameter to be chosen later, and set

$$\mathcal{A} := \{n(n+2) : n \leq x\}, \quad \mathcal{P} = \{p \leq z\}. \tag{3.8}$$

If $n, n+2$ are both prime, then either $n \leq z$ or both $n, n+2$ have only prime factors exceeding z. Consequently,

$$\pi_2(x) \leq S(\mathcal{A}, \mathcal{P}) + z. \tag{3.9}$$

Since n and $N - n$ are both prime if all their prime factors exceed $\sqrt{x+2}$, in the particular case $z = \sqrt{x+2}$ we additionally have

$$0 \leq \pi_2(x) - S(\mathcal{A}, \mathcal{P}) \leq z. \tag{3.10}$$

Estimates for $S(\mathcal{A}, \mathcal{P})$ are thus intimately connected with the quantative form of the twin prime conjecture (see Chapter 1, §1.8).

The sieve problem in its general form is too vague to be tractable, so it is necessary to make a few further assumptions. We assume $\mathcal{A}$ has "approximately" X elements and that divisibility by distinct primes $p \in \mathcal{P}$ constitute "approximately" mutually independent events, each occurring with "approximate" probability $\alpha(p)$. (All of this will be made precise in §3.2.) In this case, it is natural to expect

$$S(\mathcal{A}, \mathcal{P}) \approx X \prod\nolimits_{p \in \mathcal{P}} (1 - \alpha(p)). \tag{3.11}$$

The goal of sieve theory, from our perspective in this chapter, is to quantify and then to justify such approximations, in as wide a range of circumstances as possible.

In the classical situation described by (3.7), it is reasonable to approximate the number of integers $n \leq x$ by x and the probability such an integer is divisible by p by $1/p$. Our expectation translates into the guess $\pi(x, z) \approx x \prod_{p \leq z} (1 - 1/p)$. We have seen that when z is constant or slow-growing this holds as an asymptotic estimate, but that for $z = \sqrt{x}$, the case originally of interest, the estimate is off by a constant factor. Nevertheless, the approximation sign in (3.11) correct if read as an assertion that both sides have the same order of magnitude.

For another example, consider the situation described by (3.8). We again approximate $|\mathcal{A}|$ by x. This time the probability that a term of the sequence $\{n(n+2)\}_{n \leq x}$ is divisible by the prime p is approximately $\omega(p)/p$, where

$$\omega(p) = \begin{cases} 1 & \text{if } p = 2, \\ 2 & \text{otherwise.} \end{cases} \tag{3.12}$$

(So that $\omega(p)$ counts the number of solutions $\pmod p$ to $n(n+2) \equiv 0 \pmod p$.) Assuming $z = z(x)$ tends to infinity with x, our approximation asserts (cf. Exercise 3.1.3)

$$S(\mathcal{A}, \mathcal{P}) \approx x \prod_{p \le z} (1 - \omega(p)/p) \sim 2C_2 e^{-2\gamma} \frac{x}{\log^2 z}, \tag{3.13}$$

where

$$C_2 = \prod\nolimits_{p>2} \left(1 - (p-1)^{-2}\right) \tag{3.14}$$

is the twin prime constant. Using Legendre's methods (cf. Exercise 3.3.4), this approximation can be proved to hold as an asymptotic formula for z small (say $z \le \frac{1}{2} \log x$) and x tending to infinity; probably no method can establish the same for $z = \sqrt{x+2}$, as in combination with (3.10) this would contradict the quantitative form of the twin prime conjecture.

3.1.5 Legendre, Brun and Hooley; oh my!

We have already stated that the goal of sieve theory, for us, is to quantify and to justify estimates of the form

$$S(\mathcal{A}, \mathcal{P}) \approx X \prod\nolimits_{p \in \mathcal{P}} (1 - \alpha(p)).$$

We can get a feel for the respective power of the three sieve methods of this chapter if we consider what they say about the particular estimate

$$\pi(x, z) \approx x \prod\nolimits_{p \le z} (1 - 1/p)$$

corresponding to our initial problem. As we have noted previously, Legendre's method of successive approximation shows this is valid as an asymptotic formula for $z = \log x$. The first improvement on Legendre's formula, Brun's pure sieve, allows one to prove the same for any $z = z(x) \to \infty$ satisfying the inequality $z(x) \le x^{1/10 \log\log x}$ (for large x); in particular, choosing z as large as possible and referring to (3.4) shows

$$\pi(x) \le \pi(x, z) + z \ll \frac{x}{\log x} \log\log x. \tag{3.15}$$

Whereas the other methods yield asymptotic formulae for z in a certain limited range, the powerful Brun-Hooley sieve additionally allows one to obtain upper and lower bounds for $\pi(x, z)$ for z as large as a power of x. The upper bound aspect permits recovery of Chebyshev's estimate $\pi(x) \ll x/\log x$.[2] The lower bound aspect is also interesting; it permits deduction of bounds like $\pi(x, x^{1/1000}) \gg x/\log x$. This does *not* translate into a lower bound on the number of primes up to x. But because an integer up to x all of whose prime

[2]Take this with a grain of salt. In the derivation, we require the results of Mertens, which in turn rest on those of Chebyshev.

factors exceed $x^{1/1000}$ can have at most 1000 prime factors, it *does* give us a lower bound on the number of 1000-almost primes up to x, where an r-*almost prime* is an integer with no more than r prime divisors, counting multiplicity.

This might seem a bit silly because we already have a lower bound for $\pi(x)$ of the correct order of magnitude. But the general sieve framework is rather flexible, and therein lies the allure of this approach. We have already seen that sieve methods can be adapted to yield information about the twin prime conjecture. Developing these ideas, Brun used his pure sieve to prove (in analogy with (3.15))

$$\pi_2(x) \ll \frac{x}{\log x}(\log\log x)^2. \tag{3.16}$$

This is off by a factor of $(\log\log x)^2$ from the conjectured order of magnitude, but it still has profound implications. One such is that $\sum_p 1/p$, restricted to twin primes p, is either a finite sum or a convergent infinite series.

To remove the unwanted factor $(\log\log x)^2$ from (3.16), Brun required the complicated combinatorial apparatus of his full sieve. We will reach the same goal via the much simpler Brun-Hooley sieve. The same method will allow us to prove the following two deep theorems of Brun ([Bru20]; see [Wan84] for an English translation), approximations to the twin prime and Goldbach conjectures respectively:

- There are infinitely many pairs of 9-almost primes $n, n+2$.
- Every large even integer N is a sum of two 9-almost primes.

In the next section, we formally introduce some notions and notations arising in the general sieving situation. We then discuss the first of our sieve methods, that of Eratosthenes-Legendre. This is a straightforward adaptation of Legendre's method of successive approximation to the general sieving situation. As promised, we present a rigorous justification of Legendre's manipualtions via the well-known Principle of Inclusion-Exclusion from enumerative combinatorics. After giving a few elementary applications, we turn to a discussion of Brun's pure sieve, so-named because of its purely combinatorial nature. Indeed, it rests on the purely combinatorial observation that the approximations in Legendre's method are alternately over and under-estimates. Despite being an easy variant on Legendre's method, it is much more powerful, which we illustrate by proving the aforementioned theorem of Brun on the sum of the reciprocals of the twin primes. We conclude with a discussion of Hooley's elegant and surprisingly powerful "almost-pure" sieve. Our treatment is based on Hooley's original article [Hoo94] and the recent exposition of Ford & Halberstam [FH00].

3.1.6 Further Reading

We make no claims to comprehensiveness in this chapter, rather the opposite. For a thorough discussion of sieve methods, the reader can consult the classic text of Halberstam & Richert [HR74] or the recent monograph of Greaves [Gre01], though the novice should be warned that neither makes for particularly pleasant bedtime reading. The remarkably readable and illuminating notes of Ben Green [Gre] on Brun's pure sieve and the Selberg sieve are strongly urged on the reader. Another accessible reference for these (and more advanced topics) is Odlyzko's senior thesis [Odl71].

3.1.7 Exercises

Exercise 3.1.1 *(An Alternate Derivation of the Bateman-Horn Heuristic; cf. [Pól59]).* In this section we saw that the natural but naive guess for the number of primes up to x, viz.

$$x \prod_{p \leq x^{1/2}} \left(1 - \frac{1}{p}\right), \tag{3.17}$$

exceeds $\pi(x)$ by the multiplicative factor $2e^{-\gamma}(1 + o(1))$ (cf. (3.5), (3.6)).

a) Using the PNT and Mertens' theorem, show that if the exponent $1/2$ on x is replaced with $e^{-\gamma}$, then one obtains the correct asymptotic. That is, prove that

$$x \prod\nolimits_{p \leq x^{e^{-\gamma}}} \left(1 - \frac{1}{p}\right) \sim \frac{x}{\log x}.$$

b) Let $f_1, \ldots, f_k$ be polynomials satisfying the conditions of the Bateman-Horn conjecture (see §1.8.3), and let us attempt to estimate the number of positive values not exceeding x at which the f_i are simultaneously prime. Using $\omega_f(p)$ to denote the number of roots of a polynomial f over $\mathbf{Z}/p\mathbf{Z}$, the estimate corresponding to (3.17) is

$$x \prod_{p} \left(1 - \frac{\omega^*(p)}{p}\right), \quad \omega^*(p) = \begin{cases} \omega_{f_1 \cdots f_k}(p) & \text{if } p \leq x^{\frac{1}{2}d_i}, \\ \omega_{f_i f_{i+1} \cdots f_k}(p) & \text{if } x^{\frac{1}{2}d_{i-1}} < p \leq x^{\frac{1}{2}d_i}, \\ 0 & \text{if } p > x^{\frac{1}{2}d_k}; \end{cases}$$

here d_i denotes the degree of f_i, and the d_i are assumed to be nondecreasing. (This is probably easiest to first grok in the special case $k = 1$.)

Show that if $\frac{1}{2}$ is replaced by $e^{-\gamma}$ in the definition of ω^*, then one obtains the asymptotic formula (1.52) of the Bateman-Horn conjecture. (You should assume that the infinite product (1.53) defining the "correction factor" of that conjecture always converges.)

Exercise 3.1.2. Use the elementary lower bound (valid for $z \geq 1$)

$$\prod_{p \leq z} \left(1 - \frac{1}{p}\right)^{-1} = \prod_{p \leq z} \sum_{j=0}^{\infty} \frac{1}{p^j} \geq \sum_{n \leq z} \frac{1}{n} \geq \log z \tag{3.18}$$

to deduce that with $\pi(x, z)$ as above,

$$\pi(x, z) \leq \frac{x}{\log z} + O(x^{\log 2}). \tag{3.19}$$

Thus obtain the bound

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

without Mertens' theorem.

Exercise 3.1.3. Using Mertens' theorem, establish that with $\omega(p)$ as in (3.12) and C_2 as in (3.14), $\prod_{p\le z}(1-\omega(p)/p) \sim 2C_2 e^{-2\gamma}/\log^2 z$ $(z\to\infty)$.

Exercise 3.1.4. Let d be a positive integer. Show that the number of integers in any interval $(Y, Y+X]$ divisible by d is $X/d + O(1)$, where the implied constant is absolute. Use this to prove that for every $X \ge 3$, the number of primes in any interval of the form $(Y, Y+X]$ (where Y is nonnegative) is $O(X/\log\log X)$, where the implied constant is absolute.

3.2 The General Sieve Problem: Notations and Preliminaries

> Probability is not a notion of pure mathematics, but of philosophy or physics. – G.H. Hardy & J.E. Littlewood, *Some Problems of Partitio Numeronum, III* [HL22]

The general sieve problem (for us) takes the following form: Given a finite sequence $\mathcal{A} = \{a_i\}$ of integers and a finite set of primes $\mathcal{P}$, estimate the quantity

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

where $P := \prod_{p\in\mathcal{P}} p$.

In many situations, the sifting set of primes arises by truncating of an infinite set of primes. Consequently, it is expedient to allow the set of primes $\mathcal{P}$ to be infinite and to introduce special notation indicating that we sieve only by those primes $p \in \mathcal{P}$ with $p \le z$. We thus define

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

where

$$P(z) := \prod_{\substack{p\in\mathcal{P}\\ p\le z}} p.$$

Thus $S(\mathcal{A}, \mathcal{P}, z) = S(\mathcal{A}, \mathcal{P}\cap[2, z])$.

We use the notation A_d to denote the number of terms of $\mathcal{A}$ divisible by d, i.e.,

$$A_d = |\{a \in \mathcal{A} : d \mid a\}|.$$

The letter X denotes an approximation to the size of $\mathcal{A}$. We assume the existence of a multiplicative function α taking values in $[0, 1]$ for which

$$A_d := X\alpha(d) + r(d) \tag{3.20}$$

for each $d \mid P$ (or each $d \mid P(z)$, as the case may be). In practice, we *choose* X and α, and we *define* $r(d)$, for $d \mid P$, in order that equation (3.20) holds.

To see how this corresponds to the explanation of the introduction, note that (3.20) asserts the probability an element of $\mathcal{A}$ is divisible by d is "approximately" $\alpha(d)$, while the multiplicativity of α says that the events corresponding to divisibilities by primes $p \in \mathcal{P}$ are "approximately" mutually independent.

3.3 The Sieve of Eratosthenes-Legendre and its Applications

3.3.1 The Principle of Inclusion-Exclusion

Any rigorous study of sieve methods begins with the following fundamental principle, a well-known result from enumerative combinatorics:

Theorem 3.3.1 (Principle of Inclusion-Exclusion). *Let X be a nonempty, finite set of N objects, and let $P_1, \ldots, P_r$ be properties elements of X may have. For any subset $I \subset \{1, 2, \ldots, r\}$, let $N(I)$ denote the number of elements of X that have each of the properties indexed by the elements of I. Then with N_0 denoting the number of elements of X with none of these properties, we have*

$$N_0 = \sum_{k=0}^{r} (-1)^k \sum_{\substack{I \subset \{1,2,\ldots,r\} \\ |I|=k}} N(I) = \sum_{I \subset \{1,2,\ldots,r\}} (-1)^{|I|} N(I). \tag{3.21}$$

Proof. Suppose $x \in X$ has exactly l of these properties. If $l = 0$, then x is counted only once, in the term $N(\emptyset)$. On the other hand, if x has $1 \leq l \leq r$ properties, then the number of k element subsets I for which x is counted in $N(I)$ is exactly $\binom{l}{k}$, and the total weight with which

$$\sum_{k=0}^{r} (-1)^k \sum_{\substack{I \subset \{1,2,\ldots,r\} \\ |I|=k}} N(I)$$

counts x is given by

$$\sum_{k=0}^{l} (-1)^k \binom{l}{k} = (1-1)^l = 0,$$

by the binomial theorem. □

3.3.2 A First Sieve Result

The Principle of Inclusion-Exclusion can be applied immediately to the general situation of §3.2. As may be expected from the introduction, the utility of the resulting estimates is crippled by our need to choose parameters in such a way as to keep the error terms in check. Nevertheless, the bounds obtained are still powerful enough to yield interesting consequences.

Theorem 3.3.2 (Sieve of Eratosthenes-Legendre).

$$S(\mathcal{A},\mathcal{P}) = X\prod_{p\in\mathcal{P}}(1-\alpha(p)) + \sum_{d\mid P}\mu(d)r(d).$$

Proof. Let $p_1,\dots,p_r$ be a list of the primes in $\mathcal{P}$, and for each i let P_i be the property of being divisible by p_i. For every $d \mid P$, there are $X\alpha(d)+r(d)$ terms $a\in\mathcal{A}$ divisible by d. The number of $a\in\mathcal{A}$ divisible by no prime $p\in\mathcal{P}$ is, by the Principle of Inclusion-exclusion,

$$\begin{aligned}
\sum_{k=0}^{r}(-1)^k \sum_{\substack{I\subset\{1,2,\dots,r\}\\ |I|=k}} N(I) &= \sum_{k=0}^{r}(-1)^k \sum_{\substack{d\mid P\\ \nu(d)=k}} A_d\\
&= \sum_{k=0}^{r}\sum_{\substack{d\mid P\\ \nu(d)=k}} \mu(d)\left(X\alpha(d)+r(d)\right)\\
&= X\sum_{d\mid P}\mu(d)\alpha(d) + \sum_{d\mid P}\mu(d)r(d)\\
&= X\prod_{p\in\mathcal{P}}(1-\alpha(p)) + \sum_{d\mid P}\mu(d)r(d),
\end{aligned}$$

where in the last line we have used the multiplicativity of α to express the sum as a product. $\square$

As a first application, we prove:

Corollary 3.3.3. *Let $\mathcal{P}$ be a set of prime numbers, and let $M(\mathcal{P})$ denote the set of integers n divisible by some prime $p\in\mathcal{P}$. Then $M(\mathcal{P})$ has natural density $1-\prod_{p\in\mathcal{P}}(1-1/p)$.*

Proof. Let $\mathcal{S}$ denote the complementary set of integers divisible by no prime $p\in\mathcal{P}$. Then with $\mathcal{A} := \{n\le x\}$, we have

$$S(x) \le S(\mathcal{A},\mathcal{P},z) \tag{3.22}$$

for any choice of z. To estimate $S(\mathcal{A},\mathcal{P},z)$ we take

$$X = x, \qquad \alpha(n) = 1/n \quad (n=1,2,\dots),$$

and note that with this choice of X and α, the remainders $r(d)$ satisfy $|r(d)|\le 1$ for every $d\mid P$. Consequently, choosing $z=\log x$,

$$\begin{aligned}
S(\mathcal{A},\mathcal{P},\log x) &= x\prod\nolimits_{\substack{p\in\mathcal{P}\\ p\le\log x}}(1-1/p) + O(x^{\log 2}) &(3.23)\\
&= \left(\prod\nolimits_{p\in\mathcal{P}}(1-1/p)+o(1)\right)x = (C+o(1))x, &(3.24)
\end{aligned}$$

say. We now take two cases, according as C is vanishing or nonvanishing. In the former case, (3.22) coupled with (3.24) implies $\mathcal{S}$ has density 0, so that $M(\mathcal{P})$ has density 1, in accordance with the claim of the corollary. To treat the latter case, we begin with the observation that

$$S(x) - S(\mathcal{A}, \mathcal{P}, z) \ll |\{n \leq x : \text{there exists } p \in \mathcal{P}, z < p \leq x, p \mid n\}|. \tag{3.25}$$

The claim will follow from (3.24) as soon as we show the second term on the right hand side of (3.25) is $o(x)$. But this term is bounded above by

$$\sum_{\substack{p \in \mathcal{P} \\ z < p \leq x}} \frac{x}{p} \leq x \sum_{\substack{p \in \mathcal{P} \\ p > z}} \frac{1}{p} = o(x);$$

the final estimate here derives from the convergence of $\sum_{p \in \mathcal{P}} 1/p$, which in turn follows from the nonvanishing of C. $\square$

3.3.3 Three Number-Theoretic Applications

We now turn our attention to three problems in number theory which can be attacked by the simple methods we have developed thus far. None of the results we prove are the best of their kind, but the proofs are simple and the statements fairly striking.

Theorem 3.3.4. *The following sets have density zero:*

i. the set of integers $n > 1$ for which

$$4/n = 1/x + 1/y + 1/z \tag{3.26}$$

has no solution in positive integers x, y, z,

ii. the set of integers expressible as a sum of two squares,

iii. the set of odd perfect numbers.

It is a well-known conjecture that the exceptional set described in part iii) is empty. The same conjecture for the set described in i) is ascribed to Erdős & Strauss. We will deduce Theorem 3.3.4 from the following lemma, which in turn is a consequence of Theorem 3.3.3:

Lemma 3.3.5. *The set of positive integers divisible by no prime $p \equiv 3 \pmod 4$ has density 0.*

Proof. We know from either the general results of Chapter 2 or from Chapter 1, Exercise 1.9.3 that

$$\sum_{\substack{p \leq x \\ p \equiv 3 \pmod 4}} \frac{\log p}{p} = \frac{1}{2} \log x + O(1).$$

By partial summation (the proof is essentially the same as when deriving the estimate for $\sum_{p\le x} 1/p$ given in §5 of Chapter 1),

$$\sum_{\substack{p\le x\\ p\equiv 3 \pmod 4}} \frac{1}{p} = \frac{1}{2}\log\log x + C + O\left(\frac{1}{\log x}\right). \tag{3.27}$$

In particular, the sum diverges as $x \to \infty$. Hence $\prod_{p\equiv 3 \pmod 4}(1-1/p)$ diverges to 0, and the result follows from Corollary 3.3.3. ☐

Proof of Theorem 3.3.4, Part I. It suffices to show that (3.26) is solvable if n possesses a prime divisor $p = 4k-1 \equiv 3 \pmod 4$. In this case write $n = (4k-1)q$. Then the result follows from the algebraic identity

$$\frac{4}{q(4k-1)} = \frac{1}{2qk} + \frac{1}{2qk} + \frac{1}{q(4k^2-k)}.$$

Actually, this shows that $4/n$ can almost always be written as a sum of *two* unit fractions, since we may write $1/(2qk) + 1/(2qk) = 1/qk$, ☐

Exercise 3.3.1. Prove that there are infinitely many positive integers n for which $4/n = 1/x + 1/y$ is not solvable in positive integers x, y.

Exercise 3.3.2. Prove that (3.26) is solvable for every positive integer n in (not necessarily positive) integers x, y, z.

Hint: Use the identity $4/(4k-1) = 1/(2k)+1/(2k)+1/(4k^2-k)$, implicit in the above proof, as well as an analogous identity with $4k-1$ replaced by $4k+1$.

Schinzel has shown [Sch00] that

$$\frac{1}{aT+b} = \frac{4}{A(T)} + \frac{4}{B(T)} + \frac{4}{C(T)}$$

is solvable in polynomials $A(T), B(T), C(T) \in \mathbf{Z}[T]$ with positive leading coefficients only if b is a not a quadratic residue $\pmod a$. Thus it is not possible to prove the original $4/n$ conjecture by the method of this exercise. The best known upper estimate for the exceptional set in the $4/n$ problem is due to Vaughan [Vau70], who showed that the number of $n \le x$ for which (3.26) is not solvable is

$$\ll x\exp(-C\log^{2/3} x)$$

for a positive constant C. For the proof of a more general result (with $4/n$ replaced by a/n), see [Nar86, §1.4].

Proof of Theorem 3.3.4, Part II. Let $R(x)$ denote the counting function of the set of of sums of two squares. A positive integer can be written as a sum of two *coprime* squares precisely when it is divisible by neither 4 nor any prime $p \equiv 3 \pmod 4$. If $A(x)$ denotes the counting function of such numbers, then

$$R(x) \le A(x) + A(x/1^2) + A(x/2^2) + \dots. \tag{3.28}$$

Lemma 3.3.5 implies $A(x) = o(x)$. Now given $\epsilon > 0$, choose a positive integer N such that $A(x) < \epsilon x/4$ for $x > N$. Thinking of x as large, break the sum on the right hand side of (3.28) into two parts according as $x/k^2 > N$ or $x/k^2 \leq N$. The first of the two resulting sums is bounded by

$$\sum_{k=1}^{\infty} \epsilon \frac{x/k^2}{4} = \frac{\epsilon}{4}\zeta(2)x < \frac{\epsilon}{2}x.$$

Every term in the second sum is bounded by $A(N)$, and there are no more than $\sqrt{x}$ nonzero terms. Thus,

$$R(x) \leq \epsilon x/2 + A(N)\sqrt{x} < \epsilon x$$

for large x. As $\epsilon > 0$ was arbitrary, the result follows. $\square$

A theorem of Landau (independently discovered by Ramanujan) describes the precise asymptotic behavior of $R(x)$ (for the proof, see [LeV02, vol. II, §7-5]):

$$R(x) = \frac{1}{\sqrt{2}} \left(\prod_{p \equiv 3 \pmod 4} \left(1 - \frac{1}{p^2}\right) \right)^{-1/2} \frac{x}{\sqrt{\log x}} + O\left(\frac{x}{\log^{3/2} x}\right).$$

Proof of Theorem 3.3.4, Part III. We prove that every odd perfect number n is of the form pa^2, where p is a prime with $p \equiv 1 \pmod 4$. Since such integers are sums of two squares, the result follows from part ii).

Let $n = p_1^{e_1} \dots p_k^{e_k}$ be the canonical factorization of n into primes, so that

$$2n = \sigma(n) = \prod_{i=1}^{k} \left(1 + p_i + p_i^2 + \cdots + p_i^{e_i}\right).$$

Because $2 \| 2n$, exactly one of the factors on the right hand side is divisible by 2, and that factor is divisible only by 2^1. Looking mod 2, we see that exactly one of the e_i is odd, and that for this i,

$$1 + p_i + p_i^2 + \cdots + p_i^{e_i} \equiv 2 \pmod 4.$$

If $p_i \equiv 3 \pmod 4$, then the left hand side is congruent to $(1+3)+\cdots+(1+3) \equiv 0 \pmod 4$, a contradiction. Thus $p_i \equiv 1 \pmod 4$, and $n = p_i a^2$, where

$$a = p_i^{(e_i-1)/2} \prod_{1 \leq j \neq i \leq k} p_j^{e_j/2}. \qquad \square$$

The best upper bound on the number of perfect numbers up to x is due to Wirsing [Wir59]. Improving on earlier joint work with Hornfeck, he established the upper estimate $O(x^{c/\log\log x})$, where c is an absolute positive constant and the estimate holds for, say, $x \geq 3$. In particular, the number of odd perfect numbers not exceeding x is $O(x^{\epsilon})$ for every $\epsilon > 0$.

3.3.4 Exercises

Exercise 3.3.3. Show that

$$\sum_{d|n, d|P} \mu(d) = \begin{cases} 1 & \text{if } \gcd(n, P) = 1, \\ 0 & \text{otherwise.} \end{cases}$$

Deduce that

$$S(\mathcal{A}, \mathcal{P}) = \sum_{a \in \mathcal{A}} \sum_{d|n, d|P} \mu(d).$$

Obtain another proof of Theorem 3.3.2 by reversing the order of summation.

Exercise 3.3.4. Let $\mathcal{A} := \{n(n+2)\}_{1 \le n \le x}$ and let $\mathcal{P}$ be the set of all primes. Show that if $z = z(x) \to \infty$ as $x \to \infty$ while $z(x) \le \frac{1}{2} \log x$ for all large x, then

$$S(\mathcal{A}, \mathcal{P}, z) \sim 2C_2 e^{-2\gamma} x / \log^2 z \quad (x \to \infty).$$

Here

$$C_2 := \prod\nolimits_{p>2} \left(1 - (p-1)^{-2}\right) \approx 1.3202\ldots \tag{3.29}$$

is the twin-prime constant.

Exercise 3.3.5. The method used to prove Corollary 3.3.3 may be profitably adapted to the study of the distribution of squarefree numbers. Illustrate this by proving the following three theorems:

a) The number of squarefree $n \le x$ is asymptotic to $x/\zeta(2)(= 6x/\pi^2)$ as $x \to \infty$.

b) The number of pairs of squarefree integers $n, n+2$ with $1 \le n \le x$ is asymptotic to $x \prod_p (1 - 2/p^2)$ as $x \to \infty$.

c) The number of representations of the positive integer n as a sum of two positive squarefree integers is asymptotic to

$$n \prod_p \left(1 - \frac{2}{p^2}\right) \prod_{p^2|n} \frac{p^2 - 1}{p^2 - 2} \quad (n \to \infty).$$

Exercise 3.3.6. Show that n and $\phi(n)$ have a nontrivial common divisor for almost every n.

Exercise 3.3.7. Let $\mathcal{S} = \{s_i\}$ be a sequence of integers, and let $M(\mathcal{S})$ denote the "set of multiples" of S, i.e., the set of integers n divisible by some $s \in \mathcal{S}$. When $\mathcal{S}$ is a sequence of distinct primes, Corollary 3.3.3 tells us the density of $M(\mathcal{S})$.

The existence of this density is not obvious a priori. In this exercise, we show the density of $M(\mathcal{S})$ exists under the hypothesis that $\sum s^{-1}$ converges. We also recover the result of Corollary 3.3.3 in the case when $\sum_{p \in \mathcal{P}} 1/p$ converges.

a) Show this density exists if $S = \{s_i\}_{1 \le i \le n}$ is a finite sequence. [Hint: Show that whether $n \in M(\mathcal{S})$ is determined by a congruence condition on n (mod $\prod_{s \in S} s$).] Thus we may assume S is infinite.

b) For each j, define $\mathcal{D}_j$ as the set of positive integers divisible by s_j but not by s_i for any $i < j$, and let D_j denote the density of $\mathcal{D}_j$. Use the method of part a) to show D_j always exists.

c) Show that $D_j \le 1/s_j$; thus $\sum_{j=1}^{\infty} D_j$ converges. We will show this converges to $d(M(\mathcal{S}))$, which is what one expects owing to the relation $M(\mathcal{S}) = \dot\cup_{j=1}^{\infty} \mathcal{D}_j$.

d) Show that for each n, $\underline{d}(M(\mathcal{S})) \ge \sum_{j=1}^{n} D_j$. Thus $\underline{d}(M(\mathcal{S})) \ge \sum_{j=1}^{\infty} D_j$.

e) Show that
$$\lim_{j \to \infty} \overline{d}(M(\mathcal{S}) \setminus \cup_{j=1}^{n} \mathcal{D}_j) = 0.$$

f) Use the previous two parts to deduce $d(M(\mathcal{S})) = \sum_{j=1}^{\infty} D_j$.

g) Recover the result of Corollary 3.3.3 in the case when $\sum_{p \in \mathcal{P}} 1/p$ converges.

A set S for which $M(\mathcal{S})$ possesses a natural density is called a *Besicovitch set*, in honor of A.S. Besicovitch, who gave the first example [Bes31] of a set *without* this property.

Exercise 3.3.8. Using the Sieve of Eratosthenes-Legendre and the estimate (3.27), show that there are $\ll x/\sqrt{\log\log x}$ positive integers not exceeding x with no prime divisor $p \equiv 3 \pmod 4$. What implications does this have for the exceptional sets described in Theorem 3.3.4?

3.4 Brun's Simple Pure Sieve

Our heuristic derivation of Legendre's formula for $\pi(x)$ proceeded by successive approximation: we began by taking the total number of positive integers not exceeding x, thought of as a 0th approximation, then subtracted those divisible by any single prime $p \le \sqrt{x}$, then added back those divisible by two such, etc. We noticed that every even step seemed to produce an overestimate, while every odd step produced an underestimate. This observation, suitably generalized, forms the heart of Brun's pure sieve.

3.4.1 Preparation

To prove the appropriate generalization, we first need a technical lemma on alternating sums of symmetric functions.

Recall that if $a_1, \ldots, a_n$ is a (possibly empty) sequence of $n \ge 0$ elements belonging to a commutative ring, we define (for $k \ge 0$) the kth elementary symmetric function $\sigma_k(a_1, \ldots, a_n)$ as the sum of all possible $\binom{n}{k}$ products of the a_i taken k at a time. We adopt the usual conventions about empty sums and products, so that when $n = 0$, $\sigma_0 = 1$ and $\sigma_k = 0$ for $k > 0$. To take a less pathological example, when $n = 2$, one has

$$\sigma_0(a_1, a_2) = 1, \quad \sigma_1(a_1, a_2) = a_1 + a_2, \quad \sigma_2(a_1, a_2) = a_1 a_2,$$

and $\sigma_k(a_1, a_2) = 0$ for $k > 2$.

The following lemma on alternating sums of symmetric functions appears in [Hoo94] (though is presumably much older):

Lemma 3.4.1. *Let $a_1, \ldots, a_n$ be a finite (possibly empty) sequence of real numbers from $[0, 1]$. Then*

$$\sum_{k=0}^{m} (-1)^k \sigma_k(a_1, \ldots, a_n) - \prod_{j=1}^{n} (1 - a_j) \tag{3.30}$$

is nonnegative or nonpositive according as m is even or odd, respectively.

Remark. When $m = n$, the sum appearing above is exactly the formal expansion of the subtracted product; consequently, equality holds in (3.30) for $k \geq m$.

Proof. We induct on the length n of the sequence. When $n = 0$, the product

$$P := \prod_{i=1}^{n} (1 - a_i)$$

appearing in (3.30) is empty so takes the value 1, while

$$\sum_{k=0}^{m} (-1)^k \sigma_k = 1 - 0 + 0 - \cdots \pm 0 = 1.$$

Hence (3.30) is 0 for every m, and the result follows in this case.

Now assume the result holds for each sequence of n real numbers in $[0, 1]$ (and each m) and assume $a_1, \ldots, a_{n+1}$ is a given sequence of $n+1$ real numbers from the same interval. By the induction hypothesis, it suffices to prove

$$\left(\sum_{k=0}^{m} (-1)^k \sigma_k(a_1, \ldots, a_{n+1}) - \prod_{i=1}^{n+1} (1 - a_i) \right) \\ - \left(\sum_{k=0}^{m} (-1)^k \sigma_k(a_1, \ldots, a_n) - \prod_{i=1}^{n} (1 - a_i) \right) \tag{3.31}$$

is nonnegative or nonpositive according as whether m is even or odd respectively. This is true for $m = 0$, since in this case the expression (3.31) simplifies to

$$\prod_{i=1}^{n} (1 - a_i) - \prod_{i=1}^{n+1} (1 - a_i) = a_{n+1} \prod_{i=1}^{n} (1 - a_i),$$

which is nonnegative as $a_1, \ldots, a_{n+1} \in [0, 1]$. When $m > 0$, we recognize (3.31) as

$$\begin{aligned}\sum_{k=1}^{m}(-1)^k \left(\sigma_k(a_1, \ldots, a_{n+1}) - \sigma_k(a_1, \ldots, a_n)\right) + Pa_{n+1}\\ &= \sum_{k=1}^{m}(-1)^k a_{n+1}\sigma_{k-1}(a_1, \ldots, a_n) + Pa_{n+1}\\ &= a_{n+1}\left(P - \sum_{k=0}^{r-1}(-1)^k \sigma_k(a_1, \ldots, a_n)\right).\end{aligned}$$

Since $a_{n+1} \geq 0$, the result follows from the induction hypothesis. $\square$

An important special case occurs when n is a positive integer and $a_1 = a_2 = \cdots = a_n = 1$. Then $\prod_{i=1}^{n}(1-a_i) = (1-1)^n = 0$, while $\sigma_k(1, \ldots, 1) = \binom{n}{k}$. Thus Lemma 3.4.1 has the following consequence:

Lemma 3.4.2. *Let n be a positive integer. Then the alternating sum*

$$\sum_{k=0}^{m}(-1)^k \binom{n}{k}$$

is nonnegative or nonpositive according as m is even or odd.

Remark. For the applications of this section, we only need Lemma 3.4.2. Thus it is of interest to note that Lemma 3.4.2 admits a simple proof independent of Lemma 3.4.1. Indeed, by induction on m, one can easily prove the identity

$$\sum_{k=0}^{m}(-1)^k \binom{n}{k} = (-1)^m \binom{n-1}{m}. \tag{3.32}$$

Alternately, (3.32) follows by comparing the coefficient of x^m in both sides of the formal power series identity $(1-x)^{n-1} = (1-x)^{-1}(1-x)^n$. See also Exercise 3.4.1.

We now use Lemma 3.4.2 to establish the following variant of the Principle of Inclusion-Exclusion:

Theorem 3.4.3 (Bonferroni Inequalities). *Let X be a nonempty, finite set of N objects, and let $P_1, \ldots, P_r$ be properties elements of X may have. For any subset $I \subset \{1, 2, \ldots, r\}$, let $N(I)$ denote the number of elements of X that have each of the properties indexed by the elements of I. Let N_0 denote the number of elements of X with none of these properties. Then if m is a nonnegative even integer,*

$$N_0 \leq \sum_{k=0}^{m}(-1)^k \sum_{\substack{I\subset\{1,2,\ldots,r\}\\|I|=k}} N(I), \tag{3.33}$$

while if m is a nonnegative odd integer,

$$N_0 \geq \sum_{k=0}^{m} (-1)^k \sum_{\substack{I \subset \{1,2,\ldots,r\} \\ |I|=k}} N(I). \tag{3.34}$$

Proof. Suppose that $x \in X$ has exactly l of the properties $P_1, \ldots, P_r$. If $l = 0$, then x is counted once by both N_0 and the right hand sum above (corresponding to when $I = \emptyset$). If $l \geq 1$, then x is not counted at all by N_0, and is counted with weight

$$\sum_{k=0}^{m} (-1)^k \binom{l}{k} \begin{cases} \geq 0 & \text{if } m \text{ is even,} \\ \leq 0 & \text{otherwise,} \end{cases}$$

by Lemma 3.4.2. Summing up the contributions over $x \in X$ gives the result. □

3.4.2 A Working Version

Corollary 3.4.4 (Brun's Simple Pure Sieve, general form). *With the notation of §3.2, we have for every nonnegative even integer m,*

$$\sum_{d|P, \nu(d) \leq m-1} \mu(d) A_d \leq S(\mathcal{A}, \mathcal{P}) \leq \sum_{d|P, \nu(d) \leq m} \mu(d) A_d.$$

Proof. As in the proof of Theorem 3.3.2, let $p_1, \ldots, p_r$ be a list of the primes $p \in \mathcal{P}$, and let P_i be the property of being divisible by p_i. We aim to estimate the number $S(\mathcal{A}, \mathcal{P})$ of elements of $\mathcal{A}$ possessing none of the P_i. The right hand inequality for $S(\mathcal{A}, \mathcal{P})$ follows easily from (3.33). If $m = 0$, then the left hand inequality is trivial, while if $m > 0$ then $m - 1$ is a nonnegative odd integer, and the left hand inequality follows from (3.34). □

We can make this more useful, at the cost of sacrificing generality, by substituting $A_d = X\alpha(d) + r(d)$ and estimating the resulting terms. With a bit of manipulation, we arrive at the following:

Theorem 3.4.5 (Brun's Simple Pure Sieve). *For every even integer $m \geq 0$,*

$$S(\mathcal{A}, \mathcal{P}) = X \prod_{p \in \mathcal{P}} (1 - \alpha(p)) + O\Bigg(\sum_{d|P, \nu(d) \leq m} |r(d)| \Bigg) + O\Bigg(X \sum_{d|P, \nu(d) \geq m} \alpha(d) \Bigg).$$

Here the implied constants are absolute.

Proof. From Corollary 3.4.4,

$$\begin{aligned} S(\mathcal{A}, \mathcal{P}) &= \sum_{d|P, \nu(d) \leq m} \mu(d) A_d + O\Bigg(\sum_{d|P, \nu(d) = m} A_d \Bigg) \\ &= \sum_{d|P, \nu(d) \leq m} \mu(d)(X\alpha(d) + r(d)) + O\Bigg(\sum_{d|P, \nu(d) = m} A_d \Bigg) \\ &= X \sum_{d|P, \nu(d) \leq m} \mu(d)\alpha(d) + O\Bigg(\sum_{d|P, \nu(d) \leq m} |r(d)| \Bigg) + O\Bigg(\sum_{d|P, \nu(d) = m} A_d \Bigg). \end{aligned}$$

Writing $A_d = X\alpha(d) + r(d)$, we see the last of these error terms is

$$\ll X \sum_{d|P,\nu(d)=m} \alpha(d) + \sum_{d|P,\nu(d)=m} |r(d)|;$$

hence,

$$S(\mathcal{A},\mathcal{P}) \\ = X \sum_{d|P,\nu(d)\le m} \mu(d)\alpha(d) + O\Bigg(\sum_{d|P,\nu(d)\le m} |r(d)|\Bigg) + O\Bigg(X \sum_{d|P,\nu(d)=m} \alpha(d)\Bigg). \tag{3.35}$$

In order to handle the main term appearing here, we add back in the terms of the sum corresponding to divisors d of P with $\nu(d) > m$; we can then estimate the main term as $X\prod_{p\in\mathcal{P}}(1-\alpha(p))$, but at the cost of introducing an error which is

$$\ll X \sum_{d|P,\nu(d)>m} \alpha(d).$$

This error can be combined with the last error term appearing in (3.35) to yield

$$S(\mathcal{A},\mathcal{P}) = X\prod_{p\in\mathcal{P}}(1-\alpha(p)) + O\Bigg(\sum_{d|P,\nu(d)\le m} |r(d)|\Bigg) + O\Bigg(X \sum_{d|P,\nu(d)\ge m} \alpha(d)\Bigg),$$

exactly as the theorem asserts. □

3.4.3 Application to the Twin Prime Problem (outline)

The most famous application of Brun's pure sieve is Brun's own 1919 contribution [Bru19a] to the twin prime problem:

Theorem 3.4.6. *As $x \to \infty$,*

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

The upper estimate differs from what we expect to be true by the presence of the $\log\log x$ factor. We shall later remedy this defect. Nevertheless, it is worth noting that the estimate of Theorem 3.4.6 is already sharp enough to imply the following striking result:

Corollary 3.4.7. *If there are infinitely many primes p such that $p+2$ is also prime, then the sum*

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

taken over all such primes, converges.

Proof. By Theorem 3.4.8, $\pi_2(x) \ll x/\log^{3/2} x$ as $x \to \infty$. It follows that the same estimate holds, with perhaps a different implied constant, in the range $x \geq 3$. Letting p_n denote the nth prime p for which $p+2$ is also prime, we see that for $n \geq 1$,

$$n = \pi_2(p_n) \ll p_n/\log^{3/2} p_n,$$

so that

$$p_n \gg n\log^{3/2} p_n \geq \frac{1}{2}(n+1)\log^{3/2}(n+1).$$

The comparison and integral tests together now imply that $\sum_{n=1}^{\infty} p_n^{-1}$ converges, which is the assertion of the corollary. □

We now prove Theorem 3.4.6 as an easy consequence of the following estimate, which is more directly amenable to an approach via sieve methods:

Theorem 3.4.8. *Define*

$$\pi_2(x,z) := |\{n \leq x : p \mid n(n+2) \Longrightarrow p > z\}|.$$

Suppose $z = z(x) \to \infty$ *as* $x \to \infty$ *while* $z(x) \leq x^{1/20\log\log x}$ *for all large* x. *Then*

$$\pi_2(x,z) \sim 2C_2 e^{-2\gamma} x/\log^2 z \quad (x \to \infty).$$

Deduction of Theorem 3.4.6. For any choice of the parameter z,

$$\pi_2(x) \leq z + \pi_2(x,z);$$

indeed, if p and $p+2$ are both prime, then either $p \leq z$ or both $p, p+2$ have no prime factors not exceeding z. Now take $z = z(x) = x^{1/20\log\log x}$. Then Theorem 3.4.8 implies that as $x \to \infty$,

$$\pi_2(x) \ll x^{1/20\log\log x} + \frac{x}{\log^2 x}(\log\log x)^2 \ll \frac{x}{\log^2 x}(\log\log x)^2. \qquad \square$$

3.4.4 Proof of Theorem 3.4.8

Recall the appropriate set of sieving parameters for the twin prime problem:

$$\mathcal{A} := \{a_n = n(n+2), n \leq x\}, \quad \mathcal{P} := \{\text{all primes } p\}.$$

We aim to estimate $S(\mathcal{A}, \mathcal{P}, z)$, which is the $\pi_2(x,z)$ of Theorem 3.4.8.

For our approximation X to the size of $\mathcal{A}$, we take $X = x$. For our approximation α to the probability an element of $\mathcal{A}$ is divisible by d, we take $\alpha(d) = \omega(d)/d$, where ω is defined by

$$\omega(N) := |\{a \in \mathbf{Z}/N\mathbf{Z} : a(a+2) = 0\}|.$$

(The multiplicativity of ω, and consequently α, comes from the Chinese Remainder Theorem.) In order to estimate the remainder terms

$$\begin{aligned} r(d) &:= A_d - X\alpha(d) \\ &= |\{n \leq x : n(n+2) \equiv 0 \pmod{d}\}| - x\omega(d)/d \qquad (d \mid P), \end{aligned} \tag{3.36}$$

we need the following lemma:

Lemma 3.4.9. *Let $a_1, \dots, a_k$ be k distinct residue classes $\pmod{d}$, where d is a positive integer. Then if x is any positive real number, the number of positive integers not exceeding x falling into any of the given residue classes is $kx/d + \theta$, where $|\theta| \le k$.*

Proof. Each block of d consecutive positive integers not exceeding x contains k integers falling into the given congruence classes, and there are between $\lfloor x/d \rfloor$ and $\lceil x/d \rceil$ such blocks. □

Lemma 3.4.9 applied to (3.36) implies the remainder terms $r(d)$ satisfy

$$|r(d)| \le \omega(d) = \prod\nolimits_{p|d} \omega(p) \le 2^{\nu(d)} \qquad (d \mid P).$$

Substituting the values of our sieving parameters into Theorem 3.4.5 shows

$$\pi_2(x,z) = \\ x \prod_{p \le z} (1 - \alpha(p)) + O\Bigg(\sum_{d|P, \nu(d) \le m} 2^{\nu(d)} \Bigg) + O\Bigg(x \sum_{d|P,\nu(d) \ge m} \alpha(d) \Bigg), \quad (3.37)$$

for any choice of the nonnegative even integer m. We now think of x as large, and we set

$$m := 10 \lfloor \log\log z \rfloor .$$

Note that as x gets large, so does z, and hence so does m.

By (3.37), to prove Theorem 3.4.8 it will suffice to establish that the following three estimates hold with this choice of m:

i. As $x \to \infty$ (so that $z \to \infty$ as well),
$$x \prod\nolimits_{p \le z} (1 - \alpha(p)) \sim 2C_2 e^{-2\gamma} \frac{x}{\log^2 z}.$$

ii. For all large x,
$$E_1 := \sum_{d|P,\nu(d) \le m} 2^{\nu(d)}$$
satisfies $E_1 \le 2x^{1/2} = o(x/\log^2 z)$.

iii. As $x \to \infty$, we have
$$E_2 := x \sum_{d|P,\nu(d) \ge m} \alpha(d) \ll x/\log^5 z = o(x/\log^2 z).$$

Proof of I). As $x \to \infty$ (so that $z \to \infty$ as well), we see using using the definition (3.14) of C_2 and Mertens' theorem that

$$x \prod_{p \le z} (1 - \alpha(p)) = \frac{1}{2} x \prod_{2 < p \le z} (1 - 2/p) \quad (3.38)$$
$$= x \left(2 \prod_{2 < p \le z} \frac{1 - 2/p}{(1 - 1/p)^2} \right) \prod_{p \le z} (1 - 1/p)^2 \sim 2C_2 e^{-2\gamma} \frac{x}{\log^2 z}. \; \square$$

Proof of II). For large x,

$$\begin{aligned} E_1 = \sum_{d|P, \nu(d)\le m} 2^{\nu(d)} &= \sum_{k=0}^{m} 2^k \binom{\pi(z)}{k} \le \sum_{k=0}^{m} (2\pi(z))^k \\ &\le \sum_{k=-\infty}^{m} (2\pi(z))^k = (2\pi(z))^m \frac{1}{1-\frac{1}{2\pi(z)}} \\ &\le 2(2\pi(z))^m \le 2z^m, \end{aligned}$$

since $\pi(z) \le z/2$ once x is large. Thus

$$E_1 \le 2z^{10\log\log z} \le 2z^{10\log\log x} \le 2x^{1/2}.$$

This is certainly $o(x/\log^2 z)$, since since the trivial inequality $z \le x$ implies

$$\frac{x^{1/2}}{x/\log^2 z} \le \frac{x^{1/2}}{x/\log^2 x} = \frac{\log^2 x}{x^{1/2}} \to 0. \qquad \square$$

Proof of III). We begin by rewriting E_2 in the form

$$E_2 = x \sum_{k\ge m} \sum_{\substack{d|P \\ \nu(d)=k}} \alpha(d).$$

The inner sum can be rewritten as

$$\sum_{\substack{d|P \\ \nu(d)=k}} \alpha(d) = \sum_{p_1<p_2<\cdots<p_k\le z} \alpha(p_1)\alpha(p_2)\ldots\alpha(p_k) \le \frac{1}{k!}\left(\sum_{p\le z}\alpha(p)\right)^k,$$

because in the "multinomial expansion" of the kth power on the right hand side, each term $\alpha(p_1)\cdots\alpha(p_k)$ appears with coefficient $k!$. We estimate $\sum_{p\le z}\alpha(p)$ by recalling (cf. (1.40)) that for large z,

$$\sum_{p\le z}\frac{1}{p} \le \log\log z + c,$$

where c is an absolute constant. Since $\alpha(p) \le 2/p$ for every p,

$$\sum_{k\ge m}\frac{1}{k!}\left(\sum_{p\le z}\alpha(p)\right)^k \le \sum_{k\ge m}\frac{1}{k!}(2\log\log z + 2c)^k. \tag{3.39}$$

The ratio of the $(k+1)$th term in this series to the kth is given by

$$\frac{2\log\log z + 2c}{k+1} \le \frac{2\log\log z + 2c}{10\lfloor\log\log z\rfloor + 1} \le 1/2,$$

for large enough z, hence for large enough x. Consequently, for such x the right hand sum of (3.39) is bounded above by twice its first term. Because

$$e^m = 1 + m + m^2/2! + m^3/3! + \cdots \geq m^m/m!,$$

we have $m! \geq (m/e)^m$, so that

$$\sum_{k \geq m} \frac{1}{k!} \left(\sum_{p \leq z} \alpha(p) \right)^k \leq 2 \left(\frac{2e \log\log z + 2ce}{m} \right)^m.$$

Since $m = 10 \lfloor \log\log z \rfloor$, the parenthesized expression on the right is eventually smaller than any constant exceeding $2e/10$; in particular, it is eventually smaller than $3/5$. It follows that for large x,

$$\begin{aligned} E_2 \leq 2x(3/5)^m &= 2x(3/5)^{10\lfloor \log\log z \rfloor} \\ &\leq 2(5/3)^{10} x (3/5)^{10 \log\log z} \ll x/\log^5 z, \end{aligned}$$

since $10 \log(3/5) < -5$. Thus $E_2 = o(x/\log^2 z)$ as well. □

Remark. In our proof of Theorem 3.4.8, we needed two results of Mertens, namely

$$\sum_{p \leq x} \frac{1}{p} \leq \log\log x + O(1), \qquad \prod_{p \leq x} \left(1 - \frac{1}{p}\right) \sim \frac{e^{-\gamma}}{\log x}.$$

We leave it as an exercise to show that if we are only interested in the upper bound for $\pi_2(x)$ given by Theorem 3.4.6, then the following weaker estimates suffice:

$$\sum_{p \leq x} \frac{1}{p} \ll \log\log x, \quad \prod_{p \leq x} \left(1 - \frac{1}{p}\right) \leq \frac{1}{\log x} \qquad (x \geq x_0).$$

The second of these has already been given a simple proof (cf. (3.18)) independent of Chebyshev's results. For a simple, direct proof of the first, see Exercise 3.4.5.

3.4.5 Exercises

Exercise 3.4.1. Show that the binomial coefficients $\binom{n}{k}$ increase up to the middle term (or pair of middle terms, if n is odd) and then decrease. Use this to give another proof of Lemma 3.4.2.

Suggestion: To handle the case of Lemma 3.4.2 when $m > \lfloor n/2 \rfloor$, use the symmetry of the binomial coefficients and the relation $\sum_{k=0}^{n} (-1)^k \binom{n}{k} = 0$ (for $n \geq 1$).

Exercise 3.4.2. For N a nonzero even integer, define

$$\pi_N(x) := |\{p \leq x : p, p + N \text{ are both prime}\}|.$$

Show that for each such N,

$$\pi_N(x) \ll_N \frac{x}{\log^2 x}(\log\log x)^2$$

for $x \geq 3$. Deduce that for any fixed M, the number of primes up to x within M of another prime is $O_M(x(\log\log x)^2/\log^2 x)$ as $x \to \infty$.

Exercise 3.4.3. Show that the number of $p \leq x$ with $p, p+2, p+6$ all prime is

$$\ll \frac{x}{\log^3 x}(\log\log x)^3.$$

[Hint: Imitate the proof of Theorem 3.4.8 with $z = z(x)$ tending to infinity in the range $z(x) \leq x^{1/30\log\log x}$, and m modified appropriately.] Can you prove a general theorem on simultaneous prime values of k linear polynomials?

Exercise 3.4.4. Extend the result of Theorem 3.4.8 by showing that

$$\pi_2(x,z) \sim 2C_2 e^{-2\gamma}\frac{x}{\log^2 z} \quad (x \to \infty),$$

uniformly for $2 \leq z \leq x^{1/20\log\log x}$. Hint: Modify the proof of Theorem 3.4.8 to show this is true in the range $\log\log x \leq z \leq x^{1/20\log\log x}$, and for the remaining range estimate $\pi_2(x,z)$ using the sieve of Eratosthenes-Legendre.

Exercise 3.4.5 *(Brun [Bru17]).* For $x \geq 2$, define $N = N(x)$ as the number of positive $n \leq x$ with a prime divisor p satisfying $\sqrt{x} < p \leq x$.

a) Noting that any $n \leq x$ has at most one such prime divisor p, show that $N \geq \sum_{\sqrt{x}<p\leq x} \lfloor x/p \rfloor$.

b) Conclude from the trivial upper bound $N \leq x$ that $\sum_{\sqrt{x}<p\leq x} 1/p \leq 2$.

c) Show that if M is the smallest integer with $x^{1/2^M} < 2$, then $\sum_{p\leq x} 1/p \leq 2M$. Deduce that $\sum_{p\leq x} 1/p \leq 2(\log_2\log_2 x + 1)$.

3.5 The Brun-Hooley Sieve

3.5.1 The Sifting Function Perspective

Before presenting the Brun-Hooley sieve in the succeeding sections, it is helpful to revisit the preceding results from a slightly different perspective. Keeping the notation of §3.2, we introduce the *sifting function*

$$s(n) := \begin{cases} 1 & \text{if } \gcd(n,P) = 1, \\ 0 & \text{otherwise.} \end{cases} \tag{3.40}$$

Then

$$S(\mathcal{A},\mathcal{P}) = \sum_{a\in A} s(a). \tag{3.41}$$

The Sieve of Eratosthenes is recovered by noticing that the fundamental property of the Möbius function implies a nontrivial representation for $s(n)$, namely

$$s(n) = \sum_{d|n, d|P} \mu(d). \tag{3.42}$$

Substituting this expression into (3.41) and interchanging the order of summation, we easily arrive at Theorem 3.3.2. Proceeding similarly, we could rederive Brun's method from the following lemma:

Lemma 3.5.1. *Let n be an integer. The expression*

$$\sum_{\substack{d|n, d|P \\ \nu(d) \leq m}} \mu(d) - \sum_{d|n, d|P} \mu(d) \tag{3.43}$$

is nonnegative or nonpositive according as the nonnegative integer m is even or odd.

Not surprisingly, the proof of Lemma 3.5.1 is essentially the one we have already given of the Bonferroni inequalities of the last section. Namely, if we suppose that n is divisible by exactly l primes $p \in \mathcal{P}$, then by Lemma 3.4.1,

$$\sum_{\substack{d|n, d|P \\ \nu(d) \leq m}} \mu(d) = \sum_{k=0}^{m} (-1)^k \binom{l}{k} \begin{cases} = 1 & \text{if } l = 0 \text{ (i.e., if } \gcd(n, P) = 1), \\ \geq 0 & \text{if } l \geq 1, m \text{ even}, \\ \leq 0 & \text{if } l \geq 1, m \text{ odd}, \end{cases}$$

and the result is readily deduced from (3.40) and (3.42).

We note the following consequence of Lemma 3.5.1 for later use:

Lemma 3.5.2. *If n, m are integers with m nonnegative and even, then*

$$0 \leq \sum_{\substack{d|n, d|P \\ \nu(d) \leq m}} \mu(d) - \sum_{d|n, d|P} \mu(d) \leq \sum_{\substack{d|n, d|P \\ \nu(d) = m+1}} 1.$$

Exercise 3.5.1. Prove the following identity, valid also for odd m, from which the truth of Lemma 3.5.2 is immediate: With $p^-(d)$ denoting the smallest prime divisor of d,

$$\sum_{\substack{d|n, d|P \\ \nu(d) \leq m}} \mu(d) - \sum_{d|n, d|P} \mu(d) = (-1)^m \sum_{\substack{d|n, d|P \\ \nu(d) = m+1 \\ p^-(d) = p^-(\gcd(n, P))}} 1.$$

3.5.2 The Upper Bound

The Brun-Hooley method takes two forms, depending on whether we are after upper or lower bounds. Here we describe the simpler upper bound method. We suppose the sifting set $\mathcal{P}$ to be partitioned into r disjoint sets,

$$\mathcal{P} = \dot{\bigcup}_{j=1}^{r} \mathcal{P}_i.$$

Then n is divisible by no prime $p \in \mathcal{P}$ precisely when n is divisible by no prime $p \in \mathcal{P}_j$ for every $1 \leq j \leq r$. Consequently, setting

$$P_j := \prod_{p \in \mathcal{P}_j} p,$$

and invoking Lemma 3.5.1 (with $\mathcal{P}_j, P_j$ in place of $\mathcal{P}, P$) we see that

$$\begin{aligned} s(n) = \sum_{d \mid n, d \mid P} \mu(d) &= \prod_{j=1}^{r} \sum_{d_j \mid n, d_j \mid P_j} \mu(d) \\ &\leq \prod_{j=1}^{r} \sum_{\substack{d_j \mid n, d_j \mid P_j \\ \nu(d_j) \leq m_j}} \mu(d_j), \end{aligned}$$

for any choice of nonnegative even integers $m_1, \ldots, m_r$. Referring to (3.41), we obtain the upper bound

$$\begin{aligned} S(\mathcal{A}, \mathcal{P}) &\leq \sum_{\substack{d_1, \ldots, d_r \\ d_j \mid P_j, \nu(d_j) \leq m_j}} \mu(d_1) \cdots \mu(d_r) A_{d_1 \ldots d_r} \\ &= X \sum_{\substack{d_1, \ldots, d_r \\ d_j \mid P_j, \nu(d_j) \leq m_j}} \mu(d_1) \ldots \mu(d_r) \alpha(d_1) \ldots \alpha(d_r) \\ &\qquad + \sum_{\substack{d_1, \ldots, d_r \\ d_j \mid P_j, \nu(d_j) \leq m_j}} \mu(d_1) \ldots \mu(d_r) r(d_1 \ldots d_r). \\ &= X \prod_{j=1}^{r} \sum_{\substack{d_j \mid P_j \\ \nu(d_j) \leq m_j}} \mu(d_j) \alpha(d_j) + \sum_{\substack{d_1, \ldots, d_r \\ d_j \mid P_j, \nu(d_j) \leq m_j}} \mu(d_1) \ldots \mu(d_r) r(d_1 \ldots d_r). \end{aligned} \tag{3.44}$$

This is, essentially, the Brun-Hooley upper bound. To make this more amenable in applications, it is useful to replace the first term of (3.44), which we think of as the main term, with something more easily comparable with $X \prod_{p \in \mathcal{P}} (1 - \alpha(p))$. This can be accomplished by replacing the terms of the product with something more easily comparable with $\prod_{p \in \mathcal{P}_j} (1 - \alpha(p))$. For this, we utilize Lemma 3.4.1, which implies that for each $1 \leq j \leq r$,

$$0 \leq \sum_{\substack{d_j \mid P_j \\ \nu(d_j) \leq m_j}} \mu(d_j) \alpha(d_j) - \prod_{p \in \mathcal{P}_j} (1 - \alpha(p)) \leq \sum_{\substack{d_j \mid P_j \\ \nu(d_j) = m_j + 1}} \alpha(d_j).$$

Thus, if we set

$$\prod\nolimits^{(j)} := \prod_{p \in \mathcal{P}_j} (1 - \alpha(p)), \quad \sum\nolimits^{(j)} := \sum_{\substack{d_j \mid P_j \\ \nu(d_j) = m_j + 1}} \alpha(d_j), \tag{3.45}$$

then

$$X\prod_{j=1}^{r}\sum_{\substack{d_j\mid P_j\\ \nu(d_j)\le m_j}}\mu(d_j)\le X\prod_{j=1}^{r}\left(\prod\nolimits^{(j)}+\sum\nolimits^{(j)}\right)$$
$$=X\prod_{p\in\mathcal{P}}(1-\alpha(p))\prod_{j=1}^{r}\left(1+\sum\nolimits^{(j)}\Big/\prod\nolimits^{(j)}\right),$$

provided the division makes sense, i.e., provided $\alpha(p)<1$ for each $p\in\mathcal{P}$. We henceforth assume (as will be the case in all our applications) that this inequality holds.

Recalling that $1+T\le\exp(T)$ (which for nonnegative T is immediate from the series expansion for $\exp(T)$) and estimating the remainder term of (3.44) trivially, we arrive at:

Theorem 3.5.3 (Brun-Hooley Sieve, Upper Bound). *Let $\mathcal{P}=\dot\bigcup_{j=1}^{r}\mathcal{P}_j$ be a partition of $\mathcal{P}$. Suppose that $\alpha(p)<1$ for each $p\in\mathcal{P}$. For any choice of nonnegative even integers $m_1,\dots,m_r$, we have*

$$S(\mathcal{A},\mathcal{P})\le X\prod_{p\in\mathcal{P}}(1-\alpha(p))\exp\left(\sum_{j=1}^{r}\left(\sum\nolimits^{(j)}\Big/\prod\nolimits^{(j)}\right)\right)$$
$$+O\Bigg(\sum_{\substack{d_1,\dots,d_r\\ d_j\mid P_j,\nu(d_j)\le m_j}}|r(d_1\dots d_r)|\Bigg),\quad(3.46)$$

where $\prod^{(j)}$ and $\sum^{(j)}$ are defined, for $1\le j\le r$, by (3.45), and the implied constant is absolute.

3.5.3 Applications of the Upper Bound

Define $r(N)$ as the number of (ordered) representations of N as a sum of two primes; equivalently, as the number of ordered prime pairs $(p,N-p)$. In Chapter 1, we conjectured that as $N\to\infty$ through *even* integers,

$$r(N)\sim 2C_2\frac{N}{\log^2 N}\prod_{p\mid N,p>2}\frac{p-1}{p-2}.$$

We now use the Brun-Hooley sieve to establish an upper bound for this quantity of the conjectured order of magnitude:

Theorem 3.5.4. *For every even positive integer N,*

$$r(N)\ll\frac{N}{\log^2 N}\prod_{p\mid N}\left(1+\frac{1}{p}\right).$$

Let N be an even positive integer and define

$$\mathcal{A} := \{n(N-n) : 1 \leq n \leq N\}.$$

Then taking $\mathcal{P}$ as the set of all primes, we have for any positive choice of the parameter z,

$$r(N) \leq 2z + S(\mathcal{A}, \mathcal{P}, z).$$

Indeed, if $N = n + (N-n)$ is any representation of n as a sum of two primes, then either n or $N-n$ lies in $[2, z]$ or both have no prime factor not exceeding z. The former case occurs for not more than $2z$ values of n, and the n for which the latter hold (which necessarily satisfy $2 \leq n \leq N-2$) are counted by $S(\mathcal{A}, \mathcal{P}, z)$.

We now choose our usual sifting parameters: let $X = N$, and let $\alpha(d) = \omega(d)/d$, where

$$\omega(d) := |\{n \pmod{d} : n(N-n) \equiv 0 \pmod{d}\}|;$$

then

$$\alpha(p) = \begin{cases} 1/p & \text{if } p \mid N, \\ 2/p & \text{if } p \nmid N. \end{cases} \tag{3.47}$$

Note that as N is even, $\alpha(p) < 1$ for every prime p. By Lemma 3.4.9,

$$A_d = X\alpha(d) + r(d), \quad |r(d)| \leq \omega(d). \tag{3.48}$$

We think of $X = N$ as varying and we suppose $u > 1$ is fixed. Our first goal is to show that if u is large enough,

$$S(\mathcal{A}, \mathcal{P}, z) \ll X \prod\nolimits_{p \leq z} (1 - \alpha(p)) \quad (X \to \infty), \quad \text{where } z := X^{1/u}.$$

To apply the Brun-Hooley sieve to this situation we need a partition of $\mathcal{P} \cap [2, z]$. We introduce the notation

$$\eta = \log\log X,$$

and the choice of parameters

$$K := 1.57, \quad K_1 := 1.571. \tag{3.49}$$

Actually, for the current discussion it is only important that $1 < K < K_1$, but this choice will be particularly effective for the lower bound applications of §3.5.5.

For large X, we have $\eta < z = X^{1/u}$, so that if we define R as the minimal integer with

$$z^{1/K^R} < \eta$$

then $R \geq 1$. (Indeed, $R \to \infty$ with X.) For such X, we define

$$z_j = \begin{cases} z^{1/K^j} & \text{for } 0 \leq j \leq R-1, \\ \eta & \text{for } j = R, \\ 1 & \text{for } j = R+1. \end{cases}$$

We choose the partition described by

$$\mathcal{P} \cap [2, z] = \dot{\bigcup}_{j=1}^{R+1} \mathcal{P}_j,$$
$$\mathcal{P}_j := \{p \in \mathcal{P} : z_j < p \leq z_{j-1}\} \quad (1 \leq j \leq R+1),$$

and we define the corresponding nonnegative even integers $m_1, \ldots, m_{R+1}$ by

$$m_j = 2j \quad (j = 1, \ldots, R), \quad m_{R+1} = \infty.$$

The symbol ∞ means that m_{R+1} is chosen at least as large as the cardinality of $\mathcal{P}_{R+1}$. For definiteness, we take m_{R+1} as the smallest even integer with this property. In this way, the condition on a divisor d of P_{R+1} that it have no more than m_{R+1} prime divisors is automatically satisfied.

We are now finally in a position to apply the upper-bound (3.46) to our problem. Because we have chosen m_{R+1} at least as large as the size of $\mathcal{P}_{R+1}$,

$$\sum\nolimits^{(R+1)} = \sum_{\substack{d_{R+1} \mid P_{R+1} \\ \nu(d_{R+1}) = m_{R+1}+1}} \alpha(d_{R+1}) = 0, \tag{3.50}$$

being an empty sum. Hence $\sum^{(j)} / \prod^{(j)}$ vanishes at $j = R+1$, and to estimate the main term of (3.46) it suffices to estimate this ratio for $j = 1, \ldots, R$. The product in the denominator is handled by the following lemma:

Lemma 3.5.5. *As $x \to \infty$, we have*

$$\prod_{x<p\leq y} \left(1 - \frac{2}{p}\right) = \frac{\log^2 x}{\log^2 y} \left(1 + O\left(\frac{1}{\log x}\right)\right)$$

uniformly for $y \geq x$.

Proof. Suppose $x \geq 4$; then $2/p \leq 1/2$ for each $p \geq x$, so that $\log(1 - 2/p) = -2/p + O((-2/p)^2)$ with an absolute implied constant, and

$$\begin{aligned}
\sum_{x<p\leq y} \log\left(1 - \frac{2}{p}\right) &= -2 \sum_{x<p\leq y} \frac{1}{p} + O\left(\sum_{x<p\leq y} \frac{1}{p^2}\right) \\
&= -2\left(\log \frac{\log y}{\log x} + O\left(\frac{1}{\log x}\right)\right) + O\left(\frac{1}{x}\right) \\
&= \log \frac{\log^2 x}{\log^2 y} + O\left(\frac{1}{\log x}\right).
\end{aligned}$$

Exponentiating gives the result. $\square$

As $X \to \infty$, so do each of $z_1, \dots, z_R$ (since each is at least η). Consequently, Lemma 3.5.5 implies that for all large X,

$$\begin{aligned}\prod\nolimits^{(j)} &= \prod_{z_j < p \le z_{j-1}} (1-\alpha(p)) \ge \prod_{z_j < p \le z_{j-1}} \left(1 - \frac{2}{p}\right) \\ &= \frac{\log^2 z_{j-1}}{\log^2 z_j}\left(1 + O\left(\frac{1}{\log z_j}\right)\right) \ge \frac{1}{K^2}\left(1 + O\left(\frac{1}{\log \eta}\right)\right) \ge \frac{1}{K_1^2}. \quad (3.51)\end{aligned}$$

The sums are easier to estimate. For $1 \le j \le R$, we have

$$\begin{aligned}\sum\nolimits^{(j)} = \sum_{\substack{d_j \mid P_j \\ \nu(d_j) = m_j + 1}} \alpha(d_j) &\le \frac{1}{(m_j+1)!}\left(\sum_{p \in \mathcal{P}_j} \alpha(p)\right)^{m_j+1} \\ &\le \frac{1}{(m_j+1)!}\left(\sum_{p \in \mathcal{P}_j} \frac{2}{p}\right)^{m_j+1} \le \frac{(2\log K_1)^{m_j+1}}{(m_j+1)!}\end{aligned} \quad (3.52)$$

provided X is large enough, since then

$$\sum_{z_j < p \le z_{j-1}} \frac{2}{p} = 2 \log \frac{\log z_{j-1}}{\log z_j} + O\left(\frac{1}{\log z_j}\right) \le 2\log K + O\left(\frac{1}{\log \eta}\right) \le 2 \log K_1.$$

Putting (3.51), (3.52) together and recalling (3.50), we find that for large X,

$$\sum_{j=1}^{R+1} \left(\sum\nolimits^{(j)} \Big/ \prod\nolimits^{(j)}\right) \le K_1^2 \sum_{j=1}^{R} \frac{(2\log K_1)^{2j+1}}{(2j+1)!} \le K_1^2 \exp(2 \log K_1).$$

This shows the main term of (3.46) is bounded above by a constant multiple of $X \prod_{p \le z}(1-\alpha(p))$. For any fixed $u > 1$,

$$X \prod_{p \le X^{1/u}} (1 - \alpha(p)) = \frac{1}{2} X \prod_{2 < p \le X^{1/u}} (1 - 2/p) \asymp X/\log^2 X \quad (X \to \infty), \quad (3.53)$$

so that to obtain the estimate $S(\mathcal{A}, \mathcal{P}, z) \ll X \prod_{p \le z}(1-\alpha(p))$ we need only ensure the sum appearing in the expression for the remainder term,

$$\sum_{\substack{d_1, \dots, d_R+1 \\ d_j \mid P_j, \nu(d_j) \le m_j}} |r(d_1 \dots d_{R+1})|, \quad (3.54)$$

if of smaller order than $X/\log^2 X$. We will show that for an appropriate choice of u, the sum is $\ll X^\delta$ for a constant $\delta < 1$.

For this, first note that any product $d_1 \dots d_{R+1}$ appearing as an argument of $r(\cdot)$ in the sum (3.54) satisfies

$$\begin{aligned} d_1 \dots d_{R+1} &\le \left(\prod_{j=1}^{R} z_j^{m_j} \right) \eta^\eta \\ &= X^{\frac{1}{u}\left(\sum_{j=1}^{R} m_j/K^{j-1}\right)} X^{\log\log X \log\log\log X/\log X}. \end{aligned}$$

Also,

$$\sum_{j=1}^{R} \frac{m_j}{K^{j-1}} \le \sum_{j=1}^{\infty} \frac{2j}{K^{j-1}} = \frac{2K^2}{(K-1)^2} = 15.173\dots.$$

If we fix a choice of u exceeding this sum, say $u = 16$, then for large enough X, $d_1 \dots d_{R+1} \le X^{\delta_0}$ for every such product $d_1 \dots d_{R+1}$ and some fixed $\delta_0 < 1$. Now choose $\epsilon > 0$ with $(1+\epsilon)\delta_0 < 1$. Because (cf. (3.48))

$$|r(d)| \le \omega(d) = \prod_{p \mid d} \omega(p) \le 2^{\nu(d)} \le \tau(d) \ll d^\epsilon,$$

and because each integer admits at most one representation as $d_1 \dots d_{R+1}$ with $d_i \mid P_i$ for each i (because the d_i are products of primes from disjoint sets), the sum (3.54) above is

$$\ll \sum_{n \le X^{\delta_0}} n^\epsilon \ll \sum_{n \le X^{\delta_0}} (X^{\delta_0})^\epsilon \ll X^{\delta_0(1+\epsilon)} = X^\delta,$$

where $\delta = (1+\epsilon)\delta_0 < 1$.

Thus, for all large X,

$$\begin{aligned} S(\mathcal{A}, \mathcal{P}, X^{\frac{1}{16}}) &\ll X \prod_{p \le X^{\frac{1}{16}}} (1 - \alpha(p)) \\ &= X \prod_{\substack{p \le X^{\frac{1}{16}} \\ p \nmid N}} \left(1 - \frac{2}{p}\right) \prod_{\substack{p \le X^{\frac{1}{16}} \\ p \mid N}} \left(1 - \frac{1}{p}\right) \\ &\le X \prod_{\substack{p \le X^{\frac{1}{16}} \\ p \nmid N}} \left(1 - \frac{1}{p}\right)^2 \prod_{\substack{p \le X^{\frac{1}{16}} \\ p \mid N}} \left(1 - \frac{1}{p}\right) \\ &= X \prod_{p \le X^{\frac{1}{16}}} \left(1 - \frac{1}{p}\right)^2 \prod_{\substack{p \le X^{\frac{1}{16}} \\ p \mid N}} \left(1 - \frac{1}{p}\right)^{-1} \\ &\ll \frac{X}{\log^2 X} \prod_{p \mid N} \left(1 - \frac{1}{p}\right)^{-1}. \end{aligned}$$

Noting that

$$\prod_{p|N}\left(1-\frac{1}{p}\right)^{-1}\Big/\prod_{p|N}\left(1+\frac{1}{p}\right)=\prod_{p|N}\left(1-\frac{1}{p^2}\right)^{-1}\leq\zeta(2)<\infty,$$

we conclude that for large X,

$$S(\mathcal{A},\mathcal{P},X^{1/16})\ll\frac{X}{\log^2 X}\prod_{p|N}\left(1+\frac{1}{p}\right). \tag{3.55}$$

Consequently, for all large positive even numbers N,

$$\begin{aligned}r(N)&\leq S(\mathcal{A},\mathcal{P},X^{1/16})+2X^{1/16}\\&\ll\frac{X}{\log^2 X}\prod_{p|N}\left(1+\frac{1}{p}\right)=\frac{N}{\log^2 N}\prod_{p|N}\left(1+\frac{1}{p}\right).\end{aligned}$$

This gives the assertion of Theorem 3.5.4 for sufficiently large N, but for bounded N the theorem is trivial (fiddle with the constant).

The proof we have given applies mutatis mutandis to the generalized prime twin problem, i.e., the problem of estimating

$$\pi_N(x):=|\{p\leq x: p, p+N \text{ are both prime}\}|.$$

Let N be a positive even integer, and define the sequence

$$\mathcal{A}:=\{n(n+N):1\leq n\leq x\}.$$

Then

$$\pi_N(x)\leq z+S(\mathcal{A},\mathcal{P},z).$$

To estimate the last term, we take $X=x$ and choose $\alpha(d)=\omega(d)/d$, where here $\omega(d)$ is the number of solutions to the congruence $n(N+n)\equiv 0 \pmod d$. Then $\alpha(d)$ is again given by (3.47). If we now choose the other parameters exactly as before, the same proof as above shows that (3.55) holds for all sufficiently large X.

Since $X=x$ in our case, this estimate implies that for all positive even integers N and all $x\geq x_0$,

$$S(\mathcal{A},\mathcal{P},x^{1/16})\ll\frac{x}{\log^2 x}\prod_{p|N}\left(1+\frac{1}{p}\right).$$

Here the implied constant as well as x_0 are independent of N, and we may assume x_0 is at least 2. Thus, for $x\geq x_0$,

$$\pi_N(x)\ll x^{1/16}+\frac{x}{\log^2 x}\prod_{p|N}\left(1+\frac{1}{p}\right)\ll\frac{x}{\log^2 x}\prod_{p|N}\left(1+\frac{1}{p}\right)$$

uniformly in N. Since $\pi_N(x)$ is trivially bounded by x_0 for $2\leq x\leq x_0$, regardless of N, we can extend this estimate for $\pi_N(x)$ to all $x\geq 2$ and all positive even N, with perhaps a different implied constant. Thus we have shown:

Theorem 3.5.6. *Let N be a positive even integer. Then for $x \geq 2$,*

$$\pi_N(x) \ll \frac{x}{\log^2 x} \prod_{p|N} \left(1 + \frac{1}{p}\right),$$

where the implied constant is absolute.

Exercise 3.5.2. Let $r^*(n)$ denote the number of *unordered* representations of n as a sum of two primes. Clearly $r^*(n)$ is bounded above by the number of primes in the interval $[n/2, n-2]$, with equality holding exactly when $n-p$ is prime for each prime p with $n/2 \leq p \leq n-2$. Use the estimate (1.34) with Theorem 3.5.4 to prove that this upper bound for $r^*(n)$ is attained for at most finitely many n.

By a marriage of theory and computation, $n = 210$ has been shown to be the final example of equality ([DGNP93]).

3.5.4 The Lower Bound

A natural temptation here is is to simply parallel what we did in the upper bound case. If we suppose $m_1, \ldots, m_r$ to be r odd integers, then for each j,

$$\sum_{\substack{d_j|n, d_j|P_j \\ \nu(d_j) \leq m_j}} \mu(d_j) \leq \sum_{d_j|n, d_j|P_j} \mu(d_j).$$

But since it is (generally) not the case that both sides of this are nonnegative for each $1 \leq j \leq r$, we cannot simply take the product of both sides over j and expect the inequality to be preserved.

We thus look elsewhere. If $r = 1$, then rearranging the right hand inequality of Lemma 3.5.2 gives a lower bound for $s(n) = \sum_{d|n, d|P} \mu(d)$. The same lemma, with $\mathcal{P}, P$ replaced by $\mathcal{P}_j, P_j$ implies that for any choice of nonnegative even integers $m_1, \ldots, m_r$, we have

$$0 \leq \sum_{\substack{d_j|n, d_j|P_j \\ \nu(d_j) \leq m_j}} \mu(d_j) - \sum_{d_j|n, d_j|P} \mu(d_j) \leq \sum_{\substack{d_j|n, d_j|P \\ \nu(d_j) = m_j+1}} 1 \quad (1 \leq j \leq r). \tag{3.56}$$

These inequalities allow us to a coax a lower bound for

$$s(n) = \prod_{j=1}^{r} \sum_{d_j|n, d_j|P_j} \mu(d_j) \tag{3.57}$$

out of the following general inequality:

Lemma 3.5.7 ([FH00, Lemma 1]). *Suppose that $0 \leq x_j \leq y_j$ for $1 \leq j \leq r$. Then*

$$x_1 \ldots x_r \geq y_1 \ldots y_r - \sum_{l=1}^{r} (y_l - x_l) \prod_{\substack{j=1 \\ j \neq l}}^{r} y_j.$$

Proof. The result holds with equality when $r = 1$. If it holds for $r - 1$ for some $r \geq 2$, then

$$\begin{aligned} y_1 \dots y_r - x_1 \dots x_r &= (y_1 \dots y_{r-1} - x_1 \dots x_{r-1})y_r + (x_1 \dots x_{r-1})(y_r - x_r) \\ &\leq (y_1 \dots y_{r-1} - x_1 \dots x_{r-1})y_r + (y_1 \dots y_{r-1})(y_r - x_r) \\ &\leq \sum_{l=1}^{r-1}(y_l - x_l) \prod_{\substack{j=1 \\ j \neq l}}^{r} y_j + (y_r - x_r) \prod_{\substack{j=1 \\ j \neq r}}^{r} y_j = \sum_{l=1}^{r}(y_l - x_l) \prod_{\substack{j=1 \\ j \neq l}}^{r} y_j, \end{aligned}$$

so that the result follows by induction. ☐

We apply this with

$$x_j := \sum_{d_j | n, d_j | P_j} \mu(d_j), \quad y_j := \sum_{\substack{d_j | n, d_j | P_j \\ \nu(d_j) \leq m_j}} \mu(d_j).$$

Equation (3.56) both implies that the hypotheses of Lemma 3.5.7 are satisfied and gives us an upper bound on the terms $y_l - x_l$. Using this bound in Lemma 3.5.7 and recalling (3.57), we obtain

$$s(n) \geq \prod_{j=1}^{r} \sum_{\substack{d_j | n, d_j | P_j \\ \nu(d_j) \leq m_j}} \mu(d_j) - \sum_{l=1}^{r} \Bigg(\sum_{\substack{d_l | n, d_l | P_l \\ \nu(d_l) = m_l + 1}} 1 \Bigg) \prod_{\substack{j=1 \\ j \neq l}}^{r} \Bigg(\sum_{\substack{d_j | n, d_j | P_j \\ \nu(d_j) \leq m_j}} \mu(d_j) \Bigg).$$

Summing over $n \in \mathcal{A}$ shows

$$\begin{aligned} S(\mathcal{A}, \mathcal{P}) \geq & \sum_{\substack{d_1, \dots, d_r \\ d_j | P_j, \nu(d_j) \leq m_j}} \mu(d_1) \dots \mu(d_r) A_{d_1 \dots d_r} \\ & - \sum_{l=1}^{r} \sum_{\substack{d_1, \dots, d_r \\ d_j | P_j, \nu(d_j) \leq m_j (j \neq l) \\ d_l | P_l, \nu(d_l) = m_l + 1}} \frac{\mu(d_1) \dots \mu(d_r)}{\mu(d_l)} A_{d_1 \dots d_r}. \end{aligned} \tag{3.58}$$

Writing $A_d = X\alpha(d) + r(d)$, the right hand side of (3.58) becomes

$$X \prod_{j=1}^{r} \sum_{\substack{d_j | P_j \\ \nu(d_j) \leq m_j}} \mu(d_j)\alpha(d_j) - X \sum_{l=1}^{r} \sum_{\substack{d_l | P_l \\ \nu(d_l) = m_l + 1}} \alpha(d_l) \prod_{j \neq l} \sum_{\substack{d_j | P_j \\ \omega(d_j) \leq m_j}} \mu(d_j)\alpha(d_j), \tag{3.59}$$

up to an error term that is (with an absolute implied constant)

$$\ll \sum_{\substack{d_j | P_j (1 \leq j \leq r) \\ \theta_{d_1, \dots, d_r}}} |r(d_1 \dots d_r)|.$$

Here $\theta_{d_1,\dots,d_r}$ denotes the condition that there exist $r-1$ indices j, $1 \le j \le r$, for which $\nu(d_j) \le m_j$, while the remaining index satisfies $\nu(d_j) \le m_j + 1$.

Now assume, as in the treatment of the upper bound, that $\alpha(p) < 1$ for each $p \in \mathcal{P}$. Lemma 3.4.1 implies that for each $1 \le j \le r$,

$$\sum_{\substack{d_j \mid P_j \\ \nu(d_j) \le m_j}} \mu(d_j)\alpha(d_j) \ge \prod_{p \in \mathcal{P}_j} (1 - \alpha(p)) > 0,$$

so that we may reexpress the main term in (3.59) as

$$X\left(1 - \sum_{1 \le l \le r} \left(\sum_{\substack{d_l \mid P_l \\ \nu(d_l) = m_l + 1}} \alpha(d_l) \Big/ \sum_{\substack{d_l \mid P_l \\ \nu(d_l) \le m_l}} \mu(d_k)\alpha(d_l)\right)\right) \prod_{j=1}^{r} \sum_{\substack{d_j \mid P_j \\ \nu(d_j) \le m_j}} \mu(d_j)\alpha(d_j)$$

$$\ge X \prod_{p \in \mathcal{P}} (1 - \alpha(p)) \left(1 - \sum_{1 \le l \le r} \left(\sum_{\substack{d_l \mid P_l \\ \nu(d_l) = m_l + 1}} \alpha(d_l) \Big/ \prod_{p \in \mathcal{P}_l} (1 - \alpha(p))\right)\right).$$

Summarizing, we have proved the following theorem:

Theorem 3.5.8 (Brun-Hooley Sieve, Lower Bound). *Let $\mathcal{P} = \dot{\bigcup}_{j=1}^{r} \mathcal{P}_j$ be a partition of $\mathcal{P}$. Suppose that $\alpha(p) < 1$ for each $p \in \mathcal{P}$. For any choice of nonnegative even integers $m_1, \dots, m_r$, we have*

$$S(\mathcal{A}, \mathcal{P}) \ge X \prod_{p \in \mathcal{P}} (1 - \alpha(p)) \left(1 - \sum_{1 \le j \le r} \left(\sum\nolimits^{(j)} \Big/ \prod\nolimits^{(j)} \right)\right) + O\Bigg(\sum_{\substack{d_j \mid P_j (1 \le j \le r) \\ \theta_{d_1,\dots,d_r}}} |r(d_1 \dots d_r)| \Bigg),$$

where $\prod^{(j)}$ and $\sum^{(j)}$ are defined, for $1 \le j \le r$, by (3.45), and the implied constant is absolute.

3.5.5 Applications of the Lower Bound

We now prove the two remarkable theorems of Brun mentioned in the introduction: every large even integer is a sum of two 9-almost primes, and there exist infinitely pairs of 9-almost primes differing by 2.

Our setup for attacking these problems is the same as that used in attacking the analogous upper bound problems considered in §3.5.3. For the first of these, we assume N is a positive even integer, and we take $\mathcal{A} := \{n(N-n) : 1 \le n \le N\}$. As before, we let $\mathcal{P}$ be the set of all primes.

Suppose that we have a positive even integer N and a $u > 1$ for which

$$S(\mathcal{A}, \mathcal{P}, N^{1/u}) > 0. \tag{3.60}$$

Then there exists an n, $1 \le n \le N$, such that both n and $N-n$ have all their prime divisors exceeding $N^{1/u}$; since both n and $N-n$ are bounded by N, each must have at most u prime divisors. We will show that if we choose u large enough, (3.60) holds for all sufficiently large N (depending on u). Brun's results then follow from a quantitative determination of which u are "large enough."

For the most part, we may choose our sieving parameters as before, so that $X = N$ and α is given by (3.47). With u as a parameter to be chosen later, we define the partitions of $\mathcal{P} \cap [2, z]$ as in §3.5.3. However, the choice of the corresponding nonnegative integers m_i requires more care.

To describe this choice, suppose for the moment we have constructed a sequence $\{n_i\}_{i=1}^{\infty}$ of nonnegative even integers satisfying the two inequalities

$$\sum_{j=1}^{\infty} \frac{(2\log K_1)^{n_j+1}}{(n_j+1)!} < \frac{1}{K_1^2}, \tag{3.61}$$

$$\Gamma := 1 + \sum_{j=1}^{\infty} \frac{n_j}{K^{j-1}} < \infty, \tag{3.62}$$

where K_1, K_2 are given by (3.49). We fix $u > \Gamma$ and define (with the convention regarding "∞" of §3.5.3)

$$m_j = n_j \quad (1 \le j \le R), \qquad m_{R+1} = \infty.$$

Then for all large X, we have (recalling (3.50), (3.51), (3.52))

$$\begin{aligned}\sum_{j=1}^{R+1} \left(\sum\nolimits^{(j)} \Big/ \prod\nolimits^{(j)}\right) &= \sum_{j=1}^{R} \left(\sum\nolimits^{(j)} \Big/ \prod\nolimits^{(j)}\right) \\ &\le K_1^2 \sum_{j=1}^{R} \sum\nolimits^{(j)} \le K_1^2 \sum_{j=1}^{R} \frac{(2\log K_1)^{m_j+1}}{(m_j+1)!} \le 1-\epsilon\end{aligned}$$

for some fixed $\epsilon > 0$, by (3.61). This implies the main term in the lower bound

$$\begin{aligned}S(\mathcal{A},\mathcal{P}) \ge X \prod_{p\in\mathcal{P}} (1-\alpha(p)) &\left(1 - \sum_{1\le j\le R+1} \left(\sum\nolimits^{(j)} \Big/ \prod\nolimits^{(j)}\right)\right) \\ &+ O\Bigg(\sum_{\substack{d_j \mid P_j\,(1\le j\le r)\\ \theta_{d_1,\dots,d_{R+1}}}} |r(d_1\dots d_{R+1})|\Bigg),\end{aligned} \tag{3.63}$$

is (cf. (3.53))

$$\gg X \prod_{p\le X^{1/u}} (1-\alpha(p)) \gg X/\log^2 X \quad (X\to\infty).$$

The O-term can be treated much as before; the largest value of $d_1 \dots d_{R+1}$ appearing as an argument of $r(\cdot)$ is bounded above by

$$X^{\frac{1}{u}\left(1+\sum_{j=1}^{R} m_j/K^{j-1}\right)} X^{\log\log X \log\log\log X/\log X} \le X^{\Gamma/u+o(1)} \le X^{(1+\Gamma/u)/2}$$

for all large X. Since $(1+\Gamma/u)/2 < 1$, the argument of §3.5.3 implies the remainder term in (3.63) is $O(X^\delta)$ for some constant $\delta < 1$. Thus, with this choice of parameters, we obtain (3.60) in the stronger form

$$S(\mathcal{A}, \mathcal{P}, X^{1/u}) \gg X/\log^2 X \quad (X \to \infty).$$

It remains to construct a suitable sequence $\{n_i\}$. It is not hard to see that (3.61) and (3.62) will be satisfied with the simple choice $n_i = b + 2(i-1)$ $(i \geq 1)$, provided the initial even value b is chosen sufficiently large. However, this construction leads to an unnecessarily bloated value of Γ, so that while we still obtain a statement of the form "every large even N is a sum of two numbers with $O(1)$ prime factors," the implied constant here is larger than we might like.

We can get better results if we use the greedy algorithm in selecting the first several n_i (which are the most important terms as regards the value of Γ). We begin by choosing as many of the initial n_i to be 2 as (3.61) allows, then as many of the subsequent n_i to be 4 as allowed, etc. For example, since

$$\left\lfloor \frac{1/K_1^2}{(2\log K_1)^3/3!} \right\rfloor = 3,$$

we would choose $n_1, n_2, n_3 = 2$. We then compute

$$\left\lfloor \left(1/K_1^2 - 3\frac{(2\log K_1)^3}{3!}\right) \Big/ \frac{(2\log K_1)^5}{5!} \right\rfloor$$

to determine the number of subsequent i for which we set $n_i = 4$, etc.

Using a calculator, we find that the sequence obtained begins

$$n_1 = n_2 = n_3 = 2, \quad n_4 = \cdots = n_{10} = 4, \quad n_{11} = \cdots = n_{24} = 6.$$

Instead of continuing in this manner, we make the simple choice

$$n_{25} = 8 + 2(j-25) \quad (j \geq 25).$$

Then, setting $L := 2\log K_1$,

$$\begin{aligned}
&\frac{1}{K_1^2} - \sum_{j=1}^{\infty} \frac{(2\log K_1)^{n_j+1}}{(n_j+1)!} \\
&\quad \geq \frac{1}{K_1^2} - \sum_{j=1}^{3} \frac{L^3}{3!} - \sum_{j=4}^{10} \frac{L^5}{5!} - \sum_{j=11}^{24} \frac{L^7}{7!} - \sum_{j=25}^{\infty} \frac{L^{9+2(j-25)}}{(9+2(j-25))!} \\
&\quad \geq \frac{1}{K_1^2} - 3\frac{L^3}{3!} - 7\frac{L^5}{5!} - 14\frac{L^7}{7!} - \frac{L^9/9!}{1 - L^2/(11\cdot 10)} = 0.00003\ldots > 0,
\end{aligned}$$

so that (3.61) holds in this case. Also,

$$\begin{aligned}
\Gamma &= 1 + \sum_{j=1}^{3} \frac{2}{K^{j-1}} + \sum_{j=4}^{10} \frac{4}{K^{j-1}} + \sum_{j=11}^{24} \frac{6}{K^{j-1}} + \sum_{j=25}^{\infty} \frac{8+2(j-25)}{K^{j-1}} \\
&= 1 + \sum_{j=1}^{3} \frac{2}{K^{j-1}} + \sum_{j=4}^{10} \frac{4}{K^{j-1}} + \sum_{j=11}^{24} \frac{6}{K^{j-1}} + \frac{2(4K-3)}{K^{23}(K-1)^2} = 7.993\ldots.
\end{aligned}$$

Thus (3.62) holds. Moreover, we see we may in fact take $u = 7.995$, say. We thus obtain an even stronger theorem than that stated in the introduction: every large enough even N may be represented as a sum of two (positive) numbers each of which has no more than 7 prime divisors, and the number of such representations is

$$\gg X/\log^2 X = N/\log^2 N \quad (N \to \infty).$$

In like manner, we can show that the number of positive integers $n \le x$ for which $n, n+N$ have no prime divisor not exceeding $x^{1/7.995}$ is $\gg x/\log^2 x$ as $x \to \infty$, uniformly in the choice of the positive even integer N. Any such n satisfies

$$n \le n + N \le x + N < (x^{1/7.995})^8$$

once x is sufficiently large (depending only on N). It follows that for any fixed even positive integer N, there are

$$\gg_N x/\log^2 x \quad (x \to \infty)$$

integers $n \le x$ for which $n, n+N$ have no more than 7 prime divisors. Taking $N = 2$ gives (a bit more than) Brun's statement.

We conclude this chapter by mentioning the following remarkable theorem of Chen [Che73] which, perhaps better than any other result, illustrates the power of modern sieve methods.

Chen's Theorem. *Every large even number N is the sum of a prime and a 2-almost prime; moreover, denoting by $P_N(1,2)$ the number of primes $p \le N$ for which $N - p$ is a 2-almost prime, we have for even $N \to \infty$,*

$$P_N(1,2) \ge (0.67 + o(1))\frac{N}{\log^2 N}\prod_{p|N, p>2}\frac{p-1}{p-2}\prod_{p>2}\left(1 - \frac{1}{(p-1)^2}\right).$$

There are infinitely many primes p for which $p+2$ is a 2-almost prime. Denoting by $\pi_{1,2}(x)$ the number of such $p \le x$, we have for $x \to \infty$,

$$\pi_{1,2}(x) \ge (0.67 + o(1))\prod_{p>2}\left(1 - \frac{1}{(p-1)^2}\right)\frac{x}{\log^2 x}.$$

Apart from the constant 0.67, these remain the closest approximations to the Goldbach and twin prime conjectures proven to date. The proof of Chen's Theorem is the climax of Halberstam & Richert's classic book [HR74] on sieve methods.

Exercise 3.5.3. By modifying the proof given in this section, show that the number of representations of N as a sum of two 7-almost primes is

$$\gg \frac{N}{\log^2 N}\prod_{p|N, p>2}\frac{p-1}{p-2}.$$

3.5.6 Exercises: Further Applications of the Brun-Hooley Sieve

Exercise 3.5.4 *(Brun).* Prove the following theorems of Brun, announced in [Bru19b]:

a) Every arithmetic progression $a \pmod q$ with $\gcd(a,q) = 1$ contains infinitely many 5-almost primes. (Naturally, Dirichlet's theorem is off-limits here.)

b) If x is sufficiently large, there is always an 11-almost prime in the interval $(x, x+\sqrt{x}]$.

Suggestion: Imitate the lower bound applications of the text, including the selection of the first several m_j by the greedy algorithm, but begin instead with the values $K = 2.49, K_1 = 2.50$.

Exercise 3.5.5 *(Hardy & Littlewood [HL22]).* Show that

$$\pi(y+x) - \pi(y) \ll \frac{x}{\log x}$$

for $y \geq 0, x \geq 2$. Here the implied constant is absolute.

See Lemma 4.3.6 for a more explicit result.

The next two exercises concern a generalization of the upper bound results of §3.5.3 to several linear polynomials.

Exercise 3.5.6. Let $f_i(T) = a_i T + b_i$, $1 \leq i \leq k$ be a family of k linear polynomials with integer coefficients. Suppose that $(a_i, b_i) = 1$ for $i = 1, \ldots, k$, and that no f_i is an integral multiple of any other.

a) Show that
$$E := \prod_{i=1}^{k} a_i \prod_{1\leq i<j\leq k} (a_i b_j - a_j b_i) \neq 0.$$

b) Define, for positive integral d,
$$\omega(d) := |\{n \pmod d : \prod_{i=1}^{k} f_i(n) \equiv 0 \pmod d\}|.$$
Show that ω is multiplicative, $\omega(p) \leq k$ for every prime p, and $\omega(p) = k$ if and only if $p \mid E$.

Exercise 3.5.7 *(continuation).* Now suppose that, with the notation of the preceding problem, $\omega(p) < p$ for all primes p, i.e., that $\prod_{i=1}^{k} f_i$ has no fixed prime divisor.

a) Let
$$\mathcal{A} := \left\{ \prod_{i=1}^{k} f_i(n) : 1 \le n \le x \right\}.$$
Let $\mathcal{P}$ be the set of all primes. Show that for $x \ge 2$,
$$S(\mathcal{A}, \mathcal{P}, x^{1/16}) \ll \frac{x}{\log^k x} \prod_{p \mid E} (1 + 1/p)^{p - \omega(p)},$$
where the implied constant depends only on k.

b) Let $\pi_{f_1,\dots,f_k}(x)$ denote the number of $n \le x$ such that $|f_i(n)|$ is prime for each $i = 1, 2, \dots, k$. Show that for any $z > 0$,
$$\pi_{f_1,\dots,f_k}(x) \le S(\mathcal{A}, \mathcal{P}, z) + k(2z + 1).$$
Deduce from part a) that for $x \ge 2$,
$$\pi_{f_1,\dots,f_k}(x) \ll \frac{x}{\log^k x} \prod_{p \mid E} (1 + 1/p)^{p - \omega(p)},$$
where the implied constant depends only on k.

c) Using part b), rederive the results of §3.5.3.

The estimates we obtained in this chapter for the generalized twin prime problem were based on estimating the number of $n \le x$ not falling into either residue class $0, -N \pmod p$ for any $p \le x^{\epsilon}$, for a certain $\epsilon > 0$. Similarly, our estimates for $r(N)$ were based on "sieving out" the residue classes $0, -N$ for each prime $p \le N^{\epsilon}$. In both these cases, our estimates were facilitated by the existence of a sequence $\{a_n\}$ indexed by the positive integers up to x (or up to N) with the property that $p \mid a_n$ if and only if n falls into one of the singled-out congruence classes. Many sieve applications can be viewed in a similar light; the next exercise formulates two useful and oft-applied results along these lines:

Exercise 3.5.8. Let k be a positive integer.

a) Let $A > 0$. Suppose that to each prime $p \le x^A$, we associate $k_p \le k$ residue classes mod p. The number of positive integral $n \le x$ avoiding all of these residue classes is
$$\ll_{k,A} x \prod_{p \le x^A} \left(1 - \frac{k_p}{p}\right) \quad (x > 0),$$
uniformly in the particular choice of residue classes.

b) There exists $B = B(k) > 0$ with the following property: if we choose $k_p \le k$ residue classes mod p for each prime $p \le x^B$, then the number of positive integral $n \le x$ avoiding all these classes is
$$\asymp_k x \prod_{p \le x^B} \left(1 - \frac{k_p}{p}\right) \quad (x \to \infty),$$
uniformly in the particular choice of residue classes.

Hint: Use the Chinese Remainder Theorem to construct a polynomial f for which $p \mid f(n)$ precisely when n falls into one of the k_p chosen residue classes mod p.

Exercise 3.5.9. Using the lower bound estimate provided by part b) of the preceding exercise, show that a polynomial $f(T) \in \mathbf{Z}[T]$ of degree $k \geq 1$ with no fixed prime divisor represents infinitely many integers with $O(1)$ prime factors (counted with multiplicity), where the implied constant depends only on k.

References

[Bes31] A.S. Besicovitch, *On the density of the sum of two sequences of integers*, Math. Annalen **110** (1931), 336–341.

[Bru17] V. Brun, *Sur les nombres premiers de la forme $ap + b$*, Archiv for Mathem. og Naturw. **34** (1917), no. 8, 1–19.

[Bru19a] ———, *La série $\frac{1}{5}+\frac{1}{7}+\frac{1}{11}+\frac{1}{13}+\frac{1}{17}+\frac{1}{19}+\frac{1}{29}+\frac{1}{31}+\frac{1}{41}+\frac{1}{43}+\frac{1}{59}+\frac{1}{61}+\cdots$ où les dénominateurs sont "nombres premiers jumeaux" est convergente ou finie*, Bull. Sci. Math **43** (1919), 100–104, 124–128.

[Bru19b] ———, *Le crible d'Eratosthène et le théorème de Goldbach*, C. R. Math. Acad. Sci. Paris **168** (1919), 544–546.

[Bru20] ———, *Le crible d'Eratosthène et le théorème de Goldbach*, Christiania Vidensk. Selsk. Skr (1920), no. 3, 36pp.

[Che73] Jing Run Chen, *On the representation of a large even integer as the sum of a prime and the product of at most two primes*, Sci. Sinica **16** (1973), 157–176. MR 55 #7959

[DGNP93] Jean-Marc Deshouillers, Andrew Granville, Władysław Narkiewicz, and Carl Pomerance, *An upper bound in Goldbach's problem*, Math. Comp. **61** (1993), no. 203, 209–213. MR 94b:11101

[Erd65] P. Erdős, *Some recent advances and current problems in number theory*, Lectures on modern mathematics, Edited by T. L. Saaty, vol. III, John Wiley & Sons Inc., New York, 1965, pp. ix+321.

[FH00] Kevin Ford and H. Halberstam, *The Brun-Hooley sieve*, J. Number Theory **81** (2000), no. 2, 335–350. MR 2001d:11095

[Gra95] Andrew Granville, *Harald Cramér and the distribution of prime numbers*, Scand. Actuar. J. (1995), no. 1, 12–28, Harald Cramér Symposium (Stockholm, 1993). MR 96g:01002

[Gre] Benjamin Green, *Notes on sieve theory*, DVI files available for download online at `http://www.dpmms.cam.ac.uk/~bjg23/expos.html`.

[Gre01] George Greaves, *Sieves in number theory*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)], vol. 43, Springer-Verlag, Berlin, 2001. MR 2002i:11092

[HL22] G.H. Hardy and J.E. Littlewood, *Some problems of 'Partitio Numeronum', III: On the expression of a number as a sum of primes*, Acta Math. **44** (1922), 1–70.

[Hoo94] C. Hooley, *On an almost pure sieve*, Acta Arith. **66** (1994), no. 4, 359–368. MR 95g:11092

[HR74] H. Halberstam and H.-E. Richert, *Sieve methods*, Academic Press [A subsidiary of Harcourt Brace Jovanovich, Publishers], London-New York, 1974, London Mathematical Society Monographs, No. 4. MR 54 #12689

[LeV02] William Judson LeVeque, *Topics in number theory. Vol. I, II*, Dover Publications Inc., Mineola, NY, 2002, Reprint of the 1956 original [Addison-Wesley Publishing Co., Inc., Reading, Mass.; MR **18**, 283d], with separate errata list for this edition by the author. MR 1 942 365

[Nar86] Władysław Narkiewicz, *Classical problems in number theory*, Monografie Matematyczne [Mathematical Monographs], vol. 62, Państwowe Wydawnictwo Naukowe (PWN), Warsaw, 1986. MR 90e:11002

[Odl71] Andrew M. Odlyzko, *Sieve methods*, senior thesis, California Institute of Technology, 1971, available for online viewing at `http://auditorymodels.org/jba/BOOKS_Historical/`.

[Pól59] G. Pólya, *Heuristic reasoning in the theory of numbers*, Amer. Math. Monthly **66** (1959), 375–384. MR 21 #3392

[Sch00] A. Schinzel, *On sums of three unit fractions with polynomial denominators*, Funct. Approx. Comment. Math. **28** (2000), 187–194, Dedicated to Włodzimierz Staś on the occasion of his 75th birthday. MR 2002a:11029

[Vau70] R. C. Vaughan, *On a problem of Erdős, Straus and Schinzel*, Mathematika **17** (1970), 193–198. MR 44 #6600

[Wan84] Yuan Wang (ed.), *Goldbach conjecture*, Series in Pure Mathematics, vol. 4, World Scientific Publishing Co., Singapore, 1984. MR 87h:11102

[Wir59] Eduard Wirsing, *Bemerkung zu der Arbeit über vollkommene Zahlen*, Math. Ann. **137** (1959), 316–318. MR 21 #3389

Chapter 4

An Elementary Proof of the Prime Number Theorem

> No elementary proof of the prime number theorem is known, and one may ask whether it is reasonable to expect one. Now we know that the theorem is roughly equivalent to a theorem about an analytic function, the theorem that Riemann's zeta function has no roots on a certain line. A proof of such a theorem, not fundamentally dependent on the theory of functions, seems to me extraordinarily unlikely. It is rash to assert that a mathematical theorem cannot be proved in a particular way; but one thing seems quite clear. We have certain views about the logic of the theory; we think that some theorems, as we say lie deep and others nearer to the surface. If anyone produces an elementary proof of the prime number theorem, he will show that these views are wrong, that the subject does not hang together in the way we have supposed, and that it is time for the books to be cast aside and for the theory to be rewritten. – G.H. Hardy

4.1 Introduction

We mentioned in Chapter 1 that the prime number theorem, i.e., the assertion that

$$\pi(x) = (1 + o(1))\frac{x}{\log x} \qquad (x \to \infty), \tag{4.1}$$

was first established by Hadamard and de la Vallée Poussin (independently) in 1896, using methods of complex analysis and the ingenious ideas of Riemann forty years prior.

By the time the next forty years had passed, the analytic information required for the proof of the prime number theorem had been reduced to the nonvanishing of Riemann's function $\zeta(s)$ on the line $\Re(s) = 1$; e.g., we quote (without proof) the following general "Tauberian theorem" (which the reader unfamiliar with complex analysis can safely skip):

Theorem (Wiener-Ikehara, 1931). *Let* $\sum_{n=1}^{\infty} f(n)n^{-s}$ *be a Dirichlet series with nonnegative coefficients, convergent for* $\Re(s) > 1$. *Let* F *be the (analytic) function defined by the series in this region, and suppose* F *can be extended to a function analytic on an open set containing* $\Re(s) \geq 1$, *except possibly for a simple pole at* $s = 1$ *with residue* $r \geq 0$. *Then*

$$\frac{1}{x}\sum_{n \leq x} f(n) = (c + o(1))x \quad (x \to \infty).$$

Now partial summation shows that (Exercise 4.1.1)

$$\zeta(s) = \frac{1}{s-1} + 1 - s\int_1^{\infty} \frac{\{x\}}{x^{s+1}}\,dx \tag{4.2}$$

in the region $\Re(s) > 1$. The final term on the right is analytic in all of $\Re(s) > 0$, so this defines an analytic continuation of ζ to the half-plane $\Re(s) > 0$, except for a simple pole at $s = 1$ with residue 1. Since ζ has no zeros for $\Re(s) > 1$ (by the Euler-product expansion, for instance), *if* one can show ζ has no zeros on $\Re(s) = 1$, then $-\zeta'/\zeta$ analytically continues to an open set containing $\Re(s) \geq 1$, except for a simple pole at $s = 1$ with residue 1. Since (see Exercise 4.1.2)

$$-\frac{\zeta'}{\zeta}(s) = \sum_{n=1}^{\infty} \frac{\Lambda(n)}{n^s} \quad (\Re(s) > 1),$$

Wiener-Ikehara then yields

$$\psi(x) := \sum_{n \leq x} \Lambda(n) = (1 + o(1))x,$$

an assertion we saw was equivalent to the prime number theorem in Chapter 1 (see Corollary 1.5.5). Conversely, and more easily, if ζ does have zeros on the line $\Re(s) = 1$, then the prime number theorem cannot hold (again, see Exercise 4.1.2).

With this connection in mind, it is not hard to understand Hardy's remarks or to appreciate the later sensation resulting from Erdős and Selberg's 1949 derivations of the prime number theorem by totally elementary means ([Erd49], [Sel49]). Both Erdős and Selberg start with the *Selberg symmetry formula*, proved by Selberg in March of 1948:

$$\theta(x)\log x + \sum_{p \leq x} \log(p)\theta(x/p) = 2x\log x + O(x) \quad (x \geq 1). \tag{4.3}$$

This compact formula encapsulates a surprising amount of information about primes, including (ultimately) the prime number theorem itself. As an example of a result we can quickly skim off the top, note that if we define

$$a := \liminf_{x\to\infty} \frac{\theta(x)}{x}, \quad A := \limsup_{x\to\infty} \frac{\theta(x)}{x}, \tag{4.4}$$

and divide the symmetry formula by $x \log x$, we immediately obtain $A \leq 2$ (so we recover a Chebyshev-type upper bound). Only a little more work is required to see that in fact $a + A = 2$ (Exercise 4.1.4). We shall not say more about the symmetry formula or the clever ways Erdős and Selberg derived the prime number theorem from it, referring the reader instead to H.N. Shapiro's carefully crafted and well-motivated treatment [Sha83, Chapter 10]. For a historical viewpoint, see [Gol03].

Since 1949, a large number of papers have appeared presenting elementary proofs of the prime number theorem. Most are simplifications or variations on themes already present in Erdős and Selberg's work and employ the symmetry formula or a structurally-similar analog. The two exceptions to date are the proofs of Daboussi [Dab84] and Hildebrand [Hil86]. The goal of this chapter is a complete, reasonably self-contained exposition of Hildebrand's proof.

In the next three sections, we acquaint the reader with certain preliminary well-known results whose proofs were omitted from Hildebrand's original paper. Thus, in §4.2, we prove Landau's result (see [Lan06, §2]) that the prime number theorem is equivalent to

$$M^*(x) := \frac{1}{x}\sum_{n \leq x} \mu(n) = o(1) \qquad (x \to \infty). \tag{4.5}$$

In §4.3, we derive by means of the Selberg sieve that there are at most $(2 + o(1))y/\log y$ primes in an interval of length y (as $y \to \infty$), and in §4.4 we state and prove the Turán-Kubilius inequality for strongly additive functions. We sometimes digress briefly to discuss interesting related results not required for the proof of the PNT, e.g., the Brun-Titchmarsh Theorem and Erdős' multiplication table theorem.

With those preparations out of the way, we proceed to describe the lemmas endemic to Hildebrand's proof, concluding that section with a demonstration that (4.5) follows from a Fundamental Lemma stating that M^* varies slowly over large intervals. In the final section we give the proof of this Fundamental Lemma, thereby completing the proof of the prime number theorem.

4.1.1 Exercises

Exercise 4.1.1. Prove (4.2) by computing $\int t^{-s} dA(t)$ for $A(x) = \sum_{n \leq x} 1$.

The next two exercises assume some familiarity with complex analysis.

Exercise 4.1.2. Let $Z(s) := -\zeta'(s)/\zeta(s)$. Because $\sum n^{-s}$ is a series of analytic functions uniformly convergent on compact subsets of $\Re(s) > 1$, term-by-term differentiation is permissible, and

$$-\zeta'(s) = \sum_{n=1}^{\infty} \frac{\log n}{n^s} \quad (\Re(s) > 1).$$

The relation $\sum_{d|n}\Lambda(d) = \log n$ implies

$$\sum_{n=1}^{\infty}\frac{\log n}{n^s} = \left(\sum_{n=1}^{\infty}\frac{\Lambda(n)}{n^s}\right)\left(\sum_{n=1}^{\infty}\frac{1}{n^s}\right) \quad (\Re(s) > 1),$$

because all the involved Dirichlet series converge absolutely in this region. Hence

$$Z(s) = \sum_{n=1}^{\infty}\frac{\Lambda(n)}{n^s} \quad (\Re(s) > 1).$$

a) Prove that for $\Re(s) > 1$,

$$Z(s) = \frac{s}{s-1} + s\int_1^{\infty}(\psi(t) - t)\,\frac{dt}{t^{s+1}}.$$

b) Assuming $\psi(t) = t + o(t)$ (the PNT), show that the inner integral is $o(1/(\sigma - 1))$, as $\sigma := \Re(s) \downarrow 1$ (uniformly in s). Conclude that for fixed $t \neq 0$, one has

$$\lim_{\sigma\downarrow 1}(\sigma - 1)|Z(\sigma + it)| = 0.$$

c) On the other hand, show that if ζ has a zero of order $m \geq 0$ at $1 + it$ (so that necessarily $t \neq 0$), then

$$\lim_{\sigma\downarrow 1}(\sigma - 1)Z(\sigma + it) = -m.$$

Combining the results of b) and c), we see ζ has no zeros on the line $\Re(s) = 1$ if the prime number theorem holds.

Exercise 4.1.3. Observe that for $\Re(s) > 1$,

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

a) Show that in the same region, one has the representation

$$\sum_{n=1}^{\infty}\frac{\mu(n)}{n^s} = s\int_1^{\infty}\frac{M(t)}{t^{s+1}}\,dt,$$

where $M(x) := \sum_{n\leq x}\mu(n)$.

b) Assume $M(x) = o(x)$. Prove that the integral here is $o(1/(\sigma - 1))$ as $\sigma := \Re(s) \downarrow 1$ (uniformly in s). Use this to show ζ^{-1} has no poles on the line $\Re(s) = 1$, so that ζ is zero-free on $\Re(s) = 1$.

c) Assume $M(x) = O(x^{1/2+\epsilon})$ for a certain $\epsilon > 0$; prove that ζ^{-1} is regular for $\Re(s) > 1/2 + \epsilon$, so that ζ has no zeros there. In particular, if this estimate on M holds for each $\epsilon > 0$, then ζ is zero-free to the right of $\Re(s) = 1/2$, i.e., the *Riemann hypothesis* holds. In fact, Littlewood showed that the Riemann hypothesis is equivalent to this statement about M (see [Lan69, Satz 481]).

d) Riemann himself calculated several zeros of ζ on the line $\Re(s) = 1/2$. Assuming only that ζ has at least one zero on this line, use the method of b) to disprove the hypothetical estimate $M(x) = o(x^{1/2})$.

Exercise 4.1.4. Assume Selberg's fundamental formula (4.3). Show that $a+A = 2$, where a, A are given by (4.4).

Suggestions: Let $(x_n)_{n\geq 1}$ be a sequence of positive real with $x_n \to \infty$ and $\theta(x_n)/x_n \to A$. By combining the fundamental formula with the familiar estimate

$$\sum_{p\leq x} \frac{\log p}{p} \sim \log x,$$

show that $a + A \leq 2$. Choosing instead a sequence $(x'_n)_{n\geq 1}$ with $\theta(x'_n)/x'_n \to a$, show that $a + A \geq 2$.

Exercise 4.1.5. Recall that the Farey sequence $\mathfrak{F}_N$ of order N is the list of reduced fractions between 0 and 1 (inclusive) with denominator bounded by N, taken in increasing order. Assuming the result of Landau mentioned in the introduction, prove that the prime number theorem is equivalent to the estimate

$$S(N) := \sum_{p/q\in\mathfrak{F}_N} \cos 2\pi \frac{p}{q} = o(N).$$

Hint: Establish the identity

$$\mu(n) = \sum_{\substack{1\leq a\leq n\\ \gcd(a,n)=1}} e^{2\pi i a/n}$$

and use it to relate $S(N)$ to $M(N)$.

4.2 Some Statements Equivalent to the Prime Number Theorem

For positive real numbers x, define

$$M(x) := \sum_{n\leq x} \mu(n), \quad U(x) := \sum_{n\leq x} \frac{\mu(n)}{n}, \quad V(x) := \sum_{n\leq x} \frac{\mu(n)\log n}{n}.$$

Note that $M^*(x)$, as defined by (4.5), is simply $M(x)/x$.

Theorem 4.2.1 (Landau). *The following are equivalent:*

i. $\psi(x) \sim x$,

ii. $M(x) = o(x)$,

iii. $U(x) = o(1)$,

iv. $V(x) = o(\log x)$.

For the equivalence of the first three, see Landau's doctoral dissertation [Lan99] as well as his paper [Lan06]. The equivalence of iii) and iv) is proved in [Lan53, §156].

As we saw in Chapter 1 (see §1.5), the estimate $\psi(x) \sim x$ is equivalent to the prime number theorem, so that the same holds for each of the other items in the theorem statement.

4.2.1 An Inversion Formula and its Consequences

Theorem 4.2.2 (Generalized Möbius Inversion Formula). *Let F, G be functions defined on the positive real axis and vanishing on $(0,1)$. Suppose that*

$$F(x) = \sum_{n \le x} G(x/n) \quad (x > 0). \tag{4.6}$$

Then

$$G(x) = \sum_{n \le x} \mu(n) F(x/n) \quad (x > 0). \tag{4.7}$$

Proof. Both sides of (4.7) vanish for $0 < x < 1$, and for $x \ge 1$ we have

$$\begin{aligned}
\sum_{n \le x} \mu(n) F(x/n) &= \sum_{n \le x} \mu(n) \sum_{m \le x/n} G(x/mn) \\
&= \sum_{mn \le x} \mu(n) G(x/mn) \\
&= \sum_{k \le x} G(x/k) \sum_{d \mid k} \mu(d) \\
&= G(x).
\end{aligned}$$

□

Remark. If f, g are arithmetic functions, we may extend f, g to the positive real axis by defining them to vanish at nonintegral values. Then (one direction of) the Möbius inversion formula falls out of the above theorem.

Lemma 4.2.3. *As $x \to \infty$, we have both*

$$U(x) = O(1), \quad (\log x) U(x) = V(x) + O(1).$$

Proof. We take $G(x) = 1$ for $x > 1$; then (4.6) holds for $F(x) = \lfloor x \rfloor$, so that (4.7) shows

$$1 = \sum_{n \le x} \mu(n) \left\lfloor \frac{x}{n} \right\rfloor = xU(x) + O(x).$$

Dividing by x gives the first estimate of the lemma. For the second, take $G(x) = x$ for $x > 1$; then

$$F(x) = \sum_{n \le x} \frac{x}{n} = x \log x + \gamma x + O(1)$$

for $x \geq 1$, whence

$$\begin{aligned} x = G(x) &= \sum_{n\leq x} \mu(n) \left(\frac{x}{n} \log \frac{x}{n} + \gamma \frac{x}{n} + O(1) \right) \\ &= x \log x \sum_{n \leq x} \frac{\mu(n)}{n} - x \sum_{n\leq x} \frac{\mu(n) \log n}{n} + \gamma x \sum_{n\leq x} \frac{\mu(n)}{n} + O(x) \\ &= (x \log x) U(x) - xV(x) + xU(x) + O(x) \\ &= (x \log x) U(x) - xV(x) + O(x), \end{aligned}$$

using the estimate $U(x) = O(1)$ that we just established. Dividing by x and rearranging completes the proof. □

4.2.2 An Estimate of Dirichlet

We need an estimate on the summatory function of τ:

Lemma 4.2.4 (Dirichlet). *As $x \to \infty$,*

$$\sum_{n\leq x} \tau(n) = x \log x + (2\gamma - 1)x + O(x^{1/2}).$$

Proof. Observe that

$$T(x) := \sum_{n\leq x} \tau(n) = \sum_{n\leq x} \sum_{ab=n} 1 = \sum_{ab\leq x} 1,$$

i.e., $T = T(x)$ is just the number of lattice points (with positive integral coordinates) beneath the hyperbola $ab = x$. Such lattice points either satisfy $a \leq \sqrt{x}$ or $b \leq \sqrt{x}$, so that

$$T = T_1 + T_2 - T_3,$$

where

$$\begin{aligned} T_1 &= |\{(a, b) : a \leq \sqrt{x} : ab \leq x\}|, \\ T_2 &= |\{(a, b) : b \leq \sqrt{x} : ab \leq x\}|, \\ T_3 &= |\{(a, b) : a \leq \sqrt{x}, b \leq \sqrt{x}, ab \leq x\}|. \end{aligned}$$

Now $T_3 = \lfloor \sqrt{x} \rfloor^2 = x + O(\sqrt{x})$, while

$$\begin{aligned} T_2 = T_1 = \sum_{a\leq\sqrt{x}} \sum_{b\leq x/a} 1 = \sum_{a\leq\sqrt{x}} \lfloor x/a \rfloor &= x \sum_{a\leq\sqrt{x}} \frac{1}{a} + O(\sqrt{x}) \\ &= x(\log x^{1/2} + \gamma + O(x^{-1/2})) + O(x^{1/2}) \\ &= \frac{1}{2} x \log x + \gamma x + O(x^{1/2}). \end{aligned}$$

The result follows. □

4.2.3 Proof of the Equivalences

We will prove

$$\begin{aligned}U(x) = o(1) \Rightarrow M(x) = o(x) \Rightarrow \psi(x) \sim x \\ \Rightarrow V(x) = o(\log x) \Rightarrow U(x) = o(1).\end{aligned}$$

(This corresponds to $3 \Rightarrow 2 \Rightarrow 1 \Rightarrow 4 \Rightarrow 3$.) Of course the implication of main interest for us is

$$M(x) = o(x) \Longrightarrow \psi(x) \sim x,$$

since it is the former statement which we shall ultimately establish.

Proof $[\mathbf{U(x) = o(1) \Rightarrow M(x) = o(x)}]$. This is a straightforward partial summation:

$$\begin{aligned}M(x) = \int_{1/2}^{x} t\,dU(t) &= xU(x) - \int_{1}^{x} U(t)\,dt \\ &= o(x) - \int_{1}^{x} o(1)\,dt = o(x).\end{aligned}$$

□

Proof $[\mathbf{M(x) = o(x)} \Rightarrow \boldsymbol{\psi}\mathbf{(x) \sim x}]$. This is the most difficult of all the implications. Our strategy will be to write

$$\psi(x) - x = \sum_{qd \le x} \mu(q) f(d) + O(1)$$

for a certain function f, and then to use the relation $M(x) = o(x)$ to show the right hand sum here is also $o(x)$ as $x \to \infty$.

To construct such an f, recall the identities

$$\Lambda(n) = \sum_{d|n} \mu(d) \log \frac{n}{d}, \quad 1 = \sum_{d|n} \mu(d)\tau(d), \quad \left\lfloor \frac{1}{n} \right\rfloor = \sum_{d|n} \mu(d);$$

these should all be familiar, except possibly the second, which however follows quickly from Möbius inversion. Using these relations, we find that for any constant C and any $x \ge 1$, we have

$$\begin{aligned}\psi(x) - [x] + C &= \sum_{n \le x} \sum_{d|n} \mu(d) \left(\log \frac{n}{d} - \tau(n/d) + C\right) \\ &= \sum_{qd \le x} \mu(d) \left(\log(q) - \tau(q) + C\right). \qquad (4.8)\end{aligned}$$

We can thus take, for any constant C,

$$f(n) = \log n - \tau(n) + C;$$

however, it is convenient to choose C so that the partial sums of f are small. By Dirichlet's estimate for $\sum_{n\le x}\tau(n)$, if we choose $C=2\gamma$, then

$$F(x) := \sum_{n\le x} f(n) \le B\sqrt{x},$$

for some constant B and each $x\ge 1$.

With f defined with respect to this C, we proceed to show

$$\sum_{qd\le x}\mu(d)f(q) = o(x). \tag{4.9}$$

Let $a,b\ge 1$ with $ab=x$; we shall specify a and b more precisely later. Then we can rewrite

$$\sum_{qd\le x}\mu(d)f(q) = \sum_{n\le a}\mu(n)F(x/n) + \sum_{n\le b} f(n)M(x/n) - M(a)F(b). \tag{4.10}$$

Indeed, each lattice points (q,d) underneath the hyperbola $qd\le x$ either satisfies $q\le a$ or $d\le b$, and is taken into account by either the first or the second sum. Those which satisfy both are counted twice, and these are subtracted off in the term $M(a)Y(b)$.

The first sum on the right hand side of (4.10) is bounded in absolute value by

$$\begin{aligned} B\sum_{n\le a}\sqrt{x/n} &= B\sqrt{x}\sum_{n\le a}\frac{1}{\sqrt{n}} \\ &\le B\sqrt{x}\left(1+\int_1^a \frac{dt}{\sqrt{t}}\right) \le 2B\sqrt{x}\sqrt{a} = 2B\frac{x}{\sqrt{b}}, \end{aligned}$$

recalling that $ab=x$. Now fix $b=b(\epsilon)\ge 1$ large enough that

$$2B/\sqrt{b} < \epsilon. \tag{4.11}$$

So we can bound the first term on the right hand side of (4.10):

$$\left|\sum_{n\le x}\mu(n)F(x/n)\right| < \epsilon x.$$

To estimate the second term, choose $c=c(\epsilon,K)$ large enough that $x/n>c$ implies $|M(x/n)|<(\epsilon/K)x/n$, where K is a constant to be specified shortly. Then if $x>bc$, one has $x/n>c$ for each $n\le b$, so that the second term is bounded above in absolute value by

$$\frac{\epsilon}{K}\sum_{n\le b}|f(n)|/n.$$

So if we take $K := \sum_{n \le b} |f(n)|/n$, then the second term is majorized by ϵx. Note that K, and hence c, depends only on ϵ.

For the third term, we note

$$|M(a)F(b)| \le a(B\sqrt{b}) \le ab\frac{B}{\sqrt{b}} < \epsilon x,$$

by (4.11). It follows that the entire right hand side of (4.10) is bounded in absolute value by $3\epsilon x$ for $x > bc$, where b, c depend only on ϵ. Since ϵ was arbitrary, (4.9) follows. $\square$

Remark. The way we began this proof looks mysterious, but it can be motivated by looking at appropriate Dirichlet series: the assertion $\psi(x) \sim x$ says that the average of the coefficients of the Dirichlet series $Z(s)$ of Exercise 4.1.2 is 1, i.e., that the average of the coefficients of $Z(s) - \zeta(s)$ is 0. But

$$Z(s) - \zeta(s) = -\frac{\zeta'}{\zeta} - \zeta = (-\zeta' + \zeta^2)\frac{1}{\zeta}.$$

Now compute the coefficients of $-\zeta' + \zeta^2$ and of $1/\zeta$ and multiply the Dirichlet series to obtain (4.8), without the constant C.

Proof $[\psi(\mathbf{x}) \sim \mathbf{x} \Rightarrow \mathbf{V}(\mathbf{x}) = \mathbf{o}(\log \mathbf{x})]$. In Chapter 2, we observed already that

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

so that Möbius inversion implies

$$\mu(n)\log n = -\sum_{d|n} \Lambda(d)\mu(n/d).$$

Hence

$$\begin{aligned} V(x) = \sum_{n \le x} \frac{\mu(n)\log n}{n} &= -\sum_{n\le x}\frac{1}{n}\sum_{d|n}\Lambda(d)\mu(n/d) \\ &= -\sum_{dd' \le x} \frac{\Lambda(d)}{d}\frac{\mu(d')}{d'} = -\sum_{d\le x}\frac{\Lambda(d)}{d}U(x/d). \end{aligned}$$

Now write $\psi(x) = x + R(x)$, so that the assumption $\psi(x) \sim x$ implies $R(x) = o(x)$. Then we can rewrite this final sum in the form

$$\sum_{d \le x} \frac{\psi(d) - \psi(d-1)}{d} U(x/d) = \sum_{d\le x}\frac{1}{d}U(x/d) + \sum_{d \le x}\frac{R(d) - R(d-1)}{d}U(x/d).$$

The first sum on the right can be explicitly evaluated:

$$\sum_{d\le x}\frac{1}{d}\sum_{d' \le x/d}\frac{\mu(d')}{d'} = \sum_{dd'\le x}\frac{\mu(d)}{dd'} = \sum_{n \le x}\frac{1}{n}\sum_{d|n}\mu(d) = 1,$$

for $x > 1$, so this first sum is certainly $o(\log x)$. Handling the second is a bit trickier. We can reexpress it, as in partial summation, by

$$\frac{R(\lfloor x \rfloor)}{\lfloor x \rfloor} + \sum_{d \le x-1} \frac{R(d)}{d} \left(U\left(\frac{x}{d}\right) - U\left(\frac{x}{d+1}\right) \right) + \sum_{d \le x-1} \frac{R(d)}{d(d+1)} U\left(\frac{x}{d+1}\right).$$

The first term is $o(1)$. Keeping in mind that $U(x) = O(1)$, the last term here is seen to be

$$\begin{aligned} \ll \sum_{d \le x-1} \frac{|R(d)|}{d(d+1)} &= \sum_{d \le x-1} \frac{o(d)}{d(d+1)} \\ &= \sum_{d \le x-1} o\left(\frac{1}{d+1}\right) = o\left(\sum_{d \le x} \frac{1}{d}\right) = o(\log x), \end{aligned}$$

so all the difficulty lies in establishing the estimate $o(\log x)$ for the second term. This sum is bounded, in absolute value, by

$$\sum_{d \le x-1} \frac{|R(d)|}{d} \sum_{x/(d+1) < n \le x/d} \frac{1}{n}, \tag{4.12}$$

so it will suffice to show (4.12) is $o(\log x)$.

For this let $\epsilon > 0$ be given, and choose a positive integer x_0 sufficiently large that $|R(d)|/d < \epsilon$ for $d \ge x_0$. We split the sum (4.12) at x_0 and estimate the two pieces. Let $B = B(\epsilon)$ be an upper bound for $|R(d)|/d$ for $d \le x_0$; then the first piece of the sum (taken over $d < x_0$) is majorized in absolute value by

$$B \sum_{D \le x-1} \sum_{x/(d+1) < n \le x/d} \frac{1}{n} = B \sum_{x/x_0 < n \le x} \frac{1}{n} \le B(\log x_0 + 1) < \epsilon \log x,$$

for sufficiently large x (depending on ϵ). Similarly, the second piece is bounded in absolute value by

$$\epsilon \sum_{x/\lfloor x \rfloor < n \le x/x_0} \frac{1}{n} \le \epsilon \left(\log \frac{x}{x_0} + 1 \right) < 2\epsilon \log x$$

for x sufficiently large. It follows that the sum in (4.12) is bounded by $3\epsilon \log x$ for $x > x_1(\epsilon)$, say. As ϵ was arbitrary, the estimate $o(x \log x)$ follows. $\square$

Proof $[\mathbf{V(x) = o(\log x) \Rightarrow U(x) = o(1)}]$. By Lemma 4.2.3,

$$U(x) = \frac{V(x)}{\log x} + O\left(\frac{1}{\log x}\right) = o(1) + O\left(\frac{1}{\log x}\right) = o(1). \qquad \square$$

4.2.4 Exercises

Exercise 4.2.1. Let $f(n)$ be an arithmetic function and suppose $\sum_{n=1}^\infty f(n)/n$ converges. Prove that $\sum_{n\le x} f(n) = o(x)$. Note that this generalizes the proof that $U(x) = o(1)$ implies $M(x) = o(x)$.

Exercise 4.2.2 *(continuation).* Using the result of the previous exercise, prove that the prime number theorem follows from the convergence of

$$\sum_{n=1}^{\infty} \frac{\Lambda(n)-1}{n}. \tag{4.13}$$

Exercise 4.2.3 *(continuation).*

a) Prove that for $x \ge 1$,

$$\sum_{n\le x} \frac{\Lambda(n)-1}{n} = \sum_{qd\le x} \frac{\mu(d)f(q)}{qd} - 2\gamma,$$

where f is given (as before) by

$$f(n) = \log n - \tau(n) + 2\gamma.$$

b) By the hyperbola method, we can reexpress

$$\sum_{qd\le x} \frac{\mu(d)f(q)}{qd} = \sum_{q\le y} \frac{f(q)}{q} U(x/q) + \sum_{d\le x/y} \frac{\mu(d)}{d} G(x/d) - G(y)U(x/y), \tag{4.14}$$

where

$$G(z) := \sum_{k\le z} \frac{f(k)}{k}$$

and $1 \le y \le x$. Prove that $G(z) = C + O(1/\sqrt{z})$ for an appropriate choice of constant C.

c) Assume the prime number theorem in the form $U(x) = o(1)$. By first fixing y and letting $x \to \infty$ in (4.14), then letting $y \to \infty$, conclude that

$$\sum_{qd\le x} \frac{\mu(d)f(q)}{qd} = o(1),$$

and hence that the infinite sum (4.13) has the value -2γ.

d) Deduce from this that

$$\sum_{n\le x} \frac{\Lambda(n)}{n} = \log x - \gamma + o(1).$$

Exercise 4.2.4. Liouville's λ function is defined as the completely multiplicative function with $\lambda(p) = -1$ for every prime p. Show that

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

and use this to prove

$$L(x) := \sum_{n\le x} \lambda(n) = \sum_{d\le\sqrt{x}} M(x/d^2). \tag{4.15}$$

Deduce that $M(x) = o(x)$ implies $L(x) = o(x)$. The latter has the following pleasant interpretation: the probability an integer has an even number of prime factors (counted with multiplicity) is precisely 1/2.

Exercise 4.2.5 *(continuation).* By using either (4.15) or proceeding directly, prove that

$$M(x) = \sum_{d\le\sqrt{x}} \mu(d)L(x/d^2).$$

Deduce that $L(x) = o(x)$ implies $M(x) = o(x)$. Taken with the result of the previous exercise, this shows $L(x) = o(x)$ is equivalent to $M(x) = o(x)$, and hence to the prime number theorem.

4.3 An Upper Bound on $\pi(x+y)-\pi(x)$

Our aim here is to establish the next lemma, which is the only result of this section we shall need in the proof of the PNT.

Lemma 4.3.1. *As $y\to\infty$,*

$$\pi(x+y)-\pi(x) \le (2+o(1))\frac{y}{\log y}$$

uniformly for $x\ge 1$.

Remark. It must be emphasized that though we present this as a lemma in a proof of the prime number theorem, it is not itself as a corollary of that theorem. As illustration, let us try to derive a bound on $\pi(x+y)-\pi(x)$ from the hypothetical estimate $\pi(x) = \mathrm{li}(x) + O(x^{1/2}\log x)$, a very strong form of the prime number theorem known to be equivalent to the Riemann hypothesis. Then, assuming say that $y\le x$, the reader can check

$$\mathrm{li}(x+y)-\mathrm{li}(x) = \int_x^{x+y}\frac{dt}{\log t} = (1+o(1))\frac{y}{\log x} \quad (x\to\infty);$$

one consequence is that if $y = y(x) \le x$ tends to infnity in such a way that $(\log^2 x)x^{1/2}/y$ tends to 0, then $\pi(x+y)-\pi(x) \sim y/\log x$ (as $x\to\infty$). So far so good. On the other hand, suppose we wish to take y smaller than $x^{1/2}$, say

$y = y(x) = \log^{100} x$; then this version of the prime number theorem provides no bound on the number of primes in $(x, x+y]$ other than the trivial bound $O(x^{1/2}\log^{1/2} x)$, which is larger than the number of integers in the interval! Indeed, even the expected asymptotic $\pi(x+y) - \pi(x) \sim y/\log x$ is known to be false for this choice of y (see [Mai85]). In these situations, Lemma (4.3.1) provides useful information.

Our strategy in proving this will be as follows: instead of directly estimating the number of primes in the interval $(x, x+y]$, we will estimate instead the number of integers in the interval $(x, x+y]$ all of whose prime factors exceed z, where z is some parameter to be chosen later. The desired estimate will then follow from the inequality

$$\pi(x+y) - \pi(x) \leq z + |\{x < n \leq x+y : p \mid n \Rightarrow p > z\}|, \tag{4.16}$$

which reflects that a prime in the interval $(x, x+y]$ is either bounded by z or has no prime factors not exceeding z.

Those who have read Chapter 3 will immediately recognize this estimation as a sieving problem, corresponding to the choice

$$\mathcal{A} := \{n : x < n \leq y\}, \quad \mathcal{P} = \{\text{all primes}\}. \tag{4.17}$$

Indeed, the rightmost term of (4.16) is what in Chapter 3 was defined as $S(\mathcal{A}, \mathcal{P}, z)$: the number of elements of $\mathcal{A}$ divisible by no prime $p \in \mathcal{P}$ not exceeding z. Not surprisingly, we will ultimately prove Lemma 4.3.1 by applying a suitable upper bound sieve method, that of Selberg (introduced in [Sel47]). Note that the Brun-Hooley sieve of the last chapter can be used in this capacity also, but yields a bound of the form

$$\pi(x+y) - \pi(x) \leq (C + o(1))\frac{x}{\log x},$$

where it seems impossible to obtain $C = 2$. Unfortunately, $C \leq 2$ is essential to the proof of 4.5.2.

4.3.1 Preparatory Lemmas

Lemma 4.3.2. *Let $a_1, \ldots, a_n$ be positive real numbers, and let*

$$Q(x_1, \ldots, x_n) := a_1 x_1^2 + a_2 x_2^2 + \cdots + a_n x_n^2.$$

The minimum value of Q, subject to the linear constraint

$$b_1 x_1 + \cdots + b_n x_n = 1,$$

with the b_i given real numbers not all vanishing, is given by $(\sum_{i=1}^n b_i^2/a_i)^{-1}$. Moreover, this minimum value is uniquely attained at the point

$$x_i = \frac{b_i/a_i}{\sum_{i=1}^n b_i^2/a_i} \quad (i = 1, \ldots, n).$$

Proof. By the Cauchy-Schwarz inequality,

$$1=\left(\sum_{i=1}^{n} b_i a_i^{-1/2}\cdot a_i^{1/2}x_i\right)^2 \le \left(\sum_{i=1}^{n} b_i^2 a_i^{-1}\right)\left(\sum_{i=1}^{n} a_i x_i^2\right). \tag{4.18}$$

Equality is attained in (4.18) precisely when there exists a real number t_0 with

$$a_i^{1/2}x_i = t_0\frac{b_i}{\sqrt{a_i}} \quad (i=1,\dots,n), \tag{4.19}$$

equivalently

$$x_i = t_0 b_i/a_i \quad (i=1,\dots,n). \tag{4.20}$$

Supposing the x_i are such that equality holds here, we have

$$1=\sum_{i=1}^{n} b_i x_i = \sum_{i=1}^{n} t_0\frac{b_i^2}{a_i},$$

which forces $t_0=(\sum_{i=1}^{n} b_i^2/a_i)^{-1}$. Solving for the x_i from (4.20) yields

$$x_i=\frac{b_i/a_i}{\sum_{i=1}^{n} b_i^2/a_i}. \tag{4.21}$$

Since this choice of the x_i can be directly checked to satisfy (4.19) with the given value of t_0, it follows it does indeed give the minimum value. Its uniqueness in this respect finally follows now from (4.21). □

Lemma 4.3.3 (Möbius Inversion, Dual Form). *Let $\mathcal{D}$ be a finite set of positive integers with the property that if $d\in\mathcal{D}$ and $d'\mid d$, then $d'\in\mathcal{D}$. Suppose f,g are complex-valued functions defined on $\mathcal{D}$. Then*

$$f(n)=\sum_{\substack{d\in\mathcal{D}\\ n\mid d}} g(d) \iff g(n)=\sum_{\substack{d\in\mathcal{D}\\ n\mid d}}\mu(d/n)f(d).$$

Proof. We prove the forward implication, leaving the other direction to the reader. The proof consists of writing everything out. If $n\in\mathcal{D}$,

$$\begin{aligned}\sum_{\substack{n\mid d\\ d\in\mathcal{D}}}\mu(d/n)f(d) &= \sum_{\substack{n\mid d\\ d\in\mathcal{D}}}\mu(d/n)\sum_{\substack{d\mid d'\\ d'\in\mathcal{D}}} g(d')\\ &= \sum_{\substack{n\mid d'\\ d'\in\mathcal{D}}} g(d')\sum_{\substack{d\mid d',\, n\mid d\\ d\in\mathcal{D}}}\mu(d/n) = \sum_{\substack{n\mid d'\\ d'\in\mathcal{D}}} g(d')\sum_{\substack{k\mid d'/n\\ nk\in\mathcal{D}}}\mu(k).\end{aligned}$$

But if $k\mid d'/n$, then $nk\mid d'\in\mathcal{D}$; since $\mathcal{D}$ is divisor-closed, the condition $nk\in\mathcal{D}$ is automatic. It follows that the final inner sum vanishes unless $d'/n=1$, where it takes the value 1, so that the entire above expression simplifies to $g(n)$. □

Remark. The usual Möbius inversion formula also holds for functions on finite divisor-closed sets (even arbitrary divisor closed sets). It is worth saying a word about the connection between this usual inversion formula and what we have been calling its dual.

We suppose $\mathcal{D}$ finite. We can view Möbius inversion as a statement about linear maps, namely that

$$T\colon f \to \left(n \mapsto \sum_{d|n} f(d)\right), \qquad S\colon g \to \left(n \mapsto \sum_{d|n} g(d)\mu(n/d)\right)$$

are mutually inverse linear transformations of the vector space V of complex-valued functions on $\mathcal{D}$. Similarly, the dual Möbius inversion formula states that T' and S' are inverse, where T' and S' are defined analogously. Linear algebra shows that $T \circ S = \mathrm{id}$ if and only if $S \circ T = \mathrm{id}$, i.e., one direction of the Möbius inversion formula implies the other, and similarly for the dual formula.

Now let $(e_d)_{d\in\mathcal{D}}$ be a basis of characteristic functions of elements of $\mathcal{D}$. Then the matrix of T' with respect to this basis is the transpose of that of T, while the matrix of S' is the transpose of that of S. It follows that T is the inverse of S if and only if T' is the inverse of S'; i.e., the Möbius inversion formula holds if and only if the dual Möbius inversion formula holds.

4.3.2 Proof of Lemma 4.3.1 by Selberg's sieve

We will use notation consistent with Chapter 3, as this will facilitate the reader's adapting our results to a more general sieving situation (see §4.3.3):

We let $\mathcal{A}$ (respectively $\mathcal{P}$) denote the finite sequence of integers (respectively finite set of primes) given by (4.17). Given a parameter z, we let

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

and we write

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

We let

$$A_d = |\{a \in \mathcal{A} : d \mid a\}|.$$

Then, for any $d \mid P$ (indeed, for any positive integer d), we have

$$A_d = \left\lfloor \frac{x+y}{d} \right\rfloor - \left\lfloor \frac{x}{d} \right\rfloor = \frac{y}{d} + r(d),$$

where $|r(d)| \le 1$.

The proof begins in earnest with the observation that

$$S(\mathcal{A}, \mathcal{P}, z) \le \sum_{a\in A} \Bigg(\sum_{\substack{d|a\\ d|P}} \lambda(d)\Bigg)^2$$

for any real-valued function λ defined on the divisors of P with $\lambda(1)=1$. This can be seen by noting that the inner sum majorizes the characteristic function of the integers a prime to P. Expanding and reversing the order of summation shows

$$\begin{aligned}S(\mathcal{A},\mathcal{P},z) &\leq \sum_{d_1,d_2\mid P}\lambda(d_1)\lambda(d_2)\sum_{\substack{a\in\mathcal{A}\\ d_1\mid a,d_2\mid a}}1=\sum_{d_1,d_2\mid P}\lambda(d_1)\lambda(d_2)A_{[d_1,d_2]}.\\ &= X\sum_{d_1,d_2\mid P}\frac{\lambda(d_1)\lambda(d_2)}{d_1d_2}+\sum_{d_1,d_2\mid P}\lambda(d_1)\lambda(d_2)r([d_1,d_2])=XQ+R,\end{aligned}$$

say. We think of XQ as the main term and R as an error term we wish to control, and we attempt to choose λ so that Q is small. In order to keep the error term R in check, we tack on the requirement that $\lambda(d)=0$ for $d>w$, where $w\geq 1$ is a parameter to be chosen later.

As things stand, Q is a homogeneous degree 2 polynomial (quadratic form) in the $\lambda(d)$ (for $d\mid P$). We would like to make this a diagonal quadratic form, because we know how to minimize those (by Lemma 4.3.2). To this end, we rewrite

$$Q=\sum_{d_1,d_2\mid P}\frac{\lambda(d_1)\lambda(d_2)}{d_1d_2}\gcd(d_1,d_2),$$

and use the identity

$$n=\sum_{k\mid n}\phi(k) \tag{4.22}$$

to reexpress

$$\begin{aligned}Q&=\sum_{d_1,d_2\mid P}\frac{\lambda(d_1)\lambda(d_2)}{d_1d_2}\sum_{k\mid(d_1,d_2)}\phi(k)\\ &=\sum_{k\mid P}\phi(k)\sum_{\substack{d_1,d_2\mid P\\ k\mid d_1,d_2}}\frac{\lambda(d_1)\lambda(d_2)}{d_1d_2}=\sum_{k\mid P}\phi(k)\Bigg(\sum_{\substack{d\mid P\\ k\mid d}}\frac{\lambda(d)}{d}\Bigg)^2=\sum_{k\mid P}\phi(k)y(k)^2,\end{aligned} \tag{4.23}$$

where

$$y(k):=\sum_{\substack{d\mid P\\ k\mid d}}\frac{\lambda(d)}{d}. \tag{4.24}$$

Note that while we defined y in terms of λ, we can recover λ from y, since Lemma 4.3.3 applied to the divisor closed set $\mathcal{D}=\{d\mid P\}$ shows

$$\frac{\lambda(k)}{k}=\sum_{\substack{d\mid P\\ k\mid d}}\mu(d/k)y(d). \tag{4.25}$$

The reverse implication of the same lemma shows that given any values $y(k)$ (for $k \mid P$), there is way of defining $\lambda(d)$ (for $d \mid P$) which yields the given values $y(k)$, namely, just define λ by (4.25).

The constraint on λ that $\lambda(1) = 1$ is equivalent, by (4.25), to the constraint

$$\sum_{d \mid P} \mu(d) y(d) = 1,$$

while (4.25) and (4.24) together imply that $\lambda(d) = 0$ for $d > w$ if and only if $y(d) = 0$ for $d > w$. We thus want to minimize the diagonal form

$$\sum_{\substack{d \mid P \\ d \leq w}} \phi(d) y(d)^2, \quad \text{given} \quad \sum_{\substack{d \mid P \\ d \leq w}} \mu(d) y(d) = 1,$$

and we are in the exact situation of Lemma 4.3.2. The minimum value is $1/D$, where

$$D = D(z, w) := \sum_{d \mid P, d \leq w} \frac{\mu(d)^2}{\phi(d)} = \sum_{d \mid P, d \leq w} \frac{1}{\phi(d)},$$

and this minimum is attained (uniquely, in fact) when

$$y(d) = \frac{\mu(d)}{D \phi(d)} \qquad (d \mid P, d \leq w),$$

hence when (see (4.25))

$$\lambda(k) = \frac{k}{D} \sum_{\substack{d \mid P, k \mid d \\ d \leq w}} \frac{\mu(d/k) \mu(d)}{\phi(d)} \tag{4.26}$$

for $k \mid P, k \leq w$. So choose $\lambda(k)$ according to (4.26) for $k \mid P, k \leq w$ and define $\lambda(k) = 0$ for $k \mid P, k > w$. Actually, as (4.26) already vanishes for $k > w$, we can express our choice more succinctly by saying λ is always given by (4.26).

To complete our bound on $S(\mathcal{A}, \mathcal{P}, z)$ it is also necessary to estimate R. Since $|r| \leq 1$,

$$|R| \leq \sum_{d_1, d_2 \mid P} |\lambda(d_1)||\lambda(d_2)| \leq \left(\sum_{k \mid P} |\lambda(k)| \right)^2 = \left(\sum_{k \mid P, k \leq w} |\lambda(k)| \right)^2.$$

But

$$|\lambda(k)| = \left| \frac{k}{D} \sum_{\substack{d \mid P, k \mid d \\ d \leq w}} \frac{\mu(d/k) \mu(d)}{\phi(d)} \right| \leq \frac{k}{D} \sum_{\substack{d \mid P, k \mid d \\ d \leq w}} \frac{1}{\phi(d)} \tag{4.27}$$

$$\leq \frac{k}{D \phi(k)} \sum_{d' \mid P, d' \leq w} \frac{1}{\phi(d')} = \frac{k}{\phi(k)}, \tag{4.28}$$

where to transition between (4.27) and (4.28), we have written $d = kd'$ and noted that since d is squarefree, necessarily $(k, d') = 1$, so that $\phi(d) = \phi(k)\phi(d')$. Substituting above shows

$$|R| \leq \sum_{k \mid P, k \leq w} \frac{k^2}{\phi(k)^2} \leq w^2 \sum_{k \mid P, k \leq w} \frac{1}{\phi(k)^2} \leq w^2 D^2 = (wD)^2. \tag{4.29}$$

Since $Q = 1/D$ for our choice of λ, we have shown:

Lemma 4.3.4. *In the sieving situation of this section,*

$$S(\mathcal{A}, \mathcal{P}, z) \leq XQ + R \leq X/D + (wD)^2. \tag{4.30}$$

We now need an estimate for D:

Lemma 4.3.5. *Let $D = D(z,w) = \sum_{d \mid P, d \leq w} 1/\phi(d)$. Then for $w \geq z \geq 2$, we have*

$$\log z \leq D(z,w) \leq C \log z$$

for some absolute constant C.

Proof. For the upper bound, observe

$$D \leq \sum_{d \mid P} \frac{1}{\phi(d)} \leq \prod_{p \leq z} (1 + 1/\phi(p)) = \prod_{p \leq z} (1 - 1/p)^{-1} \ll \log z,$$

by Mertens' theorem. The lower bound comes from noting that

$$\begin{aligned} D(z,w) \geq D(z,z) &= \sum_{d \mid P, d \leq z} \frac{1}{\phi(d)} \\ &= \sum_{\substack{d \leq z \\ \mu(d) \neq 0}} \prod_{p \mid d} \left(\frac{1}{p} + \frac{1}{p^2} + \dots \right) \geq \sum_{d \leq z} \frac{1}{d} \geq \log z. \end{aligned}$$

Note that because $\mathcal{P}$ is the set of all primes, $d \mid P$ whenever the squarefree number d satisfies $d \leq z$. □

It is now a simple matter obtain the stated bound on $\pi(x+y) - \pi(y)$:

Proof of Lemma 4.3.1. By Lemmas 4.3.4 and 4.3.5,

$$S(\mathcal{A}, \mathcal{P}, z) \leq X/D + (wD)^2 \leq y/\log z + O(w^2 \log^2 z).$$

Taking $z = w = y^{1/2}/\log^2 y$, we find

$$S(\mathcal{A}, \mathcal{P}, z) \leq \frac{y}{\frac{1}{2}\log y - 2\log\log y} + O\left(\frac{y}{\log^2 y}\right) \leq (2 + o(1))\frac{y}{\log y}.$$

as $y \to \infty$. Since $z = o(y/\log y)$, the proof is completed by recalling

$$\pi(x+y) - \pi(x) =\leq z + S(\mathcal{A}, \mathcal{P}, z).$$ □

4.3.3 A General Version of Selberg's Sieve (optional)

Now consider a more general sifting situation: Let $\mathcal{A}$ be a finite sequence of integers, $\mathcal{P}$ a finite set of primes, and suppose

$$A_d := |\{a \in \mathcal{A} : d \mid a\}| = X\frac{\omega(d)}{d} + r(d) \quad (d \mid P),$$

where X is an approximation to the number of elements of $\mathcal{A}$, ω is a multiplicative function taking positive values on divisors of $P := \prod_{p \in \mathcal{P}} p$, and r is a function defined on the divisors of P which we think of as a remainder term. Moreover, suppose this remainder term satisfies

$$|r(d)| \leq \omega(d) \quad (d \mid P),$$

as is the case in the applications of the last chapter.

We leave it as an exercise to adapt our former argument to prove:

Theorem 4.3.6 (Selberg's Upper Bound Sieve). *For each $w \geq 1$,*

$$S(\mathcal{A}, \mathcal{P}) \leq X/D + (wD)^2,$$

where

$$D = D(P, w) = \sum_{d \mid P, d \leq w} \frac{1}{f(d)},$$

and

$$f(k) = \sum_{d \mid k} \frac{d\mu(k/d)}{\omega(d)}.$$

It may help to note the identity

$$g(a)g(b) = g((a,b))g([a,b]),$$

valid for any multiplicative function g.

Remark. If we compare the theorem with Lemma 4.3.4, we see our f is playing the former role of ϕ. The explanation for this is that diagonalizing our quadratic form by our previous procedure now requires a function f with

$$\frac{k}{\omega(k)} = \sum_{d \mid k} f(d) \quad (k \mid P).$$

In the special case considered before, $\omega(k) \equiv 1$, and we simply 'recalled' the well-known identity (4.22). In general there is no analogous well-known identity; this is not problematic, as f can be directly constructed by Möbius inversion (appropriately generalized to pairs of functions defined on arbitrary divisor-closed sets).

The next exercise gives an important application refining and generalizing Lemma 4.3.1 to primes in arithmetic progressions:

Exercise 4.3.1 *(Brun-Titchmarsh).*

a) Prove that there exists a positive absolute constant c with the property that if $y/q \geq c$, then

$$\pi(x+y;q,a)-\pi(x;q,a) \leq \frac{2y}{\phi(q)\log\frac{y}{q}} + O\left(\frac{y}{\phi(q)}\frac{\log\log\frac{y}{q}}{\log^2\frac{y}{q}}\right)$$

uniformly for $\gcd(a,q)=1$ and $q \geq 1$.

Suggestions: Take

$$\mathcal{A} = \{x < n \leq x+y : n \equiv a \pmod q\}, \quad \mathcal{P} = \{p \leq z : p \nmid q\};$$

then with $X = y/q$, one has $A_d = X/d + r(d)$ with $|r(d)| \leq 1$ for each $d \mid P$. (So the hypotheses of Theorem 4.3.6 are satisfied with $\omega \equiv 1$.) Arguing as before, show that (with an absolute constant C)

$$\sum_{\substack{d\leq z\\ \gcd(d,q)=1}} \frac{1}{d} \leq D = D(P,w) \leq C\log z \quad (2 \leq z \leq w),$$

and note that the left hand side can be estimated from below by observing

$$\frac{q}{\phi(q)}\sum_{\substack{d\leq z\\ \gcd(d,q)=1}}\frac{1}{d} = \prod_{p|q}\left(1-\frac{1}{p}\right)^{-1}\sum_{\substack{d\leq z\\ \gcd(d,q)=1}}\frac{1}{d} \geq \sum_{d'\leq z}\frac{1}{d'} \geq \log z.$$

Therefore

$$S(\mathcal{A},\mathcal{P}) \leq \frac{X}{\frac{\phi(q)}{q}\log z} + O(w^2\log^2 z) = \frac{y}{\phi(q)\log z} + O(w^2\log^2 z);$$

now choose $z = w = X^{1/2}/\log^2 X$.

b) Use the result of a) to prove that

$$\pi(x+y;q,a)-\pi(x;q,a) \ll \frac{y}{\phi(q)\log\frac{y}{q}},$$

with an absolute implied constant, uniformly for $(a,q)=1$ and $y > q$. In particular (taking $x = 0$),

$$\pi(y;q,a) \ll \frac{y}{\phi(q)\log\frac{y}{q}}$$

for $(a,q)=1$ and $y > q$.

Hint: the result is trivial (why?) if y/q is absolutely bounded, so you may assume y/q is large.

Exercise 4.3.2. Define the *natural density* of a set of primes $\mathcal{P}$ (relative to the set of all primes) as the limit

$$\lim_{x\to\infty} |\mathcal{P}\cap[1,x]| \Big/ \frac{x}{\log x},$$

if it exists. Then the prime number theorem asserts that the set of all primes has density 1, and the PNT for arithmetic progressions asserts that the set of primes $p \equiv a \pmod q$ has density $1/\phi(q)$ whenever $\gcd(a,q)=1$.

Assuming the PNT for arithmetic progressions, prove that the set $\mathcal{P}$ of primes p for which $p-1$ is squarefree has density $\prod_q(1-\frac{1}{q(q-1)})$, where the product is over all primes q.

Hint: Approximate $\mathcal{P}$ (from "above") by the sets $\mathcal{P}_z$, where $\mathcal{P}_z$ is the set of primes p for which $p-1$ is not divisible by q^2 for any $q \le z$. Use the PNT for APs to compute the density of each $\mathcal{P}_z$ and use the Brun-Titchmarsh inequality to bound above the counting function of $\mathcal{P}_z \setminus \mathcal{P}$.

Our discussion of Selberg's sieve follows Ben Green's excellent notes [Gre]. Green gives a number of interesting applications, including an application to small prime gaps and another derivation of the upper bounds for $r(n)$ and $\pi_N(x)$ obtained (by the Brun-Hooley sieve) in the last chapter. In addition to the general references cited in Chapter 3, we recommend to the reader the particularly pithy introductory account of Selberg's sieve given in [Elk03].

4.4 The Turán-Kubilius Inequality

Recall that an arithmetic function f is called *additive* if $f(ab)=f(a)+f(b)$ whenever a and b are relatively prime and is called strongly additive if in addition $f(p^k)=f(p)$ for every prime power p^k (p prime, k a positive integer). Note that f is strongly additive if and only if

$$f(n)=\sum_{p\mid n} f(p)$$

for each positive integer n.

Lemma 4.4.1 (Turán-Kubilius Inequality [Kub62]). *Let f be a strongly additive real-valued function. Then*

$$\sum_{n\le x}\left| f(n)-\sum_{p\le x}\frac{f(p)}{p}\right|^2 \ll x\sum_{p\le x}\frac{|f(p)|^2}{p} \tag{4.31}$$

for every real x. Here the implied constant is absolute.

Proof. It is convenient to introduce the notation

$$A(x) := \sum_{p\le x} p^{-1}f(p), \quad B(x) := \left(\sum_{p\le x} p^{-1}|f(p)|^2\right)^{1/2},$$

so that the desired inequality becomes

$$\sum_{n \le x} |f(n) - A(x)|^2 \ll xB(x)^2.$$

It is no loss of generality to assume that f is nonnegative. To see this, assume (4.31) known for nonnegative functions, and introduce f_1, f_2 defined by

$$f_1(p) = \begin{cases} f(p) & \text{if } f(p) \ge 0, \\ 0 & \text{if } f(p) < 0, \end{cases} \qquad f_2(p) = \begin{cases} 0 & \text{if } f(p) \ge 0, \\ -f(p) & \text{if } f(p) < 0, \end{cases}$$

and extended to be strongly additive. Let A_i and B_i be defined as A and B were above, but with f replaced by f_i ($i = 1, 2$). Then f_1, f_2 are nonnegative strongly additive functions, $f = f_1 - f_2$, and

$$\begin{aligned} \sum_{n \le x} |f(n) - A(x)|^2 &= \sum_{n \le x} |(f_1(n) - A_1(x)) - (f_2(n) - A_2(x))|^2 \\ &\le 2 \sum_{i=1}^{2} \sum_{n \le x} |f_i(n) - A_i(x)|^2 \\ &\ll x(B_1(x)^2 + B_2(x)^2) = xB(x)^2, \end{aligned}$$

so that (4.31) follows in general (in fact the proof shows we may take twice the original implied constant).

We now proceed to the proof, under the assumption that f is nonnegative. We may assume x is an integer and at least 2. Expanding out the square shows the left hand side of (4.31) is given by

$$\sum_{n \le x} f(n)^2 - 2A(x) \sum_{n \le x} f(n) + xA(x)^2.$$

Let us examine these terms in turn. With p, q denoting primes, we have

$$\begin{aligned} \sum_{n \le x} f(n)^2 &= \sum_{n \le x} \Big(\sum_{p \mid n} f(p) \Big) \Big(\sum_{q \mid n} f(q) \Big) \\ &= \sum_{p,q \le x} f(p) f(q) \sum_{\substack{n \le x \\ p,q \mid n}} 1 \\ &= \sum_{\substack{pq \le x \\ p \ne q}} f(p) f(q) \left\lfloor \frac{x}{pq} \right\rfloor + \sum_{p \le x} f(p)^2 \left\lfloor \frac{x}{p} \right\rfloor \\ &\le x \sum_{\substack{pq \le x \\ p \ne q}} \frac{f(p) f(q)}{pq} + x \sum_{p \le x} \frac{f(p)^2}{p} \le xA(x)^2 + xB(x)^2. \end{aligned}$$

In a similar manner, we find

$$\begin{aligned}\sum_{n\le x} f(n) &= \sum_{p\le x} f(p)\left\lfloor \frac{x}{p}\right\rfloor \\ &\ge x\sum_{p\le x}\frac{f(p)}{p} - \sum_{p\le x} f(p) = xA(x) - \sum_{p\le x} f(p),\end{aligned}$$

so that

$$-2A(x)\sum_{n\le x} f(n) \le -2xA(x)^2 + 2A(x)\sum_{p\le x} f(p).$$

Adding our three terms reveals

$$\sum_{n\le x} |f(n) - A(x)|^2 \le xB(x)^2 + 2A(x)\sum_{p\le x} f(p).$$

Now matters can be settled by using the Cauchy-Schwarz inequality to show the second term on the right hand side is $O(xB(x)^2)$. Indeed,

$$\begin{aligned}2A(x)\sum_{p\le x} f(p) &\ll \sum_{p\le x}\frac{f(p)}{p}\sum_{p\le x} f(p)\\ &= \sum_{p\le x}\frac{f(p)}{p^{1/2}}\frac{1}{p^{1/2}}\sum_{p\le x}\frac{f(p)}{p^{1/2}}p^{1/2}\\ &\ll \left(\sum_{p\le x}\frac{f(p)^2}{p}\right)^{1/2}\left(\sum_{p\le x}\frac{1}{p}\right)^{1/2}\left(\sum_{p\le x}\frac{f(p)^2}{p}\right)^{1/2}\left(\sum_{p\le x} p\right)^{1/2}\\ &\ll B(x)\left(\sum_{2\le n\le x}\frac{1}{n}\right)^{1/2} B(x)\left(x\pi(x)\right)^{1/2} \ll xB(x)^2\end{aligned}$$

on recalling the estimates

$$\sum_{2<n\le x} 1/n \ll \log x, \qquad \pi(x) \ll x/\log x \quad (x\ge 2). \qquad \square$$

Exercise 4.4.1. Show that the same estimate holds, with the same implied constant, if f is complex-valued.

4.4.1 Exercises: The Orders of ν and Ω (optional)

Recall that $\nu(n)$ denotes the number of distinct prime factors of n and $\Omega(n)$ denotes the number of prime factors of n counted with multiplicity. Thus if $n = p_1^{e_1}p_2^{e_2}\dots p_k^{e_k}$ (with distinct primes), then

$$\nu(n) = k, \quad \Omega(n) = e_1 + e_2 + \cdots + e_k.$$

Note that ν is strongly additive (in the sense introduced above) and Ω is completely additive:

$$\Omega(mn) = \Omega(m) + \Omega(n)$$

for any positive integers m, n.

The first inequality of Turán-Kubilius type appears in [Tur34], in the guise of the particular estimate

$$\sum_{n \le x} (\nu(n) - \log\log x)^2 = O(x \log\log x). \tag{4.32}$$

In the next few exercises we prove this estimate and explore some of its consequences, ending with a proof of Erdős' Multiplication Table Theorem.

Exercise 4.4.2 *(Mean Values of ν, Ω).* Show that as $x \to \infty$, one has

$$\sum_{n \le x} \nu(n) = x \log\log x + C_1 x + O(x/\log x),$$
$$\sum_{n \le x} \Omega(n) = x \log\log x + C_2 x + O(x/\log x)$$

for some constants C_1, C_2.

Exercise 4.4.3. Use the Turán-Kubilius inequality to establish (4.32).

One implication of (4.32) is that both ν and Ω seldom differ substantially from this average value. This is quantified in the following exercise:

Exercise 4.4.4. Let $\kappa(x)$ be a function tending to infinity with x. Using (4.32), show that

$$|\{n \le x : |\nu(n) - \log\log x| > \kappa(x)(\log\log x)^{1/2}\}| = o(x) \quad (x \to \infty). \tag{4.33}$$

Using the estimate

$$\sum_{n \le x} (\Omega(n) - \omega(n)) \ll x,$$

which derives from 4.4.2, deduce

$$|\{n \le x : \Omega(n) - \nu(n) > (\log\log x)^{1/2}\}| = O\left(x/(\log\log x)^{1/2}\right) = o(x).$$

Conclude that (4.33) remains valid with ν replaced by Ω.

Exercise 4.4.5 *(Erdős [Erd60]).* Let $M(N)$ denote the number of distinct elements in the $N \times N$ multiplication table: i.e.,

$$M(N) = |\{mn : 1 \le m, n \le N\}|.$$

Thus $M(N) \le N^2$ trivially, and in fact commutativity of multiplication yields the stronger bound

$$M(N) \le \binom{N}{2} + N = \frac{N^2 + N}{2},$$

so that $\limsup_{N \to \infty} M(N)/N^2 \le 1/2$. Here we use the result of the last exercise to show $M(N)/N^2 \to 0$.

a) From the previous exercise we have

$$\left|\left\{n \le N : |\Omega(n) - \log\log N| > \frac{1}{4}\log\log N\right\}\right| = o(N).$$

Using this result, show that

$$\left|\left\{(a,b) : 1 \le a, b \le N, \text{ and } |\Omega(ab) - 2\log\log N| > \frac{1}{2}\log\log N\right\}\right| = o(N^2).$$

b) On the other hand, show that

$$\left|\left\{n \le N^2 : |\Omega(ab) - \log\log N| > \frac{1}{4}\log\log N\right\}\right| = o(N^2).$$

c) By combining a) and b), prove that $M(N) = o(N^2)$.

Exercise 4.4.6 *(Schroeppel).* Establish the elementary lower bound

$$M(N) \gg N^2/\log N \quad (N \to \infty).$$

Suggestion: Consider products pm, where $N/2 < p \le N$ is prime and $m \le N$ is arbitrary. Keep in mind that $\pi(N) - \pi(N/2) \gg N/\log N$, as proved in §1.5.

4.5 Hildebrand's Lemmata

4.5.1 Preliminary Lemmas

The upper bound of Lemma 4.3.1 quickly translates into an upper bound for the sum of $\log p/p$ taken over intervals $(x, x+y]$:

Corollary 4.5.1. *Let $B > 1$ be a fixed positive real number. As $y \to \infty$, we have*

$$\sum_{x<p\le x+y} \frac{\log p}{p} \le (2 + o(1)) \log \frac{x+y}{y},$$

uniformly for $y \le x \le By$.

Proof. Assuming k large, we partition the interval $(x, x+y]$ into $k = k(y) = \lfloor \log y \rfloor$ subintervals

$$I_j := (x_j, x_j + y/k], \quad x_j = x + jy/k \qquad (j = 0, \ldots, k-1).$$

Since $\log t/t$ is eventually decreasing, as $y \to \infty$ we have

$$\begin{aligned}\sum_{x<p\le x+y} \frac{\log p}{p} &\le \sum_{j=0}^{k-1} \frac{\log x_j}{x_j} \sum_{p\in I_j} 1 \\ &\le (2 + o(1)) \sum_{j=0}^{k-1} \frac{\log x_j}{x_j} \frac{y/k}{\log(y/k)}.\end{aligned}$$

But since

$$y \le x \le x_j \le x + y \le (B+1)y,$$

one has

$$\log x_j \le \log (B+1)y = (1+o(1)) \log y,$$
$$\log (y/k) \ge \log y - \log\log y \ge (1+o(1)) \log y,$$

which leads to the estimate

$$\sum_{x<p\le x+y} \frac{\log p}{p} \le (2+o(1)) \sum_{j=0}^{k-1} \frac{1}{x+jy/k}\,\frac{y}{k}. \tag{4.34}$$

The temptation to compare this Riemann sum to the corresponding integral is one we cannot resist. Keeping in mind that $x + jy/k$ is a decreasing function of $j \ge 0$ (once x and y are given), we see

$$\begin{aligned}\sum_{j=0}^{k-1} \frac{1}{x+jy/k} &= \frac{y}{kx} + \sum_{j=1}^{k-1} \frac{1}{x+jy/k}\\ &\le \frac{y}{kx} + \int_x^{x+y\frac{k-1}{k}} \frac{dt}{t}\\ &\le \frac{y}{kx} + \int_x^{x+y} \frac{dt}{t} = \frac{y}{kx} + \log\frac{x+y}{y}.\end{aligned}$$

Since $y/kx \le 1/k = o(1)$ and $\log \frac{x+y}{y} \ge \log 2$,

$$\sum_{j=0}^{k-1} \frac{1}{x+jy/k} \le y/kx + \log\frac{x+y}{y} = (1+o(1)) \log\frac{x+y}{y}.$$

Substituting into (4.34) finishes the proof. $\square$

Lemma 4.5.2. *Let $\epsilon > 0$, and let $x, x' \ge 3$. There exists λ, $1 \le \lambda \le \lambda_0$, for which*

$$\sum_{y<p\le(1+\epsilon)y} \frac{\log p}{p} \ge \delta \tag{4.35}$$

for both $y = \lambda x$ and $y = \lambda x'$. Here $\delta = \delta(\epsilon) > 0$ and $\lambda_0 = \lambda_0(\epsilon)$ is a positive integer.

Remark. It is a consequence of the prime number theorem that actually (4.35) holds for all large y, for an appropriate $\delta = \delta(\epsilon)$. For example, this follows from the result of Exercise 4.2.3 (though there are more direct demonstrations).

Proof. We may suppose $0 < \epsilon \le 1$. We let $\epsilon_1 := \epsilon/3$, and we define the sequence $(x_i)_{i\ge0}$ by $x_i = x(1+\epsilon_1)^i$. We let δ be a positive parameter whose value will be chosen later, and we define $I = I(\delta, x)$ as the set of nonnegative indices for which

$$\sum_{x_i < p \le (1+\epsilon_1)x_i} \frac{\log p}{p} \ge \delta. \tag{4.36}$$

We also define

$$\overline{I} := \{i \ge 0 : i \in I \text{ or } i+1 \in I\}.$$

Note that because

$$(1+\epsilon_1)^2 = 1 + 2\epsilon_1 + \epsilon_1^2 \le 1 + \epsilon,$$

the inequality (4.35) holds for $y = x_i$ whenever $i \in \overline{I}$.

For an interval $A \subset \mathbf{R}$, we define

$$N(A) := |A \cap I|, \qquad \overline{N}(A) := |A \cap \overline{I}|.$$

Our strategy will be as follows: we will show that if δ is sufficiently small, depending only on ϵ, then we can find a positive integer $i_0 = i_0(\epsilon)$ with the property that $\overline{N}([0, i_0)) > i_0/2$, regardless of the particular $x \ge 3$ with respect to which $\overline{I}$ is defined. Applying this observation to the given x and x', we see there must exist an $i, 0 \le i < i_0$, with $i \in \overline{I} \cap \overline{I}'$, where $\overline{I}$ is defined with respect to x and where $\overline{I}'$ is analogously defined with respect to x'. We can then take $\lambda_0 = (1+\epsilon_1)^{i_0}$ and $\lambda = (1+\epsilon_1)^i$ in the statement of the lemma.

To carry out this strategy we start by looking for a lower bound on $N([i, i+g))$, where i, g are integers with $i \ge 0$ and $g \ge 1$. By a result of Mertens' (see (1.38)),

$$\sum_{x_i < p \le x_{i+g}} \frac{\log p}{p} = g \log \frac{x_{i+g}}{x_i} + O(1) = g \log(1+\epsilon_1) + O(1), \tag{4.37}$$

with absolute implied constants. But by Corollary 4.5.1 with $B = \epsilon_1^{-1}$, we have (writing $N = N([i, i+g))$),

$$\begin{aligned}\sum_{x_i < p \le x_{i+g}} \frac{\log p}{p} &= \sum_{j=0}^{g-1} \sum_{x_{i+j} < p \le x_{i+j+1}} \frac{\log p}{p} \\ &\le 2N \log(1+\epsilon_1) + (g-N)\delta + o(g) \\ &\le 2N \log(1+\epsilon_1) + g\delta + o(g),\end{aligned} \tag{4.38}$$

where the term $o(g)$ describes the behavior as $g \to \infty$, and is uniform in both $x \ge 3$ and $i \ge 0$. Comparing (4.37) and (4.38), we deduce

$$N([i, i+g)) \ge g\left(\frac{1}{2} - \frac{\delta}{2\log(1+\epsilon_1)} + o(1) + O\left(\frac{1}{g\log(1+\epsilon_1)}\right)\right).$$

If $\delta \le \epsilon/6$, the right hand side becomes positive for sufficiently large g (for any choice of δ in this range and any value of $i \ge 0$), since

$$\log(1+\epsilon_1) = \int_1^{1+\epsilon_1} \frac{dt}{t} \ge \epsilon_1 \frac{1}{1+\epsilon_1} = \frac{\epsilon}{3+\epsilon} \ge \frac{\epsilon}{4}.$$

Fix a positive integer $g = g(\epsilon)$ large enough that this is so. Then we have shown every interval $[i, i+g)$ (with $i \geq 0$) contains an element of I, provided $\delta \leq \epsilon/6$. We assume this inequality on δ holds in what follows.

Now suppose $i_0 > 6g$. We consider two cases, according as the inequality

$$\overline{N}([0, i_0)) - N([0, i_0)) \leq i_0/3g.$$

holds or not. If it does, then there are at most $i_0/3g + 1$ blocks of consecutive indices $i \leq i_0$ not belonging to I (since the endpoints of all such blocks not terminating at i_0 are counted by the left hand side). But by what was said above, any such block has length at most g. It follows that

$$\overline{N}([0, i_0)) \geq N([0, i_0)) \geq i_0 - g(i_0/3g + 1) > i_0/2.$$

Now suppose the inequality fails. Then

$$\begin{aligned} \overline{N}([0, i_0)) &> N([0, i_0) + \frac{i_0}{3g} \\ &\geq i_0 \left(\frac{1}{2} + \frac{1}{3g} - \frac{\delta}{2\log(1+\epsilon_1)} + o(1) + O\left(\frac{1}{i_0 \log(1+\epsilon_1)} \right) \right). \end{aligned}$$

In this case we see that if $\delta = \delta(\epsilon)$ is chosen sufficiently small, then $\overline{N}([0, i_0)) > i_0/2$ once i_0 is large. Any such large i_0 (that also exceeds $6g$) completes the proof, by our earlier remarks. □

4.5.2 The Fundamental Lemma

The proof that $M^*(x) = o(1)$ is achieved by means of our next result, whose proof occupies the entire next section:

Fundamental Lemma. *Let $\eta(x)$ be a nonnegative function with $\eta(x) \to 0$ as $x \to \infty$. Then*

$$M^*(x) - M^*(x') = o(1)$$

as $x \to \infty$, uniformly for $x \leq x' \leq x^{1+\eta(x)}$.

Proof of the Prime Number Theorem (in the form $M^*(x) = o(1)$). Let $\eta(x) = (\log x)^{-1/2}$. Now

$$\int_x^{x^{1+\eta}} \frac{M^*(x')}{x'} \, dx' = \int_x^{x^{1+\eta}} \frac{M^*(x) + o(1)}{x'} \, dx' = (\eta \log x) M^*(x) + o(\eta \log x),$$

so that

$$\begin{aligned} M^*(x) &= \frac{1}{\eta \log x} \int_x^{x^{1+\eta}} \frac{M^*(x')}{x'} \, dx' + o(1) \\ &= \frac{1}{\eta \log x} \int_x^{x^{1+\eta}} \frac{1}{x'^2} \sum_{n \leq x'} \mu(n) \, dx' + o(1) \\ &= \frac{1}{\eta \log x} \sum_{n \leq x'} \mu(n) \int_{\max\{x,n\}}^{x^{1+\eta}} \frac{dx'}{x'^2} + o(1) \end{aligned}$$

Evaluating the integral, this becomes

$$\begin{aligned}
&\frac{1}{\eta \log x} \sum_{n \le x'} \mu(n) \left(\frac{1}{\max\{x,n\}} - \frac{1}{x^{1+\eta}} \right) + o(1) \\
&\qquad = \frac{1}{\eta \log x} \sum_{x < n \le x'} \frac{\mu(n)}{n} - \frac{M^*(x')}{\eta (\log x) x^{1+\eta}} + \frac{M^*(x)}{\eta x \log x} + o(1) \\
&\qquad = \frac{1}{\eta \log x} \sum_{x < n \le x'} \frac{\mu(n)}{n} + o(1),
\end{aligned}$$

but the final sum here is bounded (see Lemma 4.2.3), so the entire expression is $o(1)$. Note that the choice $\eta = \log^{-1/2} x$ is made to ensure $\eta \log x \to \infty$. $\square$

4.6 Proof of The Fundamental Lemma

4.6.1 Preparation

The next lemma provides the identity on which Hildebrand's proof is based. It is a consequence of the Turán-Kubilius inequality.

Lemma 4.6.1. *Let $\mathcal{P}$ be a set of primes $p \le x$, where $x > 0$. Then*

$$\left(\sum_{p \in \mathcal{P}} \frac{1}{p} \right) M^*(x) = - \sum_{p \in \mathcal{P}} \frac{M^*(x/p)}{p} + O\left(\left(\sum_{p \in \mathcal{P}} \frac{1}{p} \right)^{1/2} \right), \tag{4.39}$$

where the implied constant is absolute.

Proof. Let p be any prime. Then for nonnegative x,

$$\begin{aligned}
\sum_{\substack{n \le x \\ p \mid n}} \mu(n) &= \sum_{pn' \le x} \mu(pn') = \mu(p) \sum_{\substack{n' \le x/p \\ p \nmid n'}} \mu(n') \\
&= -\left(\sum_{n' \le x/p} \mu(n') + O(x/p^2) \right) = -\frac{x}{p} M^*(x/p) + O(x/p^2).
\end{aligned}$$

Now divide by x and sum over $p \in \mathcal{P}$. Reversing the order of summation of the left hand side then yields

$$\begin{aligned}
\frac{1}{x} \sum_{n \le x} \mu(n) \nu_{\mathcal{P}}(n) &= \frac{1}{x} \sum_{p \in \mathcal{P}} \sum_{\substack{n \le x \\ p \mid n}} \mu(n) \\
&= - \sum_{p \in \mathcal{P}} \frac{M^*(x/p)}{p} + O\left(\sum_{p \in \mathcal{P}} \frac{1}{p^2} \right).
\end{aligned}$$

Subtracting this from the obvious relation

$$\frac{1}{x}\sum_{n\le x}\mu(n)\sum_{p\in\mathcal{P}}\frac{1}{p} = M^*(x)\sum_{p\in\mathcal{P}}\frac{1}{p},$$

we find

$$M^*(x)\sum_{p\in\mathcal{P}}\frac{1}{p}+\sum_{p\in\mathcal{P}}\frac{1}{p}M^*(x/p) = -\frac{1}{x}\sum_{n\le x}\mu(n)\left(\nu_{\mathcal{P}}(n)-\sum_{p\in\mathcal{P}}\frac{1}{p}\right)+O\Big(\sum_{p\in\mathcal{P}}\frac{1}{p^2}\Big).$$

By the Cauchy-Schwarz and Turán-Kubilius Inequalities (the latter applied with $f=\nu_{\mathcal{P}}$), we see the first term on the right hand side is

$$\ll \frac{1}{x}x^{1/2}\left(\sum_{n\le x}\left|\nu_{\mathcal{P}}(n)-\sum_{p\in\mathcal{P}}\frac{1}{p}\right|^2\right)^{1/2}$$
$$\ll \frac{1}{x}x^{1/2}x^{1/2}\Big(\sum_{p\in\mathcal{P}}\frac{1}{p}\Big)^{1/2} = \Big(\sum_{p\in\mathcal{P}}\frac{1}{p}\Big)^{1/2}.$$

Since also

$$\sum_{p\in\mathcal{P}}\frac{1}{p^2} = \sum_{p\in\mathcal{P}}\frac{1}{p^{1/2}}\frac{1}{p^{3/2}} \ll \Big(\sum_{p\in\mathcal{P}}\frac{1}{p}\Big)^{1/2}\Big(\sum_{p\in\mathcal{P}}\frac{1}{p^3}\Big)^{1/2} \ll \Big(\sum_{p\in\mathcal{P}}\frac{1}{p}\Big)^{1/2},$$

the lemma follows. □

We also require a trivial bound on the variation of M^*:

Lemma 4.6.2. *For $x\ge 1$ and $y\ge 0$, we have*

$$M^*(x+y)-M^*(x) \ll \frac{y+1}{x}.$$

Proof. Write

$$M^*(x+y)-M^*(x) = \frac{1}{x+y}\sum_{n\le x+y}\mu(n)-\frac{1}{x}\sum_{n\le x}\mu(n)$$
$$\ll \frac{y}{x(x+y)}\sum_{n\le x}|\mu(n)|+\frac{1}{x+y}\sum_{x<n\le x+y}|\mu(n)| \ll \frac{y}{x+y}+\frac{y+1}{x+y} \ll \frac{y+1}{x}. \quad \square$$

4.6.2 Construction of $\mathcal{P},\mathcal{P}'$

We now begin the proof of the Fundamental Lemma.

Fix $\epsilon>0$. It suffices to prove that if $\eta>0$ is sufficiently small (depending on ϵ), then

$$|M^*(x)-M^*(x')| \ll \epsilon \qquad (x\le x'\le x^{1+\eta})$$

with an absolute implied constant, whenever x is sufficiently large (depending at most on ϵ, η).

We will prove this roughly as follows: We will select sets of primes $\mathcal{P} \subset [1, x)$ and $\mathcal{P}' \subset (1, x']$. Then we have the identity (4.39), as well as a corresponding identity with $\mathcal{P}$ replaced by $\mathcal{P}'$. We will choose our sets $\mathcal{P}, \mathcal{P}'$ so that the right hand sides of these identities roughly cancel each other, while the sum S of the reciprocals of the primes in $\mathcal{P}$ is roughly the same (large number) as the sum of the reciprocals of the primes in $\mathcal{P}'$, say S'. Subtracting one identity from the other, and dividing by S, we will obtain the required bound on $|M^*(x)-M^*(x')|$.

So let $0 < \eta \leq 1/2$ and suppose $3 \leq x \leq x' \leq x^{1+\eta}$. Define increasing sequences $(x_j)_{j\geq 0}$, $(x'_j)_{j'\geq 0}$ as follows: let

$$x_0 = x^{\sqrt{\eta}}, \quad x'_0 = x_0 \frac{x'}{x},$$

and for each $j \geq 0$ choose x_{j+1} in the interval

$$x_j(1+\epsilon) \leq x_{j+1} \leq \lambda_0 x_j(1+\epsilon),$$

with the property that

$$\sum_{y<p\leq(1+\epsilon)y} \frac{\log p}{p} \geq \delta \tag{4.40}$$

holds for both $y = x_{j+1}$ and $y = x'_{j+1} := x_{j+1}(x'/x)$. Here $\lambda_0 = \lambda_0(\epsilon)$ and $\delta = \delta(\epsilon)$ are the constants of Lemma 4.5.2, and the x and x' of that lemma have been taken in this application as $x_j(1+\epsilon)$ and $x'_j(1+\epsilon)$ respectively.

Define j_0 by $x_{j_0} \leq x < x_{j_0+1}$, so that $j_0 \geq 2$ if x is large (in terms of ϵ), as we will assume. Now define the intervals

$$I_j = (x_j, x_j(1+\epsilon)], \quad I'_j := (x'_j, x'_j(1+\epsilon)].$$

Then I_j, I'_j $(1 \leq j < j_0)$ are disjoint and contained in $(x_0, x]$ and $(x'_0, x']$ respectively.

Now choose sets of primes $\mathcal{P}, \mathcal{P}'$ with

$$\mathcal{P} \subset \bigcup_{1\leq j<j_0} I_j, \qquad \mathcal{P}' \subset \bigcup_{1\leq j<j_0} I'_j,$$

and

$$\left|\sum_{p\in\mathcal{P}\cap I_j} \frac{\log p}{p} - \delta\right| \leq \frac{\log(x_j(1+\epsilon))}{x_j}, \quad \left|\sum_{p\in\mathcal{P}'\cap I'_j} \frac{\log p}{p} - \delta\right| \leq \frac{\log(x'_j(1+\epsilon))}{x'_j}. \tag{4.41}$$

This is possible because originally (4.40) holds (for $y = x_j, x'_j$) and discarding a prime $p \in I_j$ shifts the value of $\sum_{p\in\mathcal{P}\cap I_j} \log p/p$ by at most $\log(x_j(1+\epsilon))/x_j$ (and similarly for I'_j).

4.6.3 Estimation of S, S'

Define

$$S := \sum_{p \in \mathcal{P}} \frac{1}{p}, \qquad S' := \sum_{p \in \mathcal{P}'} \frac{1}{p}.$$

We would like to estimate S and S', but the information we are initially given is about $\sum \log p/p$, not about $\sum 1/p$. However,

$$\sum_{1 \le j < j_0} \frac{1}{\log\left(x_j(1+\epsilon)\right)} \sum_{p \in \mathcal{P} \cap I_j} \frac{\log p}{p} \le S \le \sum_{1 \le j < j_0} \frac{1}{\log x_j} \sum_{p \in \mathcal{P} \cap I_j} \frac{\log p}{p}, \tag{4.42}$$

from which it follows that

$$\begin{aligned} S - \sum_{1 \le j < j_0} \frac{1}{\log x_j} \sum_{p \in \mathcal{P} \cap I_j} \frac{\log p}{p} &\ll \sum_{1 \le j < j_0} \left(\frac{1}{\log x_j} - \frac{1}{\log\left(x_j(1+\epsilon)\right)} \right) \sum_{p \in \mathcal{P} \cap I_j} \frac{\log p}{p} \\ &\ll \sum_{1 \le j < j_0} \frac{1}{\log x_j \log\left(x_j(1+\epsilon)\right)} \sum_{p \in \mathcal{P} \cap I_j} \frac{\log p}{p} \\ &\ll \frac{1}{\log x_0} \sum_{1 \le j < j_0} \frac{1}{\log\left(x_j(1+\epsilon)\right)} \sum_{p \in \mathcal{P} \cap I_j} \frac{\log p}{p} \\ &\ll \frac{S}{\log x_0}. \end{aligned} \tag{4.43}$$

Recalling the right-hand inequality of (4.42), we conclude

$$S = \left(1 + O\left(\frac{1}{\log x_0}\right)\right) \sum_{1 \le j < j_0} \frac{1}{\log x_j} \sum_{p \in \mathcal{P} \cap I_j} \frac{\log p}{p} \tag{4.44}$$

Similarly,

$$S' - \sum_{1 \le j < j_0} \frac{1}{\log x_j'} \sum_{p \in \mathcal{P}' \cap I_j'} \frac{\log p}{p} \ll \frac{S'}{\log x_0'} \ll \frac{S'}{\log x_0}, \tag{4.45}$$

and

$$S' = \left(1 + O\left(\frac{1}{\log x_0}\right)\right) \sum_{1 \le j < j_0} \frac{1}{\log x_j'} \sum_{p \in \mathcal{P}' \cap I_j'} \frac{\log p}{p}.$$

But for each $j, 1 \le j < j_0$,

$$\sum_{p \in \mathcal{P} \cap I_j} \frac{\log p}{p} = \delta \left(1 + O\left(\frac{1}{\delta} \frac{\log\left(x_j(1+\epsilon)\right)}{x_j}\right)\right).$$

In particular, since $\log\left(x_j(1+\epsilon)\right)/x_j \ll 1/\log x_j \ll 1/\log x_0$, there is a positive absolute constant A with

$$\delta\left(1 - \frac{A}{\log x_0}\right) \le \sum_{p \in \mathcal{P} \cap I_j} \frac{\log p}{p} \le \delta\left(1 + \frac{A}{\log x_0}\right) \quad (1 \le j < j_0),$$

so that from (4.44),

$$S = \left(1 + O\left(\frac{1}{\log x_0}\right)\right)\left(1 + O\left(\frac{1}{\delta \log x_0}\right)\right)\sum_{1\le j<j_0}\frac{\delta}{\log x_j}$$
$$= \left(1 + O\left(\frac{1}{\delta \log x_0}\right)\right)\sum_{1\le j<j_0}\frac{\delta}{\log x_j} \tag{4.46}$$

Similarly,

$$S' = \left(1 + O\left(\frac{1}{\delta \log x_0}\right)\right)\sum_{1\le j<j_0}\frac{\delta}{\log x'_j}. \tag{4.47}$$

Since

$$0 < \frac{\log x'_j - \log x_j}{\log x_j} = \frac{\log x'/x}{\log x_j} \le \frac{\eta \log x}{\log x_0} = \sqrt{\eta},$$

we have

$$1 \le \sum_{1\le j<j_0}\frac{\delta}{\log x_j} \Big/ \sum_{1\le j<j_0}\frac{\delta}{\log x'_j} \le 1 + \sqrt{\eta};$$

dividing (4.46) by (4.47), we find that for large x_0 ("large" depending on ϵ)

$$S/S' = \left(1 + O\left(\frac{1}{\delta \log x_0}\right)\right)(1 + O(\sqrt{\eta})) = \left(1 + O(\sqrt{\eta}) + O\left(\frac{1}{\delta \log x_0}\right)\right). \tag{4.48}$$

Hence if η is sufficiently small and x_0 sufficiently large (depending only on ϵ),

$$1/2 \le S/S' \le 2. \tag{4.49}$$

We will need a lower bound on S and S' in what follows. Such a bound can now be obtained from (4.46) and (4.49), which together imply that for large x_0 (depending only on ϵ),

$$\begin{aligned} S' \gg S \gg_\epsilon & \sum_{1\le j<j_0}\frac{1}{\log x_j} \\ &\ge \sum_{1\le j<j_0}\frac{1}{\log x_0 + j\log(\lambda_0(1+\epsilon))} \\ &\ge \sum_{0\le j\le j_0}\frac{1}{\log x_0 + j\log(\lambda_0(1+\epsilon))} - \frac{2}{\log x_0} \\ &\ge \int_0^{j_0+1}\frac{dt}{\log x_0 + t\log(\lambda_0(1+\epsilon))} - \frac{2}{\log x_0} \\ &\ge \frac{1}{\log(\lambda_0(1+\epsilon))}\log\frac{\log x}{\log x_0} - \frac{2}{\log x_0} \\ &\ge \frac{1}{\log(\lambda_0(1+\epsilon))}\log\frac{1}{\sqrt{\eta}} - \frac{2}{\log x_0}. \end{aligned} \tag{4.50}$$

4.6.4 Estimation of $SM^*(x) - S'M^*(x')$

Now apply Lemma 4.6.1 to $\mathcal{P}$ and to $\mathcal{P}'$ and subtract to obtain

$$\begin{aligned} &SM^*(x) - S'M^*(x') \\ &\qquad = - \sum_{1\le j<j_0} \left(\sum_{p\in\mathcal{P}\cap I_j} \frac{M^*(x/p)}{p} - \sum_{p\in\mathcal{P}'\cap I_j'} \frac{M^*(x'/p)}{p} \right) + O(\sqrt{S}). \end{aligned}$$

The error in replacing $M^*(x/p)$ with $M^*(x/x_j)$ above can be bounded using Lemma 4.6.2 as

$$\begin{aligned} &\ll \sum_{1\le j<j_0} \sum_{p\in\mathcal{P}\cap I_j} \frac{1}{p} \frac{x/x_j - x/p + 1}{x/p} = \sum_{1\le j<j_0} \sum_{p\in\mathcal{P}\cap I_j} \frac{p/x_j - 1 + p/x}{p} \\ &\quad \ll \sum_{1\le j<j_0} \sum_{p\in\mathcal{P}\cap I_j} \frac{(1+\epsilon) - 1 + p/x}{p} \ll \epsilon S + \frac{1}{x} \sum_{p\in\mathcal{P}} 1 \ll \epsilon S + O\left(\frac{1}{\log x}\right). \end{aligned}$$

Keeping in mind that $S \asymp S'$, we see the same upper estimate holds for the error incurred when replacing $M^*(x'/p)$ by $M^*(x'/x_j')$, so that

$$\begin{aligned} SM^*(x) - S'M^*(x') = - \sum_{1\le j<j_0} \frac{M^*(x/x_j)}{x_j} \left(\sum_{p\in\mathcal{P}\cap I_j} \frac{1}{p} - \sum_{p\in\mathcal{P}'\cap I_j'} \frac{1}{p} \right) \\ + O\left(\sqrt{S} + \epsilon S + \frac{1}{\log x}\right). \end{aligned}$$

Since $|M^*| \le 1$, the estimates (4.43) and (4.45) now imply

$$\begin{aligned} SM^*(x) - S'M^*(x') = - \sum_{1\le j<j_0} \frac{M^*(x/x_j)}{x_j} \left(\sum_{p\in\mathcal{P}\cap I_j} \frac{\log p}{p} - \sum_{p\in\mathcal{P}'\cap I_j'} \frac{\log p}{p} \right) \\ + O\left(\sqrt{S} + \epsilon S + \frac{1}{\log x} + \frac{S}{\log x_0}\right). \end{aligned}$$

But the inequalities (4.41) governing the selection of $\mathcal{P}, \mathcal{P}'$ imply

$$\begin{aligned} &\sum_{1\le j<j_0} \frac{M^*(x/x_j)}{x_j} \left(\sum_{p\in\mathcal{P}\cap I_j} \frac{\log p}{p} - \sum_{p\in\mathcal{P}'\cap I_j'} \frac{\log p}{p} \right) \\ &\qquad \ll \sum_{1\le j<j_0} \frac{1}{\log x_j} \left(\frac{\log\left(x_j(1+\epsilon)\right)}{x_j} + \frac{\log\left(x_j'(1+\epsilon)\right)}{x_j'} \right) \\ &\qquad \ll \sum_{1\le <j_0} \frac{1}{\log x_j} \left(\frac{\log x_j}{x_j} \right) \ll \frac{j_0}{x_0}. \end{aligned}$$

Hence

$$SM^*(x) - S'M^*(x') \ll \frac{j_0}{x_0} + \sqrt{S} + \epsilon S + \frac{1}{\log x} + \frac{S}{\log x_0}.$$

Finally, using (4.48), we see

$$S - S' \ll S'\sqrt{\eta} + \frac{S'}{\delta \log x_0} \ll S\sqrt{\eta} + \frac{S}{\delta \log x_0},$$

so that

$$SM^*(x) - SM^*(x') \ll \frac{j_0}{x_0} + \sqrt{S} + \epsilon S + \frac{1}{\log x} + \frac{S}{\log x_0} + \frac{S}{\delta \log x_0} + \sqrt{\eta}.$$

4.6.5 Denouement

Dividing by S we have shown

$$\begin{aligned} |M^*(x) - M^*(x')| &\ll \epsilon + \frac{j_0}{x_0 S} + \frac{1}{\sqrt{S}} + \frac{1}{S \log x} + \frac{1}{\log x_0} + \frac{1}{\delta \log x_0} + \sqrt{\eta} \\ &\ll \epsilon + \frac{j_0}{x_0 S} + \frac{1}{\sqrt{S}} + \frac{\sqrt{\eta}}{S \log x_0} + \frac{1}{\log x_0} + \frac{1}{\delta \log x_0} + \sqrt{\eta}. \end{aligned} \tag{4.51}$$

We now choose η sufficiently small and x sufficiently large that the right hand side of (4.51) is $O(\epsilon)$. But let us be more precise about this:

At various points in the argument we have been content to say that certain O-estimates hold if x_0 is sufficiently large and η sufficiently small, both in terms of ϵ. That is, there are constants $N_0 = N_0(\epsilon) > 3$, say, and $\eta_0 = \eta_0(\epsilon) > 0$, such that our estimates are valid (e.g., the implied constants can be “filled in”) once $x_0 > N_0$, say, and $\eta < \eta_0$.

We now fix an $\eta < \eta_0$ with the further property that

$$\max\{1/S, 1/\sqrt{S}, \sqrt{\eta}\} < \epsilon$$

for any choice of parameters with $x_0 > N_0(\epsilon)$. (That this is possible follows from (4.50).) We can now find an $N_1 = N_1(\eta, \epsilon) > N_0$ with the property that each term on the right hand side of (4.51) is bounded by ϵ provided $x_0 > N_1$. The only term for which this is not immediately clear is $j_0/x_0 S$, but this follows since

$$\frac{j_0}{x_0} \le \frac{1}{x_0} \frac{\log(x'/x)}{\log(1+\epsilon)} \le \frac{1}{x_0} \frac{\eta \log x}{\log(1+\epsilon)} = \frac{\log x_0}{x_0} \frac{\sqrt{\eta}}{\log(1+\epsilon)}.$$

Finally, pick N_2 so that $x > N_2$ implies $x^{\sqrt{\eta}} > N_1$. It follows that for any $x > N_2$, the right hand side of (4.51) is $O(\epsilon)$ with an absolute implied constant.

Exercise 4.6.1. Modify Hildebrand's proof to directly establish that $L(x) := \sum_{n \le x} \lambda(n) = o(x)$, where λ is the arithmetic function defined in Exercise 4.2.4.

References

[Dab84] Hédi Daboussi, *Sur le théorème des nombres premiers*, C. R. Acad. Sci. Paris Sér. I Math. **298** (1984), no. 8, 161–164. MR 85f:11065

[Elk03] Noam D. Elkies, *The Selberg (quadratic) sieve and some applications*, from Lecture Notes on Analytic Number Theory. Available online at `http://www.math.harvard.edu/~elkies/M259.02/index.html`, 2003.

[Erd49] P. Erdös, *On a new method in elementary number theory which leads to an elementary proof of the prime number theorem*, Proc. Nat. Acad. Sci. U. S. A. **35** (1949), 374–384. MR 10,595c

[Erd60] ______, *An asymptotic inequality in the theory of numbers*, Vestnik Leningrad. Univ. **15** (1960), no. 13, 41–49. MR 23 #A3720

[Gol03] D. Goldfeld, *The elementary proof of the prime number theorem: An historical perspective*, Available online from `http://www.math.columbia.edu/~goldfeld/`, 2003.

[Gre] Benjamin Green, *Notes on sieve theory*, DVI files available for download online at `http://www.dpmms.cam.ac.uk/~bjg23/expos.html`.

[Hil86] Adolf Hildebrand, *The prime number theorem via the large sieve*, Mathematika **33** (1986), no. 1, 23–30. MR 88a:11085

[Kub62] I. P. Kubilyus, *Probabilistic methods in the theory of numbers*, Amer. Math. Soc. Transl. (2) **19** (1962), 47–85. MR 24 #A1266

[Lan99] Edmund Landau, *Neuer Beweis der Gleichung* $\sum_{k=1}^{\infty} \frac{\mu(k)}{k} = 0$., Ph.D. thesis, Berlin. 16 S. gr. 8°, 1899.

[Lan06] ______, *Über den Zusammenhang einiger neuerer Sätze der analytischen Zahlentheorie*, Wien. Ber. **115** (1906), 589–632.

[Lan53] ______, *Handbuch der Lehre von der Verteilung der Primzahlen. 2 Bände*, Chelsea Publishing Co., New York, 1953, 2d ed, With an appendix by Paul T. Bateman. MR 16,904d

[Lan69] ______, *Vorlesungen über Zahlentheorie. Erster Band, zweiter Teil; zweiter Band; dritter Band*, Chelsea Publishing Co., New York, 1969. MR 40 #4076

[Mai85] Helmut Maier, *Primes in short intervals*, Michigan Math. J. **32** (1985), no. 2, 221–225. MR 86i:11049

[Sel47] Atle Selberg, *On an elementary method in the theory of primes*, Norske Vid. Selsk. Forh., Trondhjem **19** (1947), no. 18, 64–67. MR 9,271h

[Sel49] ———, *An elementary proof of the prime-number theorem*, Ann. of Math. (2) **50** (1949), 305–313. MR 10,595b

[Sha83] Harold N. Shapiro, *Introduction to the theory of numbers*, Pure and Applied Mathematics, John Wiley & Sons Inc., New York, 1983, A Wiley-Interscience Publication. MR 84f:10001

[Tur34] P. Turán, *On a theorem of Hardy and Ramanujan*, Journ. London Math. Soc. **9** (1934), 274–276.

Part II

Additive and Combinatorial Number Theory

Chapter 5

A Potpourri of Additive Number Theory

5.1 Introduction

Problems relating to the addition of sets of integers belong to the realm of "additive number theory." In this chapter we survey some representative problems in an effort to impart the flavor of this rich subject.

If $\mathcal{A}_1, \dots, \mathcal{A}_h$ are subsets of a given additive semigroup, we define the sumset

$$\mathcal{A}_1 + \cdots + \mathcal{A}_h := \{a_1 + \cdots + a_h : a_1 \in \mathcal{A}_1, \dots, a_h \in \mathcal{A}_h\}.$$

For a single subset $\mathcal{A}$, the *h-fold sumset* is defined by[1]

$$h\mathcal{A} := \underbrace{\mathcal{A} + \cdots + \mathcal{A}}_{h \text{ times}}.$$

A natural question to ask when studying sumsets is how the "size" of the sumset $\mathcal{A} + \mathcal{B}$ compares to that of the component summands $\mathcal{A}$ and $\mathcal{B}$. If $\mathcal{A}, \mathcal{B}$ are nonempty, finite subsets of $\mathbf{Z}$, then it is straightforward (see Exercise 5.1.1) to verify the inequality

$$|\mathcal{A} + \mathcal{B}| \geq |\mathcal{A}| + |\mathcal{B}| - 1, \tag{5.1}$$

and to give an example showing this is best possible. A far deeper theorem of Cauchy [Cau13] asserts that (5.1) continues to hold if $\mathcal{A}, \mathcal{B}$ are subsets of $\mathbf{F}_p$, subject to the obvious necessary condition that $|\mathcal{A}| + |\mathcal{B}| - 1 \leq p$. We shall give two proofs of this, the first employing recent ideas of Alon, Nathanson & Rusza [ANR95] and the second a classical argument of I. Chowla [Cho35]. Related results and applications are also discussed.

Another major strand in the web of additive number theory is the study of additive bases. If $\mathcal{A}$ is a subset of an additive semigroup S, then we call $\mathcal{A}$ a

[1] Do not confuse this with the dilation $h \times \mathcal{A} := \{ha : a \in \mathcal{A}\}$.

basis of finite order if $h\mathcal{A} = S$ for some positive integer h, i.e., if every element of S can be expressed as a sum of h (not necessarily distinct) elements of $\mathcal{A}$. The least h with this property is called the *order* of the basis. We say that $\mathcal{A}$ is an *asymptotic basis of finite order* if $S \setminus h\mathcal{A}$ is finite for some positive integer h, i.e., if all but finitely many elements of S are the sum of h elements of $\mathcal{A}$. The minimal h is called the *order of the asymptotic basis*.

As an illustration, the set of nonnegative squares forms a basis for the natural numbers of order at most 4, by a well-known theorem of Lagrange. Since every $n \equiv 7 \pmod 8$ actually requires four squares in its representation, this set is neither a basis nor an asymptotic basis of order less than 4.

One method of studying bases is to return to the question we asked before about the size of a sumset versus the size of the summands. For example, we shall deduce from the above theorem of Cauchy that the set of kth powers always forms a basis of $\mathbf{F}_p$ of order at most k. This is a $\pmod p$ analog of Waring's Problem.

When studying additive bases for the natural numbers, the role of "size" in this strategy must be replaced by some notion of "thickness." A convenient measure of thickness for such problems is the *Schnirelmann density*, defined for sets of natural numbers $\mathcal{A}$ by

$$\delta(\mathcal{A}) := \inf_{n=1,2,\dots} \frac{A(n)}{n}.$$

Note that for any set $\mathcal{A} \subset \mathbf{N}$, one has $0 \le \delta(\mathcal{A}) \le 1$, and $\delta(\mathcal{A}) = 1$ if and only if $\{1, 2, 3, \dots\} \subset \mathcal{A}$. A deep theorem of Mann [Man42] asserts that for sets $\mathcal{A}, \mathcal{B}$ of natural numbers, both of which contain 0, one has

$$\delta(\mathcal{A} + \mathcal{B}) \ge \min\{\delta(\mathcal{A}) + \delta(\mathcal{B}), 1\}.$$

It follows that if $0 \in \mathcal{A}$ and $\delta(\mathcal{A}) > 0$, then $\mathcal{A}$ is a basis of finite order. This criterion was first observed by Schnirelmann around 1930, who gave a proof independent of Mann's theorem. Both Schnirelmann's ideas and the proof of Mann's theorem are discussed in this chapter.

We also follow Schnirelmann [Sch33] in using his criterion to prove the following beautiful result: There exists an absolute constant C with the property that every integer $n > 1$ is a sum of at most C primes. The analytic input needed for the proof is the upper bound of Chapter 3 for $r(N)$, the number of ordered representations of N as a sum of two primes, which is used to establish that the set $\{p + q : p, q \text{ prime}\}$ has positive lower density.

Similar ideas allow us to prove Romanov's theorem [Rom34] that for each fixed $a \ge 2$, a positive proportion of the positive integers can be written in the form $p + a^k$. When $a = 2$, Romanov asked if perhaps all large enough odd integers admit such a representation. This was disproved in 1950 by Erdős [Erd50] and van der Corput [vdC50], who independently showed that a positive proportion of the odd integers could not be so represented. We give Erdős' proof here. A simpler proof that there are infinitely many odd integers that do not admit such a representation was given by Crocker [Cro61], who showed this holds for $2^{2^n}-5$ for each $n \geq 3$. He later showed [Cro71] that $2^{2^n}-1$ (for $n \geq 3$) cannot be written as the sum of a prime and two distinct positive powers of 2. By combining the "covering congruence" method of Erdős with his construction, he was able to produce infinitely many positive integers that cannot be written as the sum of a prime and two positive powers of 2, distinct or otherwise. Regrettably, this final construction is too complicated to be reproduced here.

Finally, we turn to the intersection of combinatorial and additive number theory and discuss a theorem proved by Schur [Sch16] in his investigations on Fermat's last theorem: Given any k-coloring of the positive integers, there is always a monochromatic solution to $x - y = z$; moreover, this holds if one restricts the coloring to $\{1, 2, \ldots, \lfloor k!e \rfloor\}$. We present a simple graph-theoretic proof of Schur's theorem as presented by Mirsky [Mir75], discuss the sharpness of the bound $\lfloor k!e \rfloor$, and give Schur's original application of this result to the nontrivial solvability of the Fermat congruence $x^m + y^m \equiv z^m \pmod{p}$ for large p (i.e., $p > p_0(m)$).

5.1.1 Exercises

Exercise 5.1.1. Let $\mathcal{A}, \mathcal{B}$ be nonempty subsets of $\mathbf{Z}$. Show that $|\mathcal{A} + \mathcal{B}| \geq |\mathcal{A}| + |\mathcal{B}| - 1$ and give an example of sets $\mathcal{A}, \mathcal{B}$ for which equality holds. Can you determine all cases where equality holds?

Exercise 5.1.2. Show that if $\mathcal{A}$ is a basis of order h_0, then $h\mathcal{A} = \mathbf{N}$ for every $h \geq h_0$. Similarly, if $\mathcal{A}$ is an asymptotic basis of order h_0, then $\mathbf{N} \setminus h\mathcal{A}$ is finite for every $h \geq h_0$.

Exercise 5.1.3. Fix an integer $r \geq 2$. For a set $\mathcal{A}$ of natural numbers, let $\hat{\mathcal{A}}$ denote the set of integers of the form $\sum_{i \in \mathcal{A}} d_i r^i$, where $0 \leq d_i \leq r-1$ and only finitely many d_i are nonzero.

Show that if $\mathbf{N} := \dot\cup_{i=1}^{k} \mathcal{A}_i$ is a partition of the natural numbers into k sets, then $\mathcal{B} := \cup_{i=1}^{k} \hat{\mathcal{A}}_i$ is a basis of order k.

Exercise 5.1.4.

a) Show that if $\mathcal{B} \subset \mathbf{N}$ is a basis of order k, then $B(x) \gg_k x^{1/k}$ as $x \to \infty$, uniformly in the particular choice of $\mathcal{B}$.

b) Using the previous exercise, show that for each fixed positive integer k, there exists a basis $\mathcal{B}$ of order k whose counting function satisfies $B(x) \ll x^{1/k}$ as $x \to \infty$.

Exercise 5.1.5. Show that every sufficiently large positive integer is a sum of two composite positive integers. More generally, show that for each k, every large positive integer is a sum of two positive integers with at least k distinct prime factors.

Exercise 5.1.6 *(Euler).* Exhibit an odd integer greater than 3 which cannot be represented as the sum of prime and a positive power of 2.

Exercise 5.1.7 *(Nathanson [Nat80]).* If $\mathcal{A}$ is a set of positive integers, let $\mathrm{FS}(\mathcal{A})$ denote the collection of all nonempty finite sums of elements of $\mathcal{A}$.

a) Show that if $\mathcal{A}$ contains arbitrarily long runs of consecutive integers, then there is an infinite set $\mathcal{B} \subset \mathcal{A}$ with $\mathrm{FS}(\mathcal{B}) \subset \mathcal{A}$.

b) Show that the hypothesis of part a) is satisfied if $d(\mathcal{A}) = 1$.

5.2 Sumsets over $\mathbf{Z}/m\mathbf{Z}$ and $\mathbf{F}_q$

We begin with a presentation of the "polynomial method" of Alon, Nathanson & Rusza [ANR95] for attacking additive problems, beginning with their proof of Cauchy's lower-bound theorem on the size of sumsets over $\mathbf{F}_p$.

Cauchy's theorem implies that if $\mathcal{A}$ is a nonempty subset of $\mathbf{F}_p$, then

$$|\{a_1 + a_2 : a_1, a_2 \in \mathcal{A}\}| \geq \min\{p, 2|\mathcal{A}| - 1\}.$$

In the 1960s, Erdős and Heilbronn conjectured that if one imposes the extra condition that $a_1 \neq a_2$, then the corresponding inequality holds with the lower bound replaced by $\min\{p, 2|\mathcal{A}| - 3\}$. The family of examples $\mathcal{A} = \{0, 1, \ldots, k-1\}$, where $p \geq 2k - 3$, illustrates that this is best possible if true. As we shall see, within the framework set up by Alon, Nathanson & Rusza, the proof of this conjecture is no more difficult than the proof of Cauchy's classical theorem.

Finally, we present Chowla's generalization of Cauchy's theorem to composite moduli and apply these ideas to Waring's Problem for residues, proving that the set of kth powers forms an additive basis of order at most k modulo any prime p.

5.2.1 The Polynomial Method of Alon, Nathanson, Rusza

Lemma 5.2.1 (Alon & Tarsi [AT92]). *Let $\mathcal{A}, \mathcal{B}$ be nonempty subsets of a field F with $|\mathcal{A}| = k, |\mathcal{B}| = l$. Let $f(X, Y) \in F[X, Y]$ be a polynomial of degree at most $k - 1$ in X and $l - 1$ in Y. If $f(a, b) = 0$ for all $a \in \mathcal{A}, b \in \mathcal{B}$, then $f(X, Y)$ is identically 0.*

Proof. We make repeated use of the well-known fact that a nonzero univariate polynomial over F cannot have more roots than its degree. Write

$$f(X, Y) = \sum f_{ij} X^i Y^j = \sum_{i=0}^{k-1} v_i(Y) X^i,$$

where $v_i(Y) = \sum_{j=0}^{l-1} f_{ij} Y^j$. For any fixed $b \in \mathcal{B}$, $f(X, b)$ is a polynomial in X of degree at most $k - 1$ with at least k distinct roots. Consequently, $v_i(b) = 0$ for each $0 \leq i \leq k - 1$.

This implies that for $1 \leq i \leq k - 1$, $v_i(Y)$ is a polynomial of degree at most $l - 1$ with at least l distinct roots. Hence $v_i(Y)$ is identically 0 for each i. Referring to the definition of $v_i(Y)$ shows that f_{ij} vanishes for each i, j, so that $f(X, Y)$ is identically 0 as was to be shown. $\square$

Lemma 5.2.2. *Let $\mathcal{A}$ be a nonempty finite subset of a field F, and let $|\mathcal{A}| = k$. For every $m \geq 0$, there is a polynomial $r_m(X) \in F[X]$ of degree at most $k-1$ such that $r_m(a) = a^m$ for all $a \in \mathcal{A}$.*

Proof. Let $t(X) = \prod_{a\in\mathcal{A}}(X-a)$, and let $r_m(X)$ be the unique polynomial of degree at most $k-1$ with $X^m \equiv r_m(X) \pmod{t(X)}$. $\square$

Theorem 5.2.3 (Cauchy [Cau13], Davenport [Dav35]). *Let $F = \mathbf{Z}/p\mathbf{Z}$ with p prime. Let $\mathcal{A}, \mathcal{B}$ be nonempty subsets of F, and let $\mathcal{C} = \mathcal{A} + \mathcal{B}$. Then*

$$|\mathcal{C}| \geq \min\{p, |\mathcal{A}| + |\mathcal{B}| - 1\}. \tag{5.2}$$

Proof. Without loss of generality we may assume $k + l - 1 \leq p$, for otherwise $l' := p - k + 1$ satisfies $1 \leq l' < l$, and we may replace $\mathcal{B}$ with a subset $\mathcal{B}' \subset \mathcal{B}$ of size l'.

In this case, (5.2) asserts that $|\mathcal{C}| \geq k + l - 1$. Suppose otherwise, and let w be the nonnegative integer defined by $w + |\mathcal{C}| = k + l - 2$. Consider the polynomial

$$f(X,Y) := (X+Y)^w \prod_{c\in\mathcal{C}}(X+Y-c) \in F[X,Y].$$

Then $f(X,Y)$ has total degree $k+l-2$, the values $f(a,b)$ vanish for each $a \in \mathcal{A}, b \in \mathcal{B}$, and the coefficient of $X^{k-1}Y^{l-1}$ is

$$\binom{k+l-2}{k-1} \not\equiv 0 \pmod{p}, \tag{5.3}$$

since $k + l - 2 \leq p - 1$.

Our strategy will be to transform f into a polynomial f^* of degree at most $k-1$ in X and $l-1$ in Y for which

$$f^*(a,b) = f(a,b) = 0 \quad (a \in \mathcal{A}, b \in \mathcal{B}), \tag{5.4}$$

and for which the coefficient of $X^{k-1}Y^{l-1}$ is unchanged. By Lemma 5.2.1, we must have $f^*(X,Y)$ identically zero, but this contradicts (5.3).

To effect this transformation, choose (by Lemma 5.2.2) for every $m, n \geq 1$, univariate polynomials $r_m(X), s_n(Y)$ of degree at most $k-1$ and $l-1$ respectively, with $r_m(a) = a^m$ for each $a \in \mathcal{A}$ and $s_n(b) = b^n$ for each $b \in \mathcal{B}$. Replace every monomial X^mY^n for which $m > k-1$ with $r_m(X)Y^n$. Since $m > k-1$ implies $n < l-1$, and since $r_m(X)$ is a polynomial of degree at most $k-1$, this has the effect of replacing X^mY^n with a (weighted) sum of monomials X^iY^n with $0 \leq i \leq k-1$. Similarly, if we now replace every monomial X^mY^n with $n > l-1$ with $X^m s_n(Y)$, then X^mY^n has been replaced with a sum of monomials X^mY^i with $0 \leq i \leq l-1$. The polynomial $f^*(X,Y)$ resulting from these two transformations is of degree at most $k-1$ in X and $l-1$ in Y, and (5.4) holds by the choice of r_m, s_n. Moreover, the coefficient of $X^{k-1}Y^{l-1}$ is unchanged, since this monomial does not appear in any of the polynomials $r_m(X)Y^n$ or $X^m s_n(Y)$. The result follows. $\square$

Vosper [Vos56] has shown that equality holds in Theorem 5.2.3 if and only if one of the following is true:

i. $|\mathcal{A}| + |\mathcal{B}| > p$,

ii. $\min\{|\mathcal{A}|, |\mathcal{B}|\} = 1$,

iii. there exists $c \in \mathbf{Z}/p\mathbf{Z}$ such that $\mathcal{A} = \{c - b' : b' \in \mathcal{B}^c\}$, where $\mathcal{B}^c$ denotes the complement of $\mathcal{B}$,

iv. $\mathcal{A}, \mathcal{B}$ are arithmetic progressions with the same common difference.

Exercise 5.2.1. Show that the four conditions above are sufficient for equality to hold in Theorem 5.2.3.

As a simple application of Cauchy's theorem, the reader is asked to tackle:

Exercise 5.2.2. Let p be prime and let $\mathcal{N}$ be the set of quadratic nonresidues $\pmod p$. Show that $|\mathcal{N} - \mathcal{N}| \geq p - 2$ and that if $c \notin \mathcal{N} - \mathcal{N}$, then $ck^2 \notin \mathcal{N} - \mathcal{N}$ for every nonzero $k \pmod p$. Conclude that $\mathcal{N} - \mathcal{N}$ is all of $\mathbf{Z}/p\mathbf{Z}$ for $p \geq 7$.

The same method used to prove Theorem 5.2.3 also yields:

Theorem 5.2.4 (Alon, Nathanson, Ruzsa [ANR95]). *Let $F = \mathbf{Z}/p\mathbf{Z}$ with p prime. Let $\mathcal{A}, \mathcal{B}$ be nonempty subsets of F with $|\mathcal{A}| \neq |\mathcal{B}|$, and let*

$$\mathcal{C} := \{a + b : a \in \mathcal{A}, b \in \mathcal{B}, a \neq b\}.$$

Then

$$|\mathcal{C}| \geq \min\{p, |\mathcal{A}| + |\mathcal{B}| - 2\}.$$

Proof. Let $\mathcal{A} = k$, $\mathcal{B} = l$. Without loss of generality we may assume $k+l-2 \leq p$, for otherwise $l' := p - k + 2$ satisfies $2 \leq l' < l$, and we may replace $\mathcal{B}$ with a subset $\mathcal{B}' \subset \mathcal{B}$ of size l'.

In this case, (5.2) asserts that $|\mathcal{C}| \geq k + l - 2$. Suppose otherwise, and let w be the nonnegative integer defined by $w + |\mathcal{C}| = k + l - 3$. Consider the polynomial

$$f(X, Y) := (X - Y)(X + Y)^w \prod_{c \in \mathcal{C}} (X + Y - c) \in F[X, Y].$$

Then $f(X, Y)$ has total degree $k + l - 2$, the values $f(a, b)$ vanish for all $a \in \mathcal{A}, b \in \mathcal{B}$, and the coefficient of $X^{k-1}Y^{l-1}$ is

$$\binom{k+l-3}{k-2} - \binom{k+l-3}{l-2} = \frac{(k-l)(k+l-3)!}{(k-1)!(l-1)!} \not\equiv 0 \pmod p, \qquad (5.5)$$

since $1 \leq k, l \leq p$ and $k + l - 3 \leq p - 1$. The remainder of the proof is as in Theorem 5.2.3. $\square$

As a corollary, we deduce the conjecture of Erdős & Heilbronn, which was first proven by da Silva & Hamidoune [DdSH94] by a much more complicated method:

Corollary 5.2.5 (Dias da Silva & Hamidoune). *Let $F = \mathbf{Z}/p\mathbf{Z}$, where p is a prime number. Let $\mathcal{A} \subset F$ with $|\mathcal{A}| = k \geq 2$. Then with $2\hat{\ }\mathcal{A}$ denoting the set of all sums of two distinct elements of $\mathcal{A}$, we have*

$$|2\hat{\ }\mathcal{A}| \geq \min\{p, 2k-3\}.$$

Proof. Arbitrarily choose $a \in \mathcal{A}$ and apply Theorem 5.2.4 with $\mathcal{A}$ as given and $\mathcal{B} := \mathcal{A} \setminus \{a\}$. □

Exercise 5.2.3 *(Alon, Nathanson, Ruzsa [ANR95]).* Let $\mathcal{A}, \mathcal{B}$ be nonempty subsets of $F = \mathbf{Z}/p\mathbf{Z}$ with $|\mathcal{A}| = k$ and $|\mathcal{B}| = l$. Let $\mathcal{C} = \{a + b : a \in \mathcal{A}, b \in \mathcal{B}, ab \neq 1\}$. Show that $|\mathcal{C}| \geq \min\{p, k + l - 3\}$. Hint: Proceed as in the proofs of Theorems 5.2.3 and 5.2.4, taking (with $w + |\mathcal{C}| = k + l - 4$)

$$f(X,Y) = (XY-1)(X+Y)^w \prod_{c \in \mathcal{C}} (X+Y-c).$$

Exercise 5.2.4. Show that the results of Theorem 5.2.3, Theorem 5.2.4 and Corollary 5.2.5 remain true for any field F, with $p = \mathrm{char}(F)$ if $\mathrm{char}(F) < \infty$ and $p = \infty$ otherwise.

5.2.2 Chowla's Sumset Addition Theorem

Theorem 5.2.6 (Chowla [Cho35]). *Let $\mathcal{A}, \mathcal{B}$ be nonempty subsets of $\mathbf{Z}/m\mathbf{Z}$ with $|\mathcal{A}| = k$, $|\mathcal{B}| = l$. Then if $0 \in \mathcal{B}$ and every nonzero $b \in \mathcal{B}$ is an element of $\mathbf{Z}/m\mathbf{Z}^*$, then $C := \mathcal{A} + \mathcal{B}$ satisfies*

$$|C| \geq \min\{m, l + k - 1\}. \tag{5.6}$$

Proof. The proof is by induction on l, the case $l = 1$ being trivial. Suppose now the result is known for all $l' < l$ with $l > 1$, and let $\mathcal{A}, \mathcal{B}$ be as in the theorem statement. As in the proof of Theorem 5.2.3, we may assume that $l + k - 1 \leq m$, and we may also assume that $k < m$.

We first claim that $\mathcal{A} + \mathcal{B} \not\subset \mathcal{A}$. Otherwise, for each fixed $b \in \mathcal{B}$ the map $a + b \mapsto a$ would be a permutation of $\mathcal{A}$. In particular, we would have

$$\sum_{a \in \mathcal{A}} (a + b) = \sum_{a \in \mathcal{A}} a, \quad \text{so that} \quad kb \equiv 0 \pmod{m}. \tag{5.7}$$

But since $l > 1$, we can choose $0 \neq b \in \mathcal{B}$; then $b \in \mathbf{Z}/m\mathbf{Z}^*$, so that (5.7) implies $k \equiv 0 \pmod{m}$, contradicting that $k < m$.

We may thus choose $a \in \mathcal{A}$ for which $\mathcal{B}'' := \{b \in \mathcal{B} : a + b \notin \mathcal{A}\} \neq \emptyset$. Set $\mathcal{A}' := \mathcal{A} \cup (a + \mathcal{B}''), \mathcal{B}' := \mathcal{B} \setminus \mathcal{B}''$. Then $1 \leq |\mathcal{B}'| < |\mathcal{B}| = l$, $|\mathcal{A}'| + |\mathcal{B}'| = l + k$, and

$$\mathcal{A}' + \mathcal{B}' = (\mathcal{A} \cup (a + \mathcal{B}'')) + \mathcal{B} \setminus \mathcal{B}'' \subset (\mathcal{A} + \mathcal{B}) \cup ((a + \mathcal{B} \setminus \mathcal{B}'') + \mathcal{B}'') \subset \mathcal{A} + \mathcal{B},$$

since $a + \mathcal{B} \setminus \mathcal{B}'' \subset \mathcal{A}$ by the choice of $\mathcal{B}''$. □

5.2.3 Waring's Problem for Residues

An 18th century conjecture of Waring, which we shall prove in Chapter 7, asserts that the set of kth powers is a basis of finite order for the natural numbers (for each $k \geq 1$). The analogous conjecture for $\mathbf{Z}/m\mathbf{Z}$ is trivially true but the question of obtaining a nontrivial estimate for the order $g(k, m)$ of the basis is still of interest.

As a simple application of Theorem 5.2.3, we now prove:

Theorem 5.2.7. *Let k be a positive integer. For every prime p, we have $g(k,p) \leq [\mathbf{F}_p^* : \mathbf{F}_p^{*k}]$. In particular, $g(k,p) \leq k$.*

Proof. Let l denote the index in question and P denote the set of kth powers. Then Theorem 5.2.3 implies

$$|\underbrace{P + \cdots + P}_{l \text{ times}}| \geq \min\{l|P| - l + 1, p\}.$$

Since

$$l|P| - l + 1 = l\left((p-1)/l + 1\right) - l + 1 = p,$$

the result follows. □

Actually $g(k,p) \leq 2$ for all sufficiently large p; in fact $p > (k-1)^4$ suffices. This may be deduced from [IR90, Chapter 8, Theorem 5]. The same theorem implies that if $a_1, \ldots, a_r, b$ are (fixed) integers with the a_i nonvanishing, then the congruence

$$a_1 x_1^k + a_2 x_2^k + \cdots + a_r x_r^k = b$$

is solvable for all sufficiently large p, and in fact the number of solutions tends to infinity with p. For an *elementary* proof that $g(3,p) \leq 2$ for $p > 7$, see [LS89].

The accessible papers of C. Small ([Sma77b], [Sma77a]) contain a "solution" to Waring's problem for residues. His papers include explicit evaluations of $g(k,n)$, for each of $k = 2, 3, 4$, and 5, and his method extends to arbitrary k. His general results are rather complicated; as a sample, we quote the case $k = 3$:

Theorem (C. Small). *Let $r \geq 1$ and $n > 1$ be integers. Then every element of $\mathbf{Z}/n\mathbf{Z}$ is a sum of r cubes if and only if one of the following holds:*

i. $r = 1$ *and:* $9 \nmid n, p^2 \nmid n$ *for all primes* $p \equiv 2 \pmod 3$, $p \nmid n$ *for all primes* $p \equiv 1 \pmod 3$,

ii. $r = 2$ *and* $9 \nmid n, 7 \nmid n$,

iii. $r = 3$ *and* $9 \nmid n$,

iv. $r \geq 4$ *(and no condition on* n*).*

5.2.4 Exercises: More on Waring's Problem for Residues

Exercise 5.2.5. Show that if $k = p - 1$ or $k = (p-1)/2$, then $g(k, p) = k$.

Exercise 5.2.6.

a) Show that $g(k, m) = \max_{p^e \| m} g(k, p^e)$.

b) Show that if m is squarefree, the kth powers $\pmod m$ are a basis of order at most k.

Exercise 5.2.7. Prove Theorem 5.2.7 without citing the the Cauchy-Davenport-Chowla Theorem. Proceed as follows: Define

$$G_i := \{x \in \mathbf{F}_p : x \text{ is a sum of } i \text{ } k\text{th powers}\}.$$

Then

$$\mathbf{F}_p^k = G_1 \subset G_2 \subset \cdots, \tag{5.8}$$

and $G_i = \mathbf{F}_p$ for all large i.

a) Show that if $x \in G_{i+1} \setminus G_i$, then $xy \in G_{i+1} \setminus G_i$ for each nonzero kth power y.

b) Show that if $G_i \subsetneq G_{i+1}$, then $|G_{i+1}| \geq |G_i| + (p-1)/l$.

c) Conclude that the chain (5.8) stabilizes after at most $l - 1$ strict containments, and use this to deduce Theorem 5.2.7.

Exercise 5.2.8. Let F be a finite field and let k be a positive integer. Let $E \subset F$ be the set of elements expressible as a sum of finitely many kth powers.

a) Show that E is a subfield of F.

b) Show that E is a proper subfield of F if and only if k is divisible by $[F^* : K^*]$ for some proper subfield $K \subsetneq F$.

c) Show that with $l = [E^* : F^{*k}]$, every element of E is a sum of l kth powers.

Exercise 5.2.9. Let q be a prime power and let k be a positive integer. Suppose $\mathbf{Z}/q\mathbf{Z}^*$ is cyclic. Using Theorem 5.2.6, show that for $s \geq p(k, \phi(q))/(p-1)$, the congruence

$$x_1^k + x_2^k + \cdots + x_s^k \equiv n \pmod q$$

is solvable in integers x_i with $(x_1, q) = 1$, for every integer n.

If $\mathbf{Z}/q\mathbf{Z}^*$ is not cyclic, so that $q = 2^t$ with $t \geq 3$, show the same holds for $s \geq 2$ if k is odd, and for $s \geq 4(q/4, k)$ if k is even.

5.3 The Density of Sumsets

5.3.1 Schnirelmann Density and Additive Bases

We now investigate sufficient conditions for a set $\mathcal{A}$ to be a basis (or asymptotic basis) of finite order. Our results are most conveniently expressed in terms of the so-called Schnirelmann density, defined for sets $\mathcal{A} \subset \mathbf{N}$ by

$$\delta(\mathcal{A}) := \inf_{n=1,2,\ldots} \frac{A(n)}{n}. \tag{5.9}$$

Exercise 5.3.1. Determine the Schnirelmann density of each of the following sets: a) the set of squares, b) the set of primes, c) the set of primes together with 1, and d) the set of natural numbers $n \equiv a \pmod{q}$ (where a, q are any integers with $q > 0$).

Schnirelmann density is rather less natural than the upper and lower densities we have discussed before. Both of the latter are translation-invariant and are unaffected by the removal or addition of finitely many elements. The Schnirelmann density possesses neither of these properties: the set of natural numbers has Schnirelmann density 1, but $\mathbf{N} \setminus \{1\}$ has density 0. Indeed, any set $\mathcal{A} \subset \mathbf{N}$ with $1 \notin \mathcal{A}$ has Schnirelmann density 0.

Exercise 5.3.2. Establish the following equivalence, which will be used later in this section:

$$\delta(\mathcal{A}) > 0 \Longleftrightarrow 1 \in \mathcal{A} \text{ and } \underline{d}(\mathcal{A}) > 0.$$

Despite its flaws, Schnirelmann density turns out to be the "proper" measure of size in many additive problems.

Lemma 5.3.1. *Let $\mathcal{A}, \mathcal{B} \subset \mathbf{N}$ with $0 \in A, 0 \in B$. If $A(n) + B(n) \geq n$ for the integer $n \geq 0$, then $n \in \mathcal{A} + \mathcal{B}$.*

Proof. If $n \in A$ or $n \in B$, this follows from $0 \in \mathcal{A} \cap \mathcal{B}$. Otherwise $A(n) = A(n-1), B(n) = B(n-1)$, and we may write

$$\mathcal{A}' := \mathcal{A} \cap [1, n-1] = \{a_1 < \cdots < a_{A(n)}\},$$
$$\mathcal{B}' := \mathcal{B} \cap [1, n-1] = \{b_1 < \cdots < b_{B(n)}\}.$$

Consider the following list of integers from the interval $[1, n-1]$:

$$a_1, \ldots, a_{A(n)}, n - b_1, \ldots, n - b_{B(n)}.$$

There are $A(n) + B(n) > n - 1$ terms on this list but only $n - 1$ integers in the range $[1, n-1]$; this implies that for some i, j, we have $a_i = n - b_j$. But then $n = a_i + b_j \in \mathcal{A} + \mathcal{B}$. □

As a consequence, we see:

Corollary 5.3.2. *Suppose $\mathcal{A} \subset N$, $0 \in \mathcal{A}$ and $\delta(A) \geq 1/2$. Then $\mathcal{A} + \mathcal{A} = \mathbf{N}$.*

Indeed, the hypothesis implies $A(n) + A(n) \geq n/2 + n/2 \geq n$ for every nonnegative integer n.

From this we easily conclude that for a subset $\mathcal{A} \subset \mathbf{N}$ with $0 \in \mathcal{A}$ to be a basis of finite order, it is sufficient that $\delta(h\mathcal{A}) \geq 1/2$ for some positive integer h. Indeed, in this case

$$(2h)\mathcal{A} = h\mathcal{A} + h\mathcal{A} = \mathbf{N} \tag{5.10}$$

by Corollary 5.3.2, so that $\mathcal{A}$ is a basis of order not exceeding $2h$. This suggests that we study the density of sets of the form $h\mathcal{A}$, or more generally the density of sumsets $\mathcal{A}_1 + \cdots + \mathcal{A}_k$.

Lemma 5.3.3. *If $\mathcal{A}, \mathcal{B} \subset \mathbf{N}$ with $0 \in \mathcal{A}, 0 \in \mathcal{B}$, then*

$$\delta(\mathcal{A} + \mathcal{B}) \geq \delta(\mathcal{A}) + \delta(\mathcal{B}) - \delta(\mathcal{A})\delta(\mathcal{B}). \tag{5.11}$$

Proof. Let n be a positive integer, and let $1 = a_1 < a_2 < \cdots < a_{A(n)} \leq n$ be a list of the elements of $\mathcal{A} \cap [1, n]$. Between a_i and a_{i+1}, there is a gap containing $g_i := a_{i+1} - a_i - 1$ numbers. But this gap contains at least $B(g_i)$ elements of $\mathcal{A} + \mathcal{B}$, since if $b \in B \cap [1, g_i]$ then $a_i + b \in \mathcal{A} + \mathcal{B}$. Because $0 \in B$, we also have $\mathcal{A} \subset \mathcal{A} + \mathcal{B}$, whence

$$\begin{aligned}(A+B)(n) &\geq A(n) + \sum\nolimits_{i=1}^{k} B(g_i) \geq A(n) + \delta(\mathcal{B}) \sum\nolimits_{i=1}^{k} g_i \\ &= A(n) + \delta(\mathcal{B})(n - A(n)) = A(n)(1 - \delta(\mathcal{B})) + n\delta(\mathcal{B}) \\ &\geq n(\delta(\mathcal{A}) - \delta(\mathcal{A})\delta(\mathcal{B}) + \delta(\mathcal{B})).\end{aligned}$$

Thus, $(A + B)(n)/n \geq \delta(\mathcal{A}) + \delta(\mathcal{B}) - \delta(\mathcal{A})\delta(\mathcal{B})$. As n was arbitrary, (5.11) follows. □

Lemma 5.3.3 asserts that

$$1 - \delta(\mathcal{A} + \mathcal{B}) \leq (1 - \delta(\mathcal{A}))(1 - \delta(\mathcal{B})).$$

This is easily extended by induction to the following corollary, whose proof is left to the reader:

Corollary 5.3.4. *If $\mathcal{A}_1, \ldots, \mathcal{A}_k \subset \mathbf{N}$ $(k \geq 2)$ and $0 \in \mathcal{A}_1 \cap \cdots \cap \mathcal{A}_k$, then*

$$1 - \delta(\mathcal{A}_1 + \cdots + \mathcal{A}_k) \leq \prod_{i=1}^{k} (1 - \delta(A_i)).$$

In particular, if $\mathcal{A} \subset \mathbf{N}$ and $0 \in A$, then

$$\delta(k\mathcal{A}) \geq 1 - (1 - \delta(\mathcal{A}))^k.$$

Theorem 5.3.5 (Schnirelmann's Basis Theorem [Sch33]). *Let $\mathcal{A} \subset \mathbf{N}$, $0 \in \mathcal{A}$, and suppose $\delta(A) > 0$. Then $\mathcal{A}$ is a basis of finite order.*

Proof. We have $\delta(k\mathcal{A}) \geq 1/2$ as soon as

$$1 - (1-\alpha)^k \geq 1/2,$$

which occurs as soon as $k \geq (-\log 2)/\log(1-\alpha)$. Referring to (5.10) finishes the proof. In fact, we see that $2h\mathcal{A} = \mathbf{N}$ for

$$h := \max\{\lceil -\log 2/\log(1-\alpha)\rceil, 1\}. \tag{5.12}$$

□

Exercise 5.3.3 *(Landau [Lan30]).* Suppose that $0 \in \mathcal{A}$ and $\alpha := \delta(A) > 0$. Show that $\mathcal{A}$ is a basis of order not exceeding $2\lfloor 1/\alpha \rfloor$. Hint: If $\alpha > 1/2$, this follows from Corollary 5.3.2. If $\alpha \in (1 - \log 2, 1/2]$, use (5.12) and the estimate $-\log(1-\alpha) \geq -\log(\log 2)$. Finally, for $\alpha \leq 1 - \log 2$, use (5.12) and the inequality $-\log(1-\alpha) \geq \alpha$.

Exercise 5.3.4. Let $\mathcal{A}$ be a set of integers with $\delta(\mathcal{A}) > 1/2$.

a) Prove that every integer > 1 is a sum of two *positive* elements of $\mathcal{A}$.

b) Show that this hypothesis is satisfied if $\mathcal{A}$ is the sequence of squarefree integers. Hint: First show $\sum_p 1/p^2 < 1/2$. (Compare with Chapter 3, Exercise 3.3.5.)

Remark. The Schnirelmann density of the squarefree numbers is known to be 53/88 [Rog64], the infimum appearing in the definition (5.9) being attained for $n = 176$.

5.3.2 Mann's Density Theorem

Upon his 1931 return from a foreign tour, Schnirelmann reported that during a visit with Landau in Göttingen, they had jointly stumbled across a new law governing the densities of sumsets. Under the same assumptions as Lemma 5.3.3, all the examples they tried suggested that

$$\delta(\mathcal{A} + \mathcal{B}) \geq \min\{1, \delta(\mathcal{A}) + \delta(\mathcal{B})\}.$$

Proving this held for all such pairs $\mathcal{A}, \mathcal{B}$ (the "$\alpha+\beta$ conjecture") became a target of vigorous research during the next decade. In 1941, Alfred Brauer offered a course on additive number theory with the sole aim of presenting all the results on the problem to date. At the end of the semester, one of the students in the course, H.B. Mann, discovered a complete proof (see [Man42]).

In this section we present a proof of Mann's theorem (based on [Nar83, Chapter 4]) as well as that of a theorem of Lepson to the effect that Mann's result is best possible. These results are *not* needed for the rest of the chapter, so that this material could be omitted on a first reading.

Mann's theorem will be an immediate consequence of the following result on finite sets. The reader may wish to compare its proof with that of Chowla's theorem (Theorem 5.2.6).

Theorem 5.3.6. *Let n be a positive integer, and suppose $\mathcal{A}, \mathcal{B}$ are subsets of $\{0, 1, \ldots, n\}$ both containing 0. Suppose that for some $0 < c \le 1$, we have*

$$A(m) + B(m) \ge cm \tag{5.13}$$

for $m = 1, 2, \ldots, n$. Then

$$(A + B)(m) \ge cm \tag{5.14}$$

for $m = 1, 2, \ldots, n$.

Proof. The proof is by complete induction on n. When $n = 1$, (5.13) implies (taking $m = 1$) that $1 \in \mathcal{A} \cup \mathcal{B}$, so that (since $0 \in \mathcal{A} \cap \mathcal{B}$) also $1 \in \mathcal{A} + \mathcal{B}$, whence

$$(A + B)(1) = 1 \ge c \cdot 1 = c.$$

Now suppose the theorem is known for all positive integral $n' < n$, where $n \ge 2$. Supposing it false for n, choose a counterexample $\mathcal{A}, \mathcal{B}$ with $B(n)$ as small as possible. Note that $B(n) > 0$, otherwise $\mathcal{B} = \{0\}$ and the inequalities (5.13) and (5.14) coincide for each m. We proceed to construct sets $\mathcal{A}', \mathcal{B}' \subset \{0, 1, 2, \ldots, n\}$, both containing 0, with the following three properties:

i. for $m = 1, 2, \ldots, n$, we have

$$A'(m) + B'(m) \ge cm, \tag{5.15}$$

ii. $\mathcal{A}' + \mathcal{B}' \subset \mathcal{A} + \mathcal{B}$,

iii. $B'(n) < B(n)$.

Together these imply $\mathcal{A}', \mathcal{B}'$ are a counterexample (for this particular n) with $B'(n) < B(n)$, contradicting the original choice of $\mathcal{A}, \mathcal{B}$.

To construct the sets $\mathcal{A}', \mathcal{B}'$, note first that since $\mathcal{B}$ contains positive elements, $\mathcal{A} + \mathcal{B} \not\subset \mathcal{A}$ (e.g., the largest element of $\mathcal{A} + \mathcal{B}$ cannot lie in $\mathcal{A}$). We may thus choose $a_0 \in \mathcal{A}$ as small as possible so that

$$\mathcal{B}'' := \{b \in \mathcal{B} : a_0 + b \notin \mathcal{A}\} \neq \emptyset.$$

Note that by the minimality of a_0, we have for every positive $r < a_0$ and every $b \in \mathcal{B}$,

$$b + (\mathcal{A} \cap [0, r]) \subset \mathcal{A}, \tag{5.16}$$

whence

$$(\mathcal{A} + \mathcal{B}) \cap [0, r] \subset \mathcal{A}. \tag{5.17}$$

Now define

$$\mathcal{A}' := \mathcal{A} \cup ((a_0 + \mathcal{B}'') \cap [0, n]), \quad \mathcal{B}' := \mathcal{B} \setminus \mathcal{B}''.$$

With these definitions, $0 \in \mathcal{A} \cap \mathcal{B}$. Indeed $0 \in \mathcal{A} \subset \mathcal{A}'$, and if $0 \in \mathcal{B}''$ then $a_0 \notin \mathcal{A}$, which is absurd. Hence $0 \in \mathcal{B} \setminus \mathcal{B}'' =: \mathcal{B}'$. Since $\mathcal{B}'' \neq \emptyset$, property iii) is immediate, and to verify ii) we need only note that

$$\mathcal{A}' + \mathcal{B}' \subset (\mathcal{A} \cup (a_0 + \mathcal{B}'')) + \mathcal{B} \setminus \mathcal{B}'' \subset (\mathcal{A} + \mathcal{B}) \cup ((a_0 + \mathcal{B} \setminus \mathcal{B}'') + \mathcal{B}'') \subset \mathcal{A} + \mathcal{B}.$$

It remains only to establish i), i.e., to establish that (5.15) holds for each $m = 1, 2, \ldots, n$. For such m, observe

$$\begin{aligned}
&A'(m) + B'(m)\\
&\quad = A(m) + |\{b'' \in \mathcal{B}'' : 1 \leq b'' + a_0 \leq m\}| + B(m) - |\{b'' \in \mathcal{B}'' : 1 \leq b'' \leq m\}|\\
&\quad \geq A(m) + B(m) - |\{b'' \in \mathcal{B}'' : 1 \leq b'' \leq m, b'' + a_0 \geq m+1\}|\\
&\quad \geq A(m) + B(m) - |\{b \in \mathcal{B} : 1 \leq b \leq m, b + a_0 \geq m+1\}|. \qquad (5.18)
\end{aligned}$$

Let b_0 denote the smallest positive element of $\mathcal{B}$ contained in $[m - a_0 + 1, m]$; if no such element exists, then the final term of (5.18) drops out, and (5.15) follows from (5.13). Otherwise, (5.18) implies

$$A'(m) + B'(m) \geq A(m) + B(b_0 - 1).$$

Write $m = b_0 + r$, so that $0 \leq r < a_0 \leq n$. Since $r < n$, it follows from the induction hypothesis that

$$(A + B)(r) \geq cr.$$

(Otherwise $\mathcal{A} \cap [0, r]$, $\mathcal{B} \cap [0, r]$ would be a counterexample to Theorem 5.3.6 for $n' = r < n$.) Since $0 \in \mathcal{A} \cap \mathcal{B}$ and $c \leq 1$,

$$|(\mathcal{A} + \mathcal{B}) \cap [0, r]| = 1 + (A + B)(r) \geq 1 + cr \geq c(1 + r).$$

Moreover, since $r < a_0$, this result together with (5.17) shows $|\mathcal{A} \cap [0, r]| \geq c(1 + r)$. But (5.16) implies the interval $[b_0, b_0 + r]$ contains at least as many elements of $\mathcal{A}$ as $[0, r]$, so that

$$|\mathcal{A} \cap [b_0, b_0 + r]| \geq c(1 + r).$$

Consequently,

$$\begin{aligned}
A'(m) + B'(m) &\geq A(m) + B(b_0 - 1)\\
&= A(b_0 + r) - A(b_0 - 1) + A(b_0 - 1) + B(b_0 - 1)\\
&\geq c(1 + r) + c(b_0 - 1) = c(b_0 + r) = cm. \qquad \square
\end{aligned}$$

Corollary 5.3.7. *Let $\mathcal{A}, \mathcal{B}$ be sets of natural numbers, both containing* 0. *Suppose that $A(m) + B(m) \geq Cm$ for every positive integer m. Then*

$$\delta(\mathcal{A} + \mathcal{B}) \geq \min\{1, C\}. \qquad (5.19)$$

In particular, taking $C = \delta(\mathcal{A}) + \delta(\mathcal{B})$, we obtain

$$\delta(\mathcal{A} + \mathcal{B}) \geq \min\{1, \delta(\mathcal{A}) + \delta(\mathcal{B})\}. \qquad (5.20)$$

Proof. If $C = 0$ the assertion is trivial. Otherwise set $c = \min\{C, 1\}$, and note that by the preceding theorem applied to $\mathcal{A} \cap [0, n]$ and $\mathcal{B} \cap [0, n]$ (with n an arbitrary positive integer), we have $(A + B)(n) \geq cn$. $\square$

The analog of (5.20) for more than two summands follows immediately by induction. That the analog of (5.19) also holds for more than two summands follows from a theorem of Dyson [Dys45]. For a proof of this and a discussion of related results, see [HR83, Chapter 1].

We now prove a theorem of Lepson which illustrates that Mann's theorem is best possible for every value of $\delta(\mathcal{A}), \delta(\mathcal{B})$:

Theorem 5.3.8 (Lepson [Lep50]). *Let $0 \le \alpha \le 1$ and $0 \le \beta \le 1$. Then there exist $\mathcal{A}, \mathcal{B} \subset \mathbf{N}$, both containing 0, with $\delta(\mathcal{A}) = \alpha, \delta(\mathcal{B}) = \beta$, and*

$$\delta(\mathcal{A} + \mathcal{B}) = \min\{1, \alpha + \beta\}.$$

Proof. For each positive integer n, set $a_n := \lceil \alpha n! \rceil$. Let $\mathcal{A}$ consist of $0, 1$ together with each of the (possibly empty) blocks $n! + 1, n! + 2, \ldots, n! + a_{n+1} - a_n$. No two of these blocks intersect; indeed

$$0 \le a_{n+1} - a_n < (\alpha(n+1)! + 1) - \alpha n! \le (n+1)! - n! + 1,$$

so that

$$n! + a_{n+1} - a_n < (n+1)! + 1.$$

It follows that for each positive integer n,

$$\frac{A(n!)}{n!} = \frac{1 + \sum_{i=1}^{n-1} (a_{i+1} - a_i)}{n!} = \frac{a_n}{n!} = \frac{\lceil \alpha n! \rceil}{n!} \ge \alpha.$$

Now for each positive integer m, we have $A(m)/m \ge A(n!)/n! \ge \alpha$, if n is chosen maximal with $n! \le m$. On the other hand, $A(n!)/n! = \lceil \alpha n! \rceil / n! \to \alpha$ as $n \to \infty$. Hence $\delta(\mathcal{A}) = \alpha$. Now define b_n and $\mathcal{B}$ similarly, so that $\delta(\mathcal{B}) = \beta$, and set $\mathcal{C} := \mathcal{A} + \mathcal{B}$.

To prove $\delta(\mathcal{C}) \le \alpha + \beta$, we consider $C(n!)$ for $n = 2, 3, \ldots$. If $c \in \mathcal{C}$ satisfies $c \le n!$, then $c = a + b$, where $a \in \mathcal{A}, b \in \mathcal{B}$ and $a, b \le n!$. These inequalities force $a \le (n-1)! + a_n - a_{n-1}$, $b \le (n-1)! + b_n - b_{n-1}$, and thus

$$\begin{aligned} c &\le 2(n-1)! + a_n + b_n - a_{n-1} - b_{n-1} \\ &\le 2(n-1)! + a_n + b_n. \end{aligned}$$

Consequently,

$$\liminf_{n\to\infty} \frac{C(n!)}{n!} \le \liminf_{n\to\infty} \frac{2(n-1)! + a_n + b_n}{n!} = \alpha + \beta.$$

It follows that $\delta(\mathcal{C}) \le \alpha + \beta$. Since trivially $\delta(\mathcal{C}) \le 1$, we obtain

$$\delta(\mathcal{A} + \mathcal{B}) \le \min\{1, \delta(\mathcal{A}) + \delta(\mathcal{B})\}.$$

Combining this inequality with (5.20) yields the theorem. $\square$

5.3.3 Asymptotic Bases

Theorem 5.3.5 can be used to obtain a criterion for a set $\mathcal{A}$ to be an asymptotic basis of finite order in terms of its *lower density* $\underline{d}(A)$. However, formulating this criterion requires a bit of care.

It is natural to conjecture, in analogy with Theorem 5.3.5, that $\underline{d}(A) > 0$ implies $\mathcal{A}$ is an asymptotic basis of finite order. This is not quite the case; for instance, congruence considerations imply the set of nonnegative even integers cannot be an asymptotic basis of any order. A similar situation occurs whenever there is an integer $g > 1$ for which all the terms of $\mathcal{A}$ are congruent $\pmod g$. If we shift the set to include 0 (which certainly does not affect whether $\mathcal{A}$ is an asymptotic basis or not), then the existence of such a g is equivalent to the existence of a $g > 1$ dividing every term of $\mathcal{A}$. We prove:

Theorem 5.3.9. *Let $\mathcal{A} \subset \mathbf{N}$, $0 \in \mathcal{A}$ and $\underline{d}(\mathcal{A}) > 0$. Moreover, suppose* $\gcd(\mathcal{A}) = 1$. *Then $\mathcal{A}$ is an asymptotic basis of finite order.*

A slightly more general result in given in Exercise 5.3.6.

The proof of this theorem requires the following lemma from elementary number theory whose proof we leave as Exercise 5.3.5:

Lemma 5.3.10. *Let $a_1, \ldots, a_k$ be positive integers with* $\gcd(a_1, \ldots, a_k) = 1$. *There exists a positive integer $n_0 = n_0(a_1, \ldots, a_k)$ for which every integer $n \geq n_0$ admits a representation in the form*

$$n = a_1x_1 + \cdots + a_kx_k \quad (x_i \in \mathbf{N}).$$

Proof of Theorem 5.3.9. The ideal of $\mathbf{Z}$ generated by the elements of $\mathcal{A}$ is the unit ideal, hence there exist $a_1, \ldots, a_k \in \mathcal{A}$ with $\gcd(a_1, \ldots, a_k) = 1$. By discarding those $a_i = 0$, we can assume all of $a_1, \ldots, a_k$ are positive. Choose n_0 in accordance with the lemma and write

$$\begin{aligned} n_0 &= a_1x_1 + \cdots + a_kx_k, \\ n_0 + 1 &= a_1x_1' + \cdots + a_kx_k', \end{aligned}$$

where each x_i and x_i' is a natural number. Then (since $0 \in \mathcal{A}$) $n_0, n_0 + 1 \in l\mathcal{A}$ for the positive integer $l = \max\{\sum x_i, \sum x_i'\}$.

Now define $\mathcal{A}' := (l\mathcal{A} - \{n_0\}) \cap \mathbf{N}$. Then $1 \in \mathcal{A}'$ and $\underline{d}(\mathcal{A}') = \underline{d}(l\mathcal{A}) \geq \underline{d}(\mathcal{A}) > 0$. Thus $\mathcal{A}'$ is a basis of finite order h_1, say, by Theorem 5.3.5. But then

$$\mathbf{N} \subset h_1\mathcal{A}' \subset h_1(l\mathcal{A} - \{n_0\}) = h_1l\mathcal{A} - \{h_1n_0\}.$$

It follows that every $n \geq n_0h_1$ is a sum of h_1l elements of $\mathcal{A}$. $\square$

It is difficult to coax a bound for the *order* of the asymptotic basis for this proof. This is because though the statement of Theorem 5.3.9 involves only the natural density, the proof required a detour through Schnirelmann density. It is natural to seek a theorem which directly relates the lower density of a sumset to that of the component sets. The following deep result of Kneser [Kne53] is exactly such a theorem:

Theorem (Kneser). *If $\mathcal{A}_0, \ldots, \mathcal{A}_k \subset$ are infinite sets of natural numbers, then either*

$$\underline{d}(\mathcal{A}_0 + \cdots + \mathcal{A}_k) \geq \liminf_{n\to\infty} (A_0(n) + \cdots + A_k(n))/n \tag{5.21}$$

or there are positive integers $g, a_0, \ldots, a_k$ such that

i. each $\mathcal{A}_i$ is contained in the union $\mathcal{A}'_i$ of a_i distinct congruence classes $\pmod g$,

ii. there are at most finitely many positive members of $\mathcal{A}'_0 + \cdots + \mathcal{A}'_k$ not in $\mathcal{A}_0 + \cdots + \mathcal{A}_k$,

iii. $d(\mathcal{A}_0 + \cdots + \mathcal{A}_k) \geq (a_0 + \cdots + a_k - k)/g$.

Applied to the problem of estimating the order of an asymptotic basis, it yields:

Corollary (Kneser). *Under the assumptions of Theorem 5.3.9, the set $\mathcal{A}$ is an asymptotic basis of order not exceeding* $\max\{2, \lfloor 2/\underline{d}(\mathcal{A}) - 1\rfloor\}$.

For a development of Kneser's ideas and a proof of his theorem, see [HR83, Chapter 1, §7-10]. The corollary appears as [Ost56, Kapitel 14, Satz 5].

5.3.4 Exercises

Exercise 5.3.5. Prove Lemma 5.3.10. Suggestion: Show that every large integer in any fixed congruence class $\pmod{2a_1 \ldots a_k}$ admits a representation in the desired form.

Exercise 5.3.6 *(Ostmann [Ost56, Satz 4, §14]).* Let $\mathcal{A} = \{a_1 < a_2 < \ldots\} \subset \mathbf{N}$ and let d be the greatest common divisor of $a_2 - a_1, a_3 - a_2, \ldots$. Suppose that for some positive integer k, $\underline{d}(k\mathcal{A}) > 0$. Show that there exists an integer h such that $h\mathcal{A}$ contains all sufficiently large integers $n \equiv ha_1 \pmod d$.

Exercise 5.3.7. Using Mann's Theorem, improve the result of Exercise 5.3.3 to the following: If $\mathcal{A}$ is a set of natural numbers containing 0 with $\delta(\mathcal{A}) > 0$, then $\mathcal{A}$ is a basis of order at most $\lceil 1/\delta(\mathcal{A})\rceil$.

Exercise 5.3.8 *(continuation).* Show that for every $0 < \alpha \leq 1$, there is a set $\mathcal{A} \subset \mathbf{N}$ with $0 \in \mathcal{A}$ and $\delta(\mathcal{A}) = \alpha$ which is a basis of order $\lceil 1/\alpha\rceil$. Thus the estimate of the last exercise is best-possible.

Suggestion: review Lepson's construction.

Exercise 5.3.9. Let $\mathcal{A} = \{a_1 < a_2 < \ldots\}$ be an infinite collection of natural numbers with the following property: for every positive integer m, there exist infinitely many i for which $ma_i < a_{i+1}$. Show that $\mathcal{A}$ is not an asymptotic basis of finite order.

Exercise 5.3.10. Using the result of the previous exercise, show that there exists a set $\mathcal{A}$ of natural numbers containing 0 and 1 whose counting function satisfies $A(N) \geq N^{1/2}$ for all large N, but which is not a basis of finite order.

Exercise 5.3.11. Using Exercise 5.3.9, show that Theorem 5.3.9 is false if the condition that $\underline{d}(\mathcal{A}) > 0$ is replaced by $\overline{d}(\mathcal{A}) > 0$.

Exercise 5.3.12 *(Carroll [Car00]).* Suppose the positive integers are partitioned into finitely many sets. Must one of them, say $\mathcal{A}_i$, have the property that $\mathcal{A}_i \cup \{0\}$ is an asymptotic basis of finite order?

Exercise 5.3.13 *(Stöhr [Stö55, Kriterium 8], †).* Show that if $a_1, a_2, \dots$ is a p-adically convergent sequence of integers, then $\{a_i\}$ is not an asymptotic basis of finite order. Prove the same for any finite union of p-adically convergent sequences (with respect to the same p).

Exercise 5.3.14 *(Ostmann [Ost56, p.27]).* For $\mathcal{A} \subset \mathbf{N}$, let $f(\mathcal{A})$ be the least integer $h \geq 2$ for which $h\mathcal{A} \cap \mathcal{A} \neq \emptyset$. That is, $f(\mathcal{A})$ is the least h for which there exists a solution to $a_1 + \cdots + a_h = a_{h+1}$ with each $a_i \in \mathcal{A}$. Using Lemma 5.3.10, prove that $f(\mathcal{A})$ exists for every infinite set $\mathcal{A}$. Ostmann calls $f(\mathcal{A})$ the *Fermat index* of $\mathcal{A}$, since Fermat's Last Theorem is equivalent to the assertion that $f(\{n^k : n = 1, 2, \dots\}) > 2$ for $k \geq 3$.

5.4 Densities of Particular Sumsets

Up to this point our investigation of sumsets and their densities has been rather general. We now turn to particular and particularly striking special cases.

We commence our discussion with a result of Wirsing to the effect that if f is a "smooth" function whose second derivative satisfies certain (stringent) inequalities, then $\mathcal{A} := \{f(n) : n = 1, 2, \dots\}$ satisfies $\underline{d}(2\mathcal{A}) > 0$.

We then turn to Schnirelmann's remarkable theorem that every $n > 1$ is a sum of at most C primes, for an absolute constant C.

Central in the proof of Schnirelmann's theorem is the proof that $\{p + q : p, q \text{ prime}\}$ has positive lower density. The next result we discuss, due to Romanov, has a similar flavor: for fixed $a \geq 2$, the set $\{p + a^k : p \text{ prime}, k \geq 1\}$ has positive lower density. We present the proof of Romanov's result as exposited by Nathanson [Nat96]. We then discuss some theorems of Erdős and Crocker related to the special case $a = 2$.

5.4.1 A Special Class of Additive Bases

The following is a special case of a theorem of Wirsing (see [Ost56, §14, Satz 11]):

Theorem 5.4.1 (Wirsing). *Let $f : [1, \infty) \to [0, \infty)$ be a twice continuously differentiable function with $f(x) \to \infty$ as $x \to \infty$, and suppose that for some constants $C_1, C_2 > 0$,*

$$0 < C_1 x^{-\beta} \leq f''(x) \leq C_2 x^{-\alpha} \quad (x \geq 1), \tag{5.22}$$

where α, β satisfy $0 < \alpha < 1$ and $\alpha \leq \beta \leq 3\alpha - 1$.

Suppose the elements of the infinite subset $\mathcal{A} = \{a_1 < a_2 < \dots\} \subset \mathbf{N}$ satisfy $a_n = f(n) + O(1)$ for $n \geq 1$. Then $\underline{d}(2\mathcal{A}) > 0$.

Proof. We will actually prove the stronger statement that the gaps between consecutive elements of $2\mathcal{A}$ are bounded.

For this, let $\gamma \geq 1$ be a real number whose precise value will be chosen later, and set $m = m(n) = \lfloor \gamma n^\alpha \rfloor$. Since $\alpha < 1$, we have $n > m$ for all large n, say $n \geq n_0 = n_0(\gamma)$. (In what follows, we suppress the dependence of constants on the parameters in the theorem statement, but we note explicitly any dependence on γ.)

For $n > n_0$, we write down the following list of elements of $2\mathcal{A}$:

$$2a_n, a_{n+1} + a_{n-1}, a_{n+2} + a_{n-2}, \ldots, a_{n+m} + a_{n-m}. \tag{5.23}$$

The gaps between consecutive terms of this sequence are bounded: Indeed, by repeated application of the mean value theorem,

$$\begin{aligned}
&a_{n+v+1} - a_{n-v-1} - a_{n+v} - a_{n-v} \\
&\qquad = f(n+v+1) - f(n+v) - (f(n-v) - f(n-v-1)) + O(1) \\
&\qquad = f'(n+v+\theta_1) - f'(n-v-\theta_2) + O(1) = 2vf''(n+\theta_3) + O(1) \\
&\qquad \leq 2mC_2(n+\theta_3)^{-\alpha} + O(1) \leq 2\gamma C_2 \left(\frac{n}{n-\gamma n^\alpha}\right)^\alpha + O(1),
\end{aligned}$$

where $\theta_1, \theta_2 < 1$ and $|\theta_3| \leq m$. Since $\alpha < 1$, the first term on the right tends to $2\gamma C_2$ as $n \to \infty$, and thus is $O_\gamma(1)$ for $n \geq n_0$.

We now show the final element of (5.23) exceeds $2a_{n+1}$ if our initial choice of γ is sufficiently large. We have

$$\begin{aligned}
a_{n-m} + a_{n+m} - 2a_n &= f(n-m) + f(n+m) - 2f(n) + O(1) \\
&= \int_0^m \int_0^m f''(n-m+x+y)\,dx\,dy + O(1) \\
&\geq \int_0^m \int_0^m \frac{C_1}{(n+m)^\beta}\,dx\,dy + O(1) = C_1 \frac{m^2}{(n+m)^\beta} + O(1) \\
&\geq C_1 \frac{(\gamma n^\alpha - 1)^2}{(2n)^\beta} + O(1) \sim C_1 \frac{\gamma^2}{2^\beta} n^{2\alpha-\beta} \quad (n \to \infty),
\end{aligned}$$

since $2\alpha - \beta \geq 1 - \alpha > 0$. But we also have (with some $\theta < 1$)

$$\begin{aligned}
2a_{n+1} - 2a_n &= 2(f(n+1) - f(n)) + O(1) \\
&= 2f'(n+\theta) + O(1) \leq 2f'(n+1) + O(1) \\
&= 2\int_1^{n+1} f''(x)\,dx + O(1) \leq 2C_2 \int_1^{n+1} \frac{dx}{x^\alpha} + O(1) \\
&= 2\frac{C_2}{1-\alpha} n^{1-\alpha} + O(1).
\end{aligned}$$

Since $2\alpha - \beta \geq 1 - \alpha$, we see that if γ is chosen so that

$$C_1 \gamma^2 2^{-\beta} > 2C_2(1-\alpha)^{-1},$$

then $a_{n-m} + a_{n+m} - 2a_n > 2a_{n+1} - 2a_n$ for all $n \geq n_1(\gamma)$. That is, we have

$$a_{n-m} + a_{n+m} > 2a_{n+1} \quad (n \geq n_1(\gamma)). \tag{5.24}$$

Finally, choose a positive integer $n_2 \geq n_1$ such that $f(x)$ is increasing for $x \geq n_2$; this is possible because $f(x) \to \infty$ and $f''(x) > 0$.

To complete the proof that $2\mathcal{A}$ has bounded gaps, it suffices to show that starting at some element of $2\mathcal{A}$ you can walk to infinity along the elements of $2\mathcal{A}$ taking only steps of bounded length. We do this by showing that for $n \geq n_2$, one can walk from a_n to a_{n+1} taking steps of only bounded length; since $a_n = f(n) + O(1) \to \infty$, the result follows.

To prove this last claim, notice that if $a_{n+1} \leq a_n$, then

$$0 \leq a_n - a_{n+1} \leq f(n) - f(n+1) + O(1) \leq O(1),$$

so that walking from a_n to a_{n+1} is a step of length bounded independently of n. If $a_{n+1} > a_n$, then we instead walk along consecutive terms of the sequence (5.23), stopping just before the first term exceeding a_{n+1}; the inequality (5.24) guarantees the existence of such a stopping point. Walking from this term to the next is a step of bounded length (where the bound perhaps depends on γ), so walking from this term to the nearer point a_{n+1} is as well. □

To illustrate, let

$$f(x) = x \log x \quad \text{and} \quad a_n = \lfloor n \log n \rfloor.$$

Then $f''(x) = 1/x$, so that (5.22) holds (with $C_1 = C_2 = 1$) for any choice of $\alpha < 1 < \beta$. We can satisfy the additional hypothesis $\beta \leq 3\alpha - 1$ with, e.g., the choice $\alpha = 3/4, \beta = 9/4$. We therefore conclude that a positive proportion of the natural numbers can be written as a sum of two integers of the form $\lfloor n \log n \rfloor$ $(n \geq 1)$. The set of integers of this form contains 0 and 1, so that Schnirelmann's basis theorem implies this is a basis of finite order for the natural numbers.

One consequence of the prime number theorem is that the nth prime p_n satisfies $p_n \sim n \log n$; on our way to the proof of Schnirelmann's theorem, we will see that, as this example suggests, the set of integers that can be written as a sum of two primes also has positive (lower) density.

Exercise 5.4.1. Suppose $1 < c \leq 3/2$. Show that $\underline{d}(2\mathcal{A}) > 0$, where $\mathcal{A} := \{[n^c] : n = 1, 2, \dots\}$. Conclude that $\mathcal{A}$ is an asymptotic basis of finite order.

5.4.2 Schnirelmann's Contribution to Goldbach's Problem

Let $r(N)$ denote the number of solutions to $N = p + q$, where p, q are primes. Goldbach's conjecture is that $r(N) > 0$ for even $N \geq 4$.

In Chapter 3, §3.5.3 we established the *upper bound*

$$r(N) \ll \frac{N}{\log^2 N} \prod_{p \mid N} \left(1 + \frac{1}{p}\right). \tag{5.25}$$

We actually established (5.25) only for *even* $N \geq 2$, but since $r(N) \leq 2$ for every odd positive integer N, (5.25) is valid for *every* $N \geq 2$.

An upper bound for $r(N)$ may seem more like an anti-Goldbach theorem. Actually, this upper bound is the main analytic input needed to achieve our goal in this section: proving Schnirelmann's result that every $n > 1$ is a sum of at most C primes for some constant C.

The plan of attack is as follows: The upper bound (5.25) enables us to bound above the mean square of $r(N)$ over $N \leq x$; Chebyshev's lower bound (Chapter 1, §1.5) for $\pi(x)$ gives us a lower bound on the mean of $r(N)$ over the same range, and the Schwarz inequality allows us to deduce from these two results a lower bound on the number of $n \leq x$ for which $r(N) > 0$, i.e., a lower bound on the counting function of the set of integers which are sums of two primes. Applying the results of the last section to this set yields the result. We now fill in the details.

Lemma 5.4.2. *As $x \to \infty$,*

$$\sum_{N \leq x} r(N)^2 \ll \frac{x^3}{\log^4 x}.$$

Proof. Substituting the upper bound (5.25), noting that $N/\log^2 N \ll x/\log^2 x$ uniformly for $2 \leq N \leq x$, we find

$$\begin{aligned}\sum_{N \leq x} r(N)^2 &\ll \sum_{2 \leq N \leq x} \left(\frac{N}{\log^2 N} \prod_{p|N}\left(1 + \frac{1}{p}\right)\right)^2 \\ &\ll \frac{x^2}{\log^4 x} \sum_{2 \leq N \leq x} \left(\prod_{p|N}\left(1 + \frac{1}{p}\right)\right)^2 \ll \frac{x^2}{\log^4 x} \sum_{2 \leq N \leq x} \left(\sum_{d|N} \frac{1}{d}\right)^2.\end{aligned}$$

It remains to show the outer sum on the right is $O(x)$. But owing to the trivial inequality

$$[d_1, d_2] \geq \max\{d_1, d_2\} \geq (d_1 d_2)^{1/2},$$

we have

$$\begin{aligned}\sum_{N \leq x} \left(\sum_{d|N} \frac{1}{d}\right)^2 &= \sum_{N \leq x} \sum_{d_1|N} \sum_{d_2|N} \frac{1}{d_1 d_2} = \sum_{d_1, d_2 \leq x} \sum_{\substack{N \leq x \\ d_1|N, d_2|N}} 1 \\ &\leq \sum_{d_1, d_2 \leq x} \frac{1}{d_1 d_2} \frac{x}{[d_1, d_2]} \leq x \sum_{d_1, d_2 \leq x} \frac{1}{(d_1 d_2)^{\frac{3}{2}}} \leq x \left(\sum_{d=1}^{\infty} d^{-\frac{3}{2}}\right)^2 \ll x.\end{aligned}$$

$\square$

Lemma 5.4.3. *As $x \to \infty$,*

$$\sum_{N \le x} r(N) \gg \frac{x^2}{\log^2 x}.$$

Proof. From Chebyshev's theorem we have $\pi(x) \gg x/\log x$ as $x \to \infty$. Thus

$$\begin{aligned}\sum_{N \le x} r(N) = \sum_{N \le x} \sum_{p+q=N} 1 &= \sum_{p+q \le x} 1 \\ &\ge \sum_{p \le x/2, q \le x/2} 1 = \pi(x/2)^2 \gg \frac{(x/2)^2}{\log^2 (x/2)} \gg \frac{x^2}{\log^2 x}. \quad \square\end{aligned}$$

Lemma 5.4.4. *Let $\mathcal{A}, \mathcal{B} \subset \mathbf{N}$ and let $r_{\mathcal{A},\mathcal{B}}(n)$ denote the number of solutions to $n = a + b$ with $a \in \mathcal{A}, b \in \mathcal{B}$. Set $\mathcal{S} := \mathcal{A} + \mathcal{B}$. Then for $x \ge 1$,*

$$\left(\sum_{n \le x} r_{\mathcal{A},\mathcal{B}}(n) \right)^2 \le S(x) \sum_{n \le x} r_{\mathcal{A},\mathcal{B}}(n)^2.$$

Proof. By the Cauchy-Schwarz inequality,

$$\begin{aligned}\left(\sum_{n \le x} r_{\mathcal{A},\mathcal{B}}(n) \right)^2 &= \left(\sum_{\substack{n \le x \\ r_{\mathcal{A},\mathcal{B}}(n) > 0}} r_{\mathcal{A},\mathcal{B}}(n) \right)^2 \\ &\le \left(\sum_{\substack{n \le x \\ r_{\mathcal{A},\mathcal{B}}(n) > 0}} 1 \right) \left(\sum_{\substack{n \le x \\ r_{\mathcal{A},\mathcal{B}}(n) > 0}} r_{\mathcal{A},\mathcal{B}}(n)^2 \right) = S(x) \sum_{n \le x} r_{\mathcal{A},\mathcal{B}}(n)^2. \quad \square\end{aligned}$$

Theorem 5.4.5. *Let $\mathcal{S}$ be the set of integers representable as a sum of two primes. Then $\underline{d}(\mathcal{S}) > 0$.*

Proof. As $x \to \infty$, we have by Lemmas 5.4.3, 5.4.2, 5.4.4,

$$S(x) \ge \left(\sum_{n \le x} r_{\mathcal{A}}(n) \right)^2 \Big/ \sum_{n \le x} r_{\mathcal{A}}(n)^2 \gg \left(\frac{x^2}{\log^2 x} \right)^2 \Big/ \frac{x^3}{\log^4 x} \gg x. \quad \square$$

As a corollary, we obtain:

Theorem 5.4.6. *There exists a constant C such that every integer $n > 1$ is a sum of at most C primes.*

Proof. Let $\mathcal{A}_0$ be the subset of the natural numbers consisting of 0 together with the primes, and let $\mathcal{A} := 2\mathcal{A}_0$. Then $0 \in \mathcal{A}$, $\gcd(\mathcal{A}) = 1$ (since $\mathcal{A}$ contains the primes), and Theorem 5.4.5 implies $\underline{d}(\mathcal{A}) > 0$. Thus $\mathcal{A}$ is an asymptotic basis of finite order h, say, and every $n \ge n_0$ is a sum of at most $c_0 := 2h$ primes, for some positive integer n_0. But each of the finitely many n, $1 < n < n_0$, can be written as a sum of at most c_1 primes for some c_1, using only the primes $2, 3$. Thus we may take $C := \max\{c_0, c_1\}$. $\square$

The same proof we have described in this section yields a similar result for any "thick" subset of the primes:

Theorem 5.4.7 (Landau [Lan30], see also [Nat87]). *If $\mathcal{P}$ is a subset of the primes for which*

$$\pi_{\mathcal{P}}(x) := \#\{p \le x : p \in \mathcal{P}\} \gg x/\log x \quad (x \to \infty),$$

then there exist constants C, n_0 (depending on $\mathcal{P}$) for which every $n \ge n_0$ is a sum of at most C primes $p \in \mathcal{P}$.

As an example, our results in Chapter 2 imply we may take for $\mathcal{P}$ the set of primes congruent to $a \pmod q$ for any integers a, q with $q > 0$ and $\gcd(a, q) = 1$ (cf. Chapter 2, Exercise 2.6.1).

Exercise 5.4.2. Prove Theorem 5.4.7. Hint: For any set of primes $\mathcal{P}$, we have $r_{\mathcal{P}}(n) \le r(n)$, so that the analog of Lemma 5.4.2 is immediate.

We close this section by mentioning a celebrated 1937 result of Vinogradov:

Vinogradov's Three Primes Theorem. *Let $r(N)$ denote the number of ordered ways of writing N as a sum of three primes. As $N \to \infty$ through odd integers, we have*

$$r(N) \sim \prod_p \left(1 + \frac{1}{(p-1)^3}\right) \prod_{p|N} \left(1 - \frac{1}{p^2 - 3p + 3}\right) \frac{N^2}{2\log^3 N}.$$

As a consequence, the primes form an asymptotic basis of order at most 4. Though this theorem has the same flavor as Theorem 5.4.6, the proof (by the circle method; cf. [Nat96, Chapter 8]) utilizes entirely different methods and requires much more significant input from analytic number theory (specifically a version of the prime number theorem on arithmetic progressions with a large range of uniformity in the modulus).

5.4.3 Romanov's Theorem

In this section we fix a positive integer $a \ge 2$; all implied constants may depend on this choice of a.

Lemma 5.4.8. *As $n \to \infty$, we have*

$$\sum_{d|n} \frac{1}{d} \ll \log\log n.$$

Proof. Let $k = \nu(n)$. Since

$$n = \prod_{p^{e_p} \| n} p^{e_p} \ge \prod_{p^{e_p} \| n} 2 = 2^k,$$

we have

$$k \le \log n / \log 2. \tag{5.26}$$

Then with p_i denoting the ith prime,

$$\sum_{d|n} \frac{1}{d} = \prod_{p^{e_p}\|n} \left(1 + \frac{1}{p} + \cdots + \frac{1}{p^{e_p}}\right)$$
$$\leq \prod_{i=1}^{k} \left(1 + \frac{1}{p_i} + \frac{1}{p_i^2} + \dots\right) = \prod_{i=1}^{k} \left(1 - \frac{1}{p_i}\right)^{-1} \ll \log p_k, \tag{5.27}$$

by Mertens' Theorem (cf. Chapter 1, §1.7). By Chebyshev's lower bound for $\pi(x)$ (cf. Chapter 1, §1.5),

$$\sqrt{p_k} \ll \frac{p_k}{\log p_k} \ll \pi(p_k) = k,$$

whence $p_k \ll k^2$, and

$$\log p_k \leq 2\log k + O(1) \ll \log k + 1 \ll \log\log n,$$

by (5.26). Substituting into (5.27) gives the result. □

Lemma 5.4.9. *Let $e(d)$ denote the order of a $\pmod d$. Then*

$$\sum_{\substack{d=1\\(a,d)=1}}^{\infty} \frac{1}{de(d)} < \infty.$$

Proof. Set $D(x) = \prod_{k\leq x} (a^k - 1)$, where we think of x as large. Then

$$D(x) < \prod_{k\leq x} a^k \leq a^{x^2}.$$

If $e(d) = k$ for some $k \leq x$, then $d \mid a^k - 1 \mid D(x)$. Thus, by Lemma 5.4.8,

$$E(x) := \sum_{k\leq x} \sum_{\substack{e(d)=k\\(a,d)=1}} \frac{1}{d} \leq \sum_{d|D(x)} \frac{1}{d} \ll \log\log D(x) \ll \log(x^2 \log a) \ll \log x$$

for large x. It follows that for all $t \geq 1$, $E(t) \ll 1 + \log t$. By partial summation, we have for $x \geq 1$,

$$\sum_{k\leq x} \frac{1}{k} \left(\sum_{\substack{e(d)=k\\(a,d)=1}} \frac{1}{d} \right) = \int_{1/2}^{x} \frac{dE(t)}{t} = \frac{E(x)}{x} + \int_1^x \frac{E(t)}{t^2} dt$$
$$\ll \frac{1+\log x}{x} + \int_1^x \frac{1+\log t}{t^2} dt \ll 1.$$

Consequently,

$$\sum_{\substack{d=1\\(a,d)=1}}^{\infty} \frac{1}{de(d)} = \sum_{k=1}^{\infty} \frac{1}{k} \left(\sum_{\substack{e(d)=k\\(a,d)=1}} \frac{1}{d} \right) < \infty,$$

as was to be shown. □

Before proceeding, we recall the following sieve result from Chapter 3, §3.5.3: For $x \geq 2$,

$$\pi_N(x) := \#\{p \leq x : p + N \text{ prime}\} \ll \frac{x}{\log^2 x} \prod_{p|N} \left(1 + \frac{1}{p}\right), \tag{5.28}$$

uniformly for even $N \geq 2$.

Lemma 5.4.10. *Let $r(N)$ denote the number of solutions to $N = p + a^k$, where p is prime and k is a positive integer. Then*

$$\sum_{N \leq x} r(N)^2 \ll x \quad (x \to \infty).$$

Proof. Since $r(N)^2$ counts the number of solutions to $p_1 + a^{k_1} = p_2 + a^{k_2} = N$, $\sum_{N \leq x} r(N)^2$ counts the number of solutions to

$$p_1 + a^{k_1} = p_2 + a^{k_2} \leq x \quad (p_1, p_2 \text{ prime}, k_1, k_2 \in \mathbf{Z}^+). \tag{5.29}$$

In any such solution, one has $k_1, k_2 \leq \log x / \log a$ and $p_1, p_2 \leq x$. The number of solutions with $k_1 = k_2$ is thus bounded by

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

It remains to establish the same estimate for the solutions with $k_1 \neq k_2$; by symmetry, we can restrict our attention to the case $k_1 < k_2$.

For fixed k_1 and k_2 with $k_1 < k_2$, the number of solutions to (5.29) is bounded above by the number of $p_1 \leq x$ with $p_1 = p_2 + (a^{k_2} - a^{k_1})$; by (5.25), this latter quantity is

$$\ll \frac{x}{\log^2 x} \prod_{p|a^{k_2} - a^{k_1}} \left(1 + \frac{1}{p}\right) \ll \frac{x}{\log^2 x} \prod_{p|a^{k_2-k_1}-1} \left(1 + \frac{1}{p}\right).$$

For $k_1 < k_2 \leq \log x / \log a$, one has $k_2 - k_1 \leq \log x / \log a$, and for any fixed $j \leq \log x / \log a$, there are not more than $\log x / \log a$ pairs $k_1 < k_2$ with $k_2 - k_1 = j$. It follows that the number of solutions to (5.29) for $k_1 < k_2$ is

$$\begin{aligned} \ll \frac{x}{\log^2 x} \log x \sum_{j \leq \frac{\log x}{\log a}} \prod_{p|a^j-1} \left(1 + \frac{1}{p}\right) &\ll \frac{x}{\log x} \sum_{j \leq \frac{\log x}{\log a}} \sum_{e(d)|j} \frac{1}{d} \\ &\ll \frac{x}{\log x} \sum_{e(d) \leq \frac{\log x}{\log a}} \frac{1}{d} \frac{\log x}{e(d) \log a} \ll x, \end{aligned}$$

by Lemma 5.4.9. □

Lemma 5.4.11. *With $r(N)$ as in the last Lemma, we have*

$$\sum_{N \leq x} r(N) \gg x \quad (x \to \infty).$$

Proof. Indeed, as $x \to \infty$,

$$\sum_{N \le x} r(N) = \sum_{N \le x} \sum_{p+a^k=N} 1 = \sum_{p+a^k \le x} 1 \ge \sum_{p \le x/2, a^k \le x/2} 1$$
$$= \pi(x/2) \left\lfloor \frac{\log(x/2)}{\log a} \right\rfloor \gg \frac{x}{\log x} \log x = x. \qquad \square$$

Theorem 5.4.12 (Romanov). *The set $\mathcal{S} := \{p + a^k : p$ prime and $k \ge 1\}$ has positive lower density.*

Proof. Taking $\mathcal{A}$ as the set of primes and $\mathcal{B}$ as the set of positive powers of a in Lemma 5.4.4 yields

$$\left(\sum_{N \le x} r(N) \right)^2 \le S(x) \sum_{N \le x} r(N)^2.$$

Thus by Lemmas 5.4.10 and 5.4.11, we have as $x \to \infty$,

$$S(x) \gg \left(\sum_{N \le x} r(N) \right)^2 \Big/ \sum_{N \le x} r(N)^2 \gg x^2/x = x. \qquad \square$$

Erdős has shown the following more general result:

Theorem (Erdős [Erd50]). *Let $a_1 < a_2 < \dots$ be an infinite sequence of integers satisfying $a_k \mid a_{k+1}$. Then a necessary and sufficient condition that the sequence $p + a_k$ should have positive density is that*

$$\limsup_{k \to \infty} \frac{\log a_k}{k} < \infty \quad \text{and} \quad \sum_{d \mid a_i} \frac{1}{d} \ll 1.$$

5.4.4 Theorems of Erdős and Crocker on the Sum of a Prime and Powers of 2

Only $O(\log x) = o(x)$ *even* integers up to x can be expressed in the form $p + 2^k$. Thus Romanov's theorem implies a positive proportion of *odd* integers admit such a representation. One consequence of our next theorem is that a positive proportion of odd integers *cannot* be so expressed. It is an open question of Erdős to to decide whether the set of odd integers admitting such a representation possesses a natural density.

Theorem 5.4.13 (Erdős). *There exists an infinite arithmetic progression of odd numbers none of which can be written as the sum of a prime and a positive power of 2.*

Proof. We begin by noting that every integer k falls into one of the following congruence classes:

$$\begin{aligned} &0 \pmod{2}, \quad 0 \pmod{3}, \quad 1 \pmod{4}, \\ &3 \pmod{8}, \quad 7 \pmod{12}, \quad 23 \pmod{24}. \end{aligned} \tag{5.30}$$

This is easily be checked by considering the various possibilities for $k \pmod{24}$. From this, we deduce that every power of 2 falls into one of the corresponding congruence classes

$$\begin{aligned} &2^0 \pmod{3}, \quad 2^0 \pmod{7}, \quad 2^1 \pmod{5}, \\ &2^3 \pmod{17}, \quad 2^7 \pmod{13}, \quad 2^{23} \pmod{241}. \end{aligned} \tag{5.31}$$

To see this, note for example that if $k \equiv 0 \pmod{2}$ then $2^k \equiv 2^0 \pmod{3}$, since $\mathrm{ord}_3 2 \mid 2$. In general, each congruence $k \equiv a_i \pmod{m_i}$ of (5.30) implies a corresponding congruence (of (5.31)) $2^k \equiv 2^{a_i} \pmod{p_i}$ for a prime p_i with $\mathrm{ord}_{p_i} 2 \mid m_i$.

Now let $a \pmod{P}$ be the congruence class mod

$$P := 2^9 \cdot 3 \cdot 7 \cdot 6 \cdot 17 \cdot 13 \cdot 241$$

which is the intersection of the congruence classes (5.31) with the congruence class $1 \pmod{2^9}$. We claim the positive integers in this class meet the conditions of the theorem. If $m \equiv a \pmod{P}$ then $\gcd(m - 2^k, P) > 1$ for every positive integer k, by our discussion above, so it only remains to show that we cannot have $m - 2^k = p$ for some prime $p \mid P$.

This is taken care of by the additional congruence $m \equiv 1 \pmod{2^9}$. Indeed, suppose $m - 2^k = p$ for some $p \mid P$. If $k \geq 9$, then looking at this as a congruence $\pmod{2^9}$ shows $1 \equiv m \equiv p \pmod{2^9}$, which is false as $1 < p < 2^9$ for each $p \mid P$. If $k < 9$, then

$$1 < m = 2^k + p \leq 2^8 + p < 2^8 + 2^8 = 2^9,$$

which contradicts the congruence $m \equiv 1 \pmod{2^9}$. □

With two powers of 2 instead of one, the situation is less well-understood. For distinct powers, Crocker showed:

Theorem 5.4.14 (Crocker [Cro71]). *For $n \geq 3$, $2^{2^n} - 1$ cannot be expressed as the sum of a prime and two distinct positive powers of 2.*

Proof. Let $n \geq 3$; we want to show that for $a < b < 2^n$, the numbers $2^{2^n} - 1 - 2^a - 2^b$ are all composite. Choose r with $2^r \| b - a$ (so that $r < n$) and then note that

$$2^{2^r} + 1 \mid (2^{2^n} - 1) - 2^a(1 + 2^{b-a}).$$

Indeed, the left hand side divides both terms on the right. But since $n \geq 3$,

$$\begin{aligned} 2^{2^n} - 1 - 2^a - 2^b &\geq 2^{2^n} - 1 - 2^{2^n - 1} - 2^{2^n - 2} \\ &= 2^{2^n - 2} - 1 > 2^{2^{n-1}} + 1 \geq 2^{2^r} + 1, \end{aligned}$$

implying that the factor we discovered above is nontrivial. □

In the same paper, Crocker exhibited an infinite family of odd integers that could not be expressed as the sum of a prime and two positive powers of two, distinct or otherwise. It is an open question to decide whether the set of such odd integers has positive lower density.

Exercise 5.4.3 *(Crocker [Cro61]).* Show that none of the integers $2^{2^n} - 5$, $n \geq 3$, can be written as the sum of a prime and a positive power of 2.

Exercise 5.4.4 *(Sierpiński ([Sie60], [Sie88, p. 446]) and Riesel ([Rie56])).* Let $P_1 = 3 \cdot 5 \cdot 7 \cdot 13 \cdot 17 \cdot 241$.

a) Referring to the proof of Theorem 5.4.13, show that there exist infinitely many positive integers n for which $\gcd(n + 2^k, P_1) > 1$ for every (nonnegative) integer k.

b) Show that for these n, we also have $\gcd(n \cdot 2^j + 1, P_1) > 1$ for every j. From this show there are infinitely many n for which $n \cdot 2^j + 1$ is composite for every j. These are called *Sierpinski numbers* (of the second kind).

c) Prove the cognate result that there are infinitely many (positive) n for which $n \cdot 2^j - 1$ is composite for every j; these are called *Riesel numbers*.

Exercise 5.4.5 *(†).* Prove that there are infinitely many even k not divisible by 3 with the property that $3^n + k$ is composite for all positive integers n. (For the analog with an arbitrary positive integer in place of 3, see Aigner [Aig61].)

5.5 Schur's Regularity Lemma

> We have here a statement of the type: 'if a system is partitioned arbitrarily into a finite number of subsystems, then at least one subsystem possesses a certain specified property.' To the best of my knowledge, there is no earlier result which bears even a remote resemblance to Schur's theorem. It is this element of novelty that impresses itself so forcefully on the reader. − L. Mirsky, *The Combinatorics of Arbitrary Partitions* [Mir75]

5.5.1 A Lemma in Graph Theory

The formulation of Schur's proof that we present here utilizes ideas from graph theory. Our presentation follows Mirsky [Mir75], who credits Greenwood & Gleason and Abbott & Moser.

We first recall a few basic definitions from the subject: A graph G is an ordered pair (V, E), where V (the set of *vertices*) is a set, and E (the set of *edges*) is a collection of two-element subsets V. A graph (V, E) is said to be a *complete graph on n vertices* if $|V| = n$ and E contains every pair of distinct elements of V. We say a graph is *edge-colored* (hereafter, simply *colored*) if we assign a color to each element of E. By a *triangle* in a complete graph, we mean a set of three of vertices $\{v_r, v_s, v_t\} \subset V$ together with the three edges $\{v_r, v_s\}, \{v_r, v_t\}, \{v_s, v_t\} \in E$.

We use $\mathcal{G}(n, k)$ to denote the set of k-colored complete graphs on n vertices.

For $k = 1, 2, \ldots$, set $\mu_k := \lfloor k!e \rfloor$. We shall deduce Schur's lemma as a corollary of the following lemma on graph colorings:

Lemma 5.5.1. *Let $k \geq 1$. Then every graph in $\mathcal{G}(\mu_k + 1, k)$ contains a monochromatic triangle.*

We first observe:

Lemma 5.5.2. *For $k \geq 2$, we have the recurrence relation*

$$\mu_k = k\mu_{k-1} + 1.$$

Proof. For $k = 1, 2, \ldots$, set

$$v_k := k! \left(1 + \frac{1}{1!} + \frac{1}{2!} + \cdots + \frac{1}{k!}\right).$$

Then v_k is an integer with $v_k \leq k!e$; moreover,

$$\begin{aligned} k!e - v_k &= \frac{1}{k+1} + \frac{1}{(k+1)(k+2)} + \ldots \\ &< \frac{1}{(k+1)}\left(1 + \frac{1}{k+1} + \frac{1}{(k+1)^2} + \ldots\right) = 1/k \leq 1. \end{aligned}$$

Consequently, $\mu_k = v_k$ (for $k \geq 1$). The result follows. □

Proof of Lemma 5.5.1. The proof is by induction on k. When $k = 1$, we are considering 1-colorings of a complete graph on 3 vertices, and the result is obvious.

Suppose the result holds for $k-1$ (with $k \geq 2$). Let G be a k-colored complete graph on $\mu_k + 1$ vertices, and label the vertices with the integers $1, 2, \ldots, \mu_k + 1$. Consider the $\mu_k = k\mu_{k-1} + 1$ edges connected to the vertex labeled 1; by the pigeonhole principle, at least $\mu_{k-1} + 1$ of these must share the same color, say red. By relabeling, we can assume these edges are

$$\{1, 2\}, \{1, 3\}, \ldots, \{1, \mu_{k-1} + 2\}.$$

If there is a red edge between any two of the vertices $2, 3, \ldots, \mu_{k-1} + 2$, say between i, j, then clearly $\{1, i, j\}$ is a monochromatic triangle.

Otherwise, the subgraph on $\{2, 3, \ldots, \mu_{k-1} + 2\}$ is a complete graph on $\mu_{k-1} + 1$ vertices colored with only $k - 1$ colors. By the induction hypothesis, this subgraph contains a monochromatic triangle, so G does as well. □

5.5.2 Combinatorial Consequences

Lemma 5.5.3. *Let $a_1 < a_2 < \cdots < a_N$ be a sequence of $N \geq \mu_k + 1$ positive integers. If the integers $1, 2, \ldots, a_N - 1$ are k-colored, there is always a monochromatic solution to $x - y = z$, with $1 \leq x, y, z \leq a_N - 1$.*

Moreover, if no three of the a_i are in arithmetic progression, then x, y, z can be chosen to be distinct.

Proof. Without loss of generality, we may assume $N = \mu_k + 1$.

We k-color the complete graph with vertices $\{a_1, \ldots, a_N\}$ according to the rule that the edge $\{a_i, a_j\}$ (with $i < j$) receives the color of $a_j - a_i$. By Lemma 5.5.1, there is a monochromatic triangle connecting the vertices a_i, a_j, a_k for some $1 \leq i < j < k \leq N$. Then $x = a_k - a_i, y = a_j - a_i, z = a_k - a_j$ gives a monochromatic solution to $x - y = z$.

Clearly $x \neq y, x \neq z$. If $y = z$, then a_i, a_j, a_k are three terms in terms arithmetic progression; the final part of the theorem follows. □

Taking $a_1, \ldots, a_N$ as the sequence of the first N integers, we immediately obtain the following corollary:

Corollary 5.5.4 (Schur). *Suppose the integers $1, 2, \ldots, \lfloor k!e \rfloor$ are k-colored. Then there exists a monochromatic solution to $x - y = z$, with $1 \leq x, y, z \leq \lfloor k!e \rfloor$.*

Now choose $a_i = 2^{i-1}$ for $1 \leq i \leq \lfloor k!e \rfloor + 1$. It is easy to check that a_i contain no three terms in arithmetic progression, so that we obtain:

Corollary 5.5.5. *If the integers $1, 2, \ldots, 2^{\lfloor k!e \rfloor} - 1$ are k-colored there exists a monochromatic solution to $x - y = z$, with $1 \leq x, y, z \leq 2^{\lfloor k!e \rfloor} - 1$ and x, y, z pairwise distinct.*

It is natural to ask to what extent these bounds can be improved. Let $N = N_k$ denote the smallest N for which any k-coloring of $\{1, 2, \ldots, N\}$ admits a monochromatic solution to $x - y = z$. Then Wan [Wan97] has shown

$$N < k!\frac{e - e^{-1} + 3}{2} \quad (k \geq 4), \qquad N < k!\frac{e - e^{-1} + 3}{2} - n + 2 \quad (\text{even } k \geq 6).$$

In the opposite direction, the original paper of Schur contains a proof that it is possible to k-color the integers $1, 2, \ldots, M_k := (3^k - 1)/2$ in such a way that there is no monochromatic solution to $x - y = z$. The proof begins with the observation that if $\mathcal{C}_1, \ldots, \mathcal{C}_k$ is a partition of $\{1, 2, \ldots, M\}$ into k sets, none of which contains a solution to $x - y = z$, then

$$\mathcal{C}_i' := (3 \times C_i) \cup ((3 \times \mathcal{C}_i) - 1), \quad \mathcal{C}_{k+1}' := \{1, 4, 7, \ldots, 3M + 1\}$$

is a partition of $\{1, 2, \ldots, 3M + 1\}$ into $k + 1$ sets with the same property.

For example, from the trivial partition of $\{1\}$ into a single (difference-free) set, we obtain a difference-free partition of $\{1, 2, 3, 4\}$ into $\{1, 4\}$ and $\{2, 3\}$. The quoted bound of Schur, $N_k > (3^k - 1)/2$, follows easily from induction and the relation $M_{k+1} = 3M_k + 1$. Fredricksen [Fre79] has shown the lower bound $(3^k - 1)/2$ may be replaced by $c(315)^{k/5}$ for some positive constant c.

The bound of Corollary 5.5.5 stemmed from our particular choice of a sequence free of three-term arithmetic progressions, and a denser sequence would lead to a better bound. In Chapter 6, we will present Behrend's construction of a "dense" arithmetic progression-free sequence. Our result there implies $2^{\lfloor k!e \rfloor} - 1$ may be replaced by $(k!e)^{1+o(1)}$. Bornsztein [Bor02] has recently shown it may be replaced by $\lfloor k!ke \rfloor + 1$.

We now explain the title of this section. Schur's student, R. Rado, took up the task of generalizing Schur's results to other equations

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = 0, \tag{5.32}$$

where the a_i are nonzero integers. He called such an equation *k-fold regular* if for any k-coloring of the positive integers, there is always a monochromatic solution to (5.32). An equation which is k-fold regular for each k was simply called *regular*. In this language, Schur's theorem is a finitary version of the statement that $x_1 + x_2 - x_3$ is regular.

We close by stating without proof the following criterion of Rado [Rad33] for the regularity of (5.32):

Theorem (Rado). *The equation* (5.32) *is regular if and only if there exists a nonempty set* $I \subset \{1, 2, \ldots, n\}$ *such that*

$$\sum_{r \in I} a_r = 0.$$

For a proof of this theorem and for further discussion of Rado's results, see [Adh02]. For a cognate result, see Exercise 6.7.2.

5.5.3 Application to the Fermat Congruence

Consider the (in)famous Fermat equation,

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

A standard way to investigate integer solutions to Diophantine equations is to begin by looking at solutions modulo primes and prime powers. As an example, take $m = 3$, and consider the corresponding Fermat equation $\pmod 7$. The cubes $\pmod 7$ are $0, 1$ and -1, and there is no solution to $\pm 1 + \pm 1 \equiv \pm 1 \pmod 7$ (for any choice of sign). It follows that in any solution x, y, z to the Fermat equation, one of x^3, y^3 or z^3 is $0 \pmod 7$, so that $7 \mid xyz$. If we could show the same held for infinitely many primes in place of 7, it would force $xyz = 0$, and we would have proven Fermat's Last Theorem in the case $m = 3$.

Unfortunately, this game-plan is doomed to failure, for $m = 3$ and any other exponent. For example, Dickson, improving on an earlier estimate of Cornacchia, showed that for $p > (m-1)^2(m-2)^2 + 6m - 3$, the "Fermat congruence"

$$x^m + y^m \equiv z^m \pmod p \tag{5.33}$$

always has a nontrivial solution, i.e., a solution in integers x, y, z with $xyz \not\equiv 0 \pmod{p}$. (Strictly speaking, he proved this only for prime exponents m.) This bound can be deduced, for example, from the theory of Jacobi sums (as developed, e.g., in [IR90, Chapter 8]).

A simpler proof that (5.33) is nontrivially solvable for $p > p_0(m)$ was given by Schur, utilizing his Corollary 5.5.4:

Theorem 5.5.6. *Let m be a positive integer. The congruence* (5.33) *is nontrivially solvable for $p \geq 1 + \lfloor m!e \rfloor$.*

Proof. Since the subgroup of mth powers in $\mathbf{Z}/p\mathbf{Z}^*$ coincides with the subgroup of $(m, p-1)$th powers, it suffices to treat the case when $m \mid p-1$.

In this case, the cosets of this subgroup partition $\mathbf{Z}/m\mathbf{Z}^*$ into m sets. Identifying an integer with the coset it represents, we get an induced m-coloring of $\{1, 2, \ldots, p-1\}$; since $p - 1 \geq \lfloor m!e \rfloor$, it follows from Corollary 5.5.4 that there exists a monochromatic solution to $Z - Y = X$.

Referring to the definition of this coloring, we see there exist integers g, x', y', and z' all coprime to p with $X \equiv gx^m \pmod{p}, Y \equiv gy^m \pmod{p}, Z \equiv gz^m \pmod{p}$. Then x, y, z are a nontrivial solution to (5.33). □

Exercise 5.5.1.

a) Show that there are no consecutive 5th power residues (mod 25). Deduce that there are no solutions to $x^5 + y^5 \equiv z^5 \pmod{25}$ with x, y, z all units (mod 25). Thus $5 \mid xyz$ if $x^5 + y^5 = z^5$.

b) Show that the situation of part a) is atypical: for $k, m \geq 1$, there is always a solution to $x^m + y^m \equiv z^m \pmod{p^k}$ with x, y, z coprime to p provided p is sufficiently large, $p > p_0(k, m)$.

Several other aspects of the Fermat congruence been investigated (such as the p-adic solvability of the Fermat equation). For an elegant survey, see [Rib79, Lecture 12].

References

[Adh02] Sukumar Das Adhikari, *Aspects of combinatorics and combinatorial number theory*, Narosa Publishing House, New Delhi, 2002. MR 1 920 868

[Aig61] Alexander Aigner, *Folgen der Art $ar^n + b$, welche nur teilbare Zahlen liefern.*, Math. Nachr. **23** (1961), 259–264.

[ANR95] Noga Alon, Melvyn B. Nathanson, and Imre Ruzsa, *Adding distinct congruence classes modulo a prime*, Amer. Math. Monthly **102** (1995), no. 3, 250–255. MR 95k:11009

[AT92] N. Alon and M. Tarsi, *Colorings and orientations of graphs*, Combinatorica **12** (1992), no. 2, 125–134. MR 93h:05067

[Bor02] Pierre Bornsztein, *On an extension of a theorem of Schur*, Acta Arith. **101** (2002), no. 4, 395–399. MR 2002k:11028

[Car00] Gabriel D. Carroll, *Combinatorial number theory*, Presentation to the Berkeley Math Circle. Available online from `http://www.people.fas.harvard.edu/~gcarroll/math/`, March 2000.

[Cau13] A.L. Cauchy, *Recherches sur les nombres*, J. École polytech. **9** (1813), 99–116.

[Cho35] Inder Chowla, *A theorem on the addition of residue classes: Application to the number $\Gamma(k)$ in Waring's problem.*, Proc. Indian Acad. Sci., Sect. A **2** (1935), 242–243 (English).

[Cro61] Roger Crocker, *A theorem concerning prime numbers*, Math. Mag. **34** (1960/1961), 316, 344. MR 25 #2994

[Cro71] ———, *On the sum of a prime and of two powers of two*, Pacific J. Math. **36** (1971), 103–107. MR 43 #3200

[Dav35] H. Davenport, *On the addition of residue classes*, J. London Math. Soc. **10** (1935), 30–32.

[DdSH94] J. A. Dias da Silva and Y. O. Hamidoune, *Cyclic spaces for Grassmann derivatives and additive theory*, Bull. London Math. Soc. **26** (1994), no. 2, 140–146. MR 95i:11007

[Dys45] F. J. Dyson, *A theorem on the densities of sets of integers*, J. London Math. Soc. **20** (1945), 8–14. MR 7,365b

[Erd50] P. Erdös, *On integers of the form $2^k + p$ and some related problems*, Summa Brasil. Math. **2** (1950), 113–123. MR 13,437i

[Fre79] Harold Fredricksen, *Schur numbers and the Ramsey numbers $N(3, 3, \cdots, 3; 2)$*, J. Combin. Theory Ser. A **27** (1979), no. 3, 376–377. MR 81d:05001

[HR83] Heini Halberstam and Klaus Friedrich Roth, *Sequences*, second ed., Springer-Verlag, New York, 1983. MR 83m:10094

[IR90] Kenneth Ireland and Michael Rosen, *A classical introduction to modern number theory*, second ed., Graduate Texts in Mathematics, vol. 84, Springer-Verlag, New York, 1990. MR 92e:11001

[Kne53] Martin Kneser, *Abschätzung der asymptotischen Dichte von Summenmengen*, Math. Z. **58** (1953), 459–484. MR 15,104c

[Lan30] E. Landau, *Die Goldbachsche Vermutung und der Schnirelmannsche Satz*, Nachr. Akad. Wiss. Göttingen Math.-Phys. Kl. (1930), 255–276.

[Lep50] Benjamin Lepson, *Certain best possible results in the theory of Schnirelmann density*, Proc. Amer. Math. Soc. **1** (1950), 592–594. MR 12,392h

[LS89] David B. Leep and Daniel B. Shapiro, *Multiplicative subgroups of index three in a field*, Proc. Amer. Math. Soc. **105** (1989), no. 4, 802–807. MR 89m:11127

[Man42] Henry B. Mann, *A proof of the fundamental theorem on the density of sums of sets of positive integers*, Ann. of Math. (2) **43** (1942), 523–527. MR 4,35e

[Mir75] L. Mirsky, *The combinatorics of arbitrary partitions*, Bull. Inst. Math. Appl. **11** (1975), no. 1–2, 6–9. MR 56 #148

[Nar83] W. Narkiewicz, *Number theory*, World Scientific Publishing Co., Singapore, 1983, Translated from the Polish by S. Kanemitsu. MR 85j:11002

[Nat80] Melvyn B. Nathanson, *Sumsets contained in infinite sets of integers*, J. Combin. Theory Ser. A **28** (1980), no. 2, 150–155. MR 81j:10085

[Nat87] ———, *A generalization of the Goldbach-Shnirel′man theorem*, Amer. Math. Monthly **94** (1987), no. 8, 768–771. MR 90b:11098

[Nat96] ———, *Additive number theory*, Graduate Texts in Mathematics, vol. 164, Springer-Verlag, New York, 1996, The classical bases. MR 97e:11004

[Ost56] Hans-Heinrich Ostmann, *Additive Zahlentheorie. Erster Teil: Allgemeine Untersuchungen. Zweiter Teil: Spezielle Zahlenmengen*, Ergebnisse der Mathematik und ihrer Grenzgebiete (N.F.), Hefte 7, vol. 11, Springer-Verlag, Berlin, 1956. MR 20 #5176

[Rad33] Richard Rado, *Studieren zur Kombinatorik*, Math. Zeitschrift **36** (1933), 424–480.

[Rib79] Paulo Ribenboim, *13 lectures on Fermat's last theorem*, Springer-Verlag, New York, 1979. MR 81f:10023

[Rie56] H. Riesel, *Några stora primtal*, Elementa **39** (1956), 258–260.

[Rog64] Kenneth Rogers, *The Schnirelmann density of the squarefree integers*, Proc. Amer. Math. Soc. **15** (1964), 515–516. MR 29 #1192

[Rom34] N.P. Romanov, *Über einige Sätze der additiven Zahlentheorie*, Math. Annalen **109** (1934), 668–678.

[Sch16] I. Schur, *Über die Kongruenz* $x^m + y^m \equiv z^m \pmod{p}$, Jahresber. d. Deutschen Math. Verein. **25** (1916), 114–117.

[Sch33] L.G. Schnirelmann, *Über additive Eigenschaften von Zahlen*, Math. Ann. **107** (1933), 649–690.

[Sie60] W. Sierpiński, *Sur un problème concernant les nombres* $k \cdot 2^n + 1$, Elem. Math. **15** (1960), 73–74. MR 22 #7983

[Sie88] ———, *Elementary theory of numbers*, second ed., North-Holland Mathematical Library, vol. 31, North-Holland Publishing Co., Amsterdam, 1988, Edited and with a preface by Andrzej Schinzel. MR 89f:11003

[Sma77a] Charles Small, *Solution of Waring's problem* mod n, Amer. Math. Monthly **84** (1977), no. 5, 356–359. MR 55 #12671

[Sma77b] ———, *Waring's problem* mod n, Amer. Math. Monthly **84** (1977), no. 1, 12–25. MR 54 #12693

[Stö55] Alfred Stöhr, *Gelöste und ungelöste Fragen über Basen der natürlichen Zahlenreihe. I, II*, J. Reine Angew. Math. **194** (1955), 40–65, 111–140. MR 17,713a

[vdC50] J. G. van der Corput, *On de Polignac's conjecture*, Simon Stevin **27** (1950), 99–105. MR 11,714e

[Vos56] A. G. Vosper, *The critical pairs of subsets of a group of prime order*, J. London Math. Soc. **31** (1956), 200–205. MR 17,1056c

[Wan97] Honghui Wan, *Upper bounds for Ramsey numbers* $R(3, 3, \cdots, 3)$ *and Schur numbers*, J. Graph Theory **26** (1997), no. 3, 119–122. MR 98f:05114

Chapter 6

Sequences Without Arithmetic Progressions

6.1 Introduction

We begin with a result (apparently) first conjectured by Schur [Soi95] which later became a celebrated theorem of van der Waerden:

Theorem 6.1.1 (van der Waerden [vdW27], 1927). *If the integers are partitioned into finitely many (possibly empty) classes, then one of these classes contains arbitrarily long arithmetic progressions.*

One might guess that the reason van der Waerden's theorem holds is that in any finite partition of the positive integers, at least one set in the partition has to be "large," in the sense of possessing positive upper density. To see this, observe that if all of the sets had asymptotic density 0, then their union would also have density 0. This line of thinking led Erdős and Turán to formulate the following conjecture:

Conjecture 6.1.1 (Erdős & Turán [ET36], 1936). *Let S be a subset of the positive integers possessing positive upper density. Then S contains arbitrarily long arithmetic progressions.*

This conjecture implies van der Waerden's theorem; indeed, it says the "largest" subset in any finite partition of the positive integers always contains arbitrarily long arithmetic progressions.

The first progress on the Erdős-Turán conjecture was made by K.F. Roth, who showed that a set with positive upper density has to contain a three-term arithmetic progression. Actually, Roth proved something stronger:

Theorem 6.1.2 (Roth [Rot52], 1952). *If $r_3(n)$ denotes the maximal size of a subset of $\{1, 2, \ldots, n\}$ containing no 3 terms in arithmetic progression, then $r_3(n)/n \to 0$ as $n \to \infty$.*

To see how this implies the original statement, observe that if no 3 terms of $\mathcal{A}$ are in arithmetic progression, then $|\mathcal{A} \cap \{1, 2, \ldots, n\}| \leq r_3(n)$ for each n.

About fifteen years later, Szemerédi proved ([Sze90], [Sze70]) that, with $r_4(n)$ defined analogously, $r_4(n)/n \to 0$. Thus a set with positive upper density contains a four-term arithmetic progression. Finally, in 1975, Szemerédi [Sze75] succeeded in proving Conjecture 6.1.1 by showing that $r_k(n)/n \to 0$ for every fixed positive integer k.

Roth's result above is the upper estimate $r_3(n) = o(n)$. Stronger, explicit upper estimates are known; the best at the time of writing is due to Bourgain [Bou99]:

$$r_3(n) \ll n\sqrt{\frac{\log\log n}{\log n}},$$

One can also ask for lower estimates on $r_3(n)$. It was noted by Szekeres that the sequence of nonnegative integers containing only 0 and 1 in their ternary expansion is free of any three-term arithmetic progression. It follows that there is an arithmetic-progression-free subset of $\{0, 1, \ldots, (3^k - 1)/2\}$ consisting of 2^k elements; translating everything by 1 shows that

$$r_3\left(\frac{3^k+1}{2}\right) \geq 2^k, \tag{6.1}$$

and sandwiching n between two numbers of the form $(3^k + 1)/2$ then shows (see Exercise 6.1.3)

$$r_3(n) \gg n^{\log 2/\log 3} \qquad (n \to \infty).$$

Szekeres verified that equality holds in (6.1) for $k = 1, 2, 3$, and 4 (see [ET36]) and conjectured this was generally true. If this were the case, then one would also have $r_3(n) = O(n^{\log 2/\log 3})$. However, in 1942 Salem & Spencer [SS42] showed that for every $\epsilon > 0$,

$$r_3(n) > n^{1-\epsilon}, \qquad n > n_0(\epsilon).$$

In fact, they proved a more precise result, with ϵ replaced by a function of n tending slowly to 0. Building upon their work, F.A. Behrend [Beh46] established the best lower bound to date,

$$r_3(n) \geq ne^{-C\sqrt{\log n}}$$

for some constant C and all large n.

The study of sequences without arithmetic progressions is currently a fertile research area with many tantalizing unsolved problems. As a sample, we mention the following conjecture (carrying an Erdős prize of $3000):

Conjecture 6.1.2 (Erdős). *If $\mathcal{A}$ is a set of positive integers for which the sum $\sum_{a\in\mathcal{A}} a^{-1}$ diverges, then $\mathcal{A}$ contains arbitrarily long arithmetic progressions.*

This is wide-open; it is not even known whether the hypothesis of guarantees the existence of a *three-term* arithmetic progression. Note that this conjecture would have a number of very interesting consequences; e.g., in the literature the observation is often made that this would imply the primes contain arbitrarily long arithmetic progressions. Very recently this latter statement has been unconditionally established by Ben Green & Terence Tao [GT04].

In this chapter we present proofs of the theorems of van der Waerden, Roth and Behrend. Our proof of van der Waerden's theorem follows the short proof outlined in Graham and Rothschild [GR74] as expounded in [PS95]. We give two proofs of Roth's theorem. The first is purely combinatorial and is based on on Szemerédi's proof of his more general (and very difficult) result mentioned above. Our treatment follows very closely that of [Gra81, Chapter 5]. The second proof given here employs the *circle method*, and in this form is due to D.J. Newman [New81], [New98]. Our presentation of Behrend's construction is based on his original paper [Beh46]. I first learned this result from Ernie Croot, and his perspective is also in evidence there.

6.1.1 Exercises

Exercise 6.1.1. Show that $r_3(n)$ is a nondecreasing function satisfying $r_3(m+n) \le r_3(m) + r_3(n)$ for every pair of positive integers m, n.

Exercise 6.1.2 *(Erdős & Turán [ET36]; see also [ES03, §6.19]).* Determine $r_3(n)$ for $n \le 8$. Suggestion: When $n < 5$, this is easy. To handle $5 \le n \le 8$, first produce a 4-element A.P. free subset of $\{1, 2, \ldots, 5\}$. Since $r_3(n)$ is nondecreasing (Exercise 6.1.1), showing $r_3(8) \le 4$ will imply $r_3(5) = r_3(6) = r_3(7) = r_3(8) = 4$. To prove this inequality, observe that any 5 element subset of $\{1, 2, \ldots, 8\}$ must contain either three elements not greater than 4 or three elements greater than 4. Show that one can assume the first case always holds (by flipping about 9/2). Now examine the 3-element A.P.-free subsets of $\{1, 2, 3, 4\}$ and show none of them can be extended to a 5-element A.P. free subset of $\{1, 2, \ldots, 8\}$.

Exercise 6.1.3 *(Szekeres).*

a) Show that the sequence of nonnegative integers containing only the digits 0 and 1 in their ternary expansion contains no non-trivial three-term arithmetic progression.

b) Deduce that $r_3((3^k+1)/2) \ge 2^k$. From this, prove $r_3(n) \gg n^{\log 2/\log 3}$. Assuming $r_3((3^k+1)/2) = 2^k$ for every k, show that also $r_3(n) \ll n^{\log 2/\log 3}$.

Exercise 6.1.4 *([Ruz99, §13]).* Let S_l be the set of nonnegative integers containing only the digits 0, 1 and 2 in their base 5 expansion and containing exactly l 1's.

a) Show that S_l contains no nontrivial three-term arithmetic progressions.

b) Using part a), show that if k is a positive integer and $0 \le l \le k$,

$$r_3\left(\frac{5^k+1}{2}\right) \ge \binom{k}{l} 2^{k-l}.$$

c) Taking $l = \lfloor k/3 \rfloor$ in part b), show that there is a constant $C \geq 0$ for which

$$r_3\left(\frac{5^k+1}{2}\right) \geq k^{-C} 3^k$$

whenever $k \geq 2$.

d) Using part c), deduce that

$$r_3(n) \gg n^{\log 3/\log 5 - o(1)}$$

as $n \to \infty$. Thus Szekeres' conjecture cannot hold.

6.2 The Theorem of van der Waerden

6.2.1 Equivalent Forms of van der Waerden's Theorem

> After lunch we [van der Waerden, Schreier, and Artin] went into Artin's office in the Mathematics Department at the University of Hamburg and tried to find a proof. We drew some diagrams on the blackboard. We had what the Germans call 'Einfälle': sudden ideas that flash into one's mind. Several times such new ideas gave the discussion a new turn, and one of the ideas finally led to the solution. – van der Waerden, *How the Proof of Baudet's Conjecture was Found* [vdW71]

We begin with some reformulations of van der Waerden's theorem due to Emil Artin and Otto Schreier. In the next sequel, we shall prove van der Waerden's theorem in the third form given below, following a method of Graham and Rothschild.

Theorem 6.2.1 (Artin & Schreier). *The following statements are equivalent:*

i. If the positive integers are partitioned into two subsets, then one of the two contains arbitrarily long arithmetic progressions.

ii. For every positive integer k, there exists a positive integer $W(k)$ such that if the numbers $\{1, 2, \ldots, W(k)]\}$ are partitioned into two subsets, then one of the subsets contains an arithmetic progression of length k.

iii. For every pair of positive integers k, r, there exists a positive integer $W(k, r)$ such that if the numbers $\{1, 2, \ldots, W(k, r)\}$ are partitioned into r subsets, then one of the subsets contains a k-term A.P.

iv. For every positive integer r, if the positive integers are partitioned into r subsets, then one of them contains arbitrarily long arithmetic progressions.

Proof. i) $\Rightarrow$ ii): We employ something called the *compactness principle*. Suppose ii) fails for the positive integer k; then for every positive integer n, there is a partition $\{1, 2, \ldots, n\} = \mathcal{A}_n \cup \mathcal{B}_n$ with neither $\mathcal{A}_n$ nor $\mathcal{B}_n$ containing a k-term arithmetic progression. We may choose an infinite subsequence of the positive integers $N_{11} < N_{12} < \ldots$ such that

$$\mathcal{L}_1 := \mathcal{A}_{N_{11}} \cap \{1\} = \mathcal{A}_{N_{12}} \cap \{1\} = \ldots,$$

since there are only two possibilities for the intersection and there are infinitely many positive integers. Similarly, there is an infinite subsequence $N_{21} < N_{22} < \ldots$ of $\{N_{1i}\}$ with the property that

$$\mathcal{L}_2 := \mathcal{A}_{N_{21}} \cap \{1, 2\} = \mathcal{A}_{N_{22}} \cap \{1, 2\} = \ldots,$$

and we can assume $N_{21} > N_{11}$ We similarly define the sequence $\{N_{kj}\}_{j \geq 1}$ and $\mathcal{L}_k$ for each $k > 2$; as above, we may assume that

$$N_{11} < N_{21} < \ldots,$$

so that in particular $N_{k1} \geq k$ for each k.

Consider now the set $\mathcal{L} := \cup_{k=1}^{\infty} \mathcal{L}_k$. We claim that neither $\mathcal{L}$ nor its complement contains arbitrarily long arithmetic progressions. Indeed, suppose $\mathcal{L}$ contains a progression of length k, all of whose terms do not exceed M. Since

$$\mathcal{L} \cap \{1, 2, \ldots, M\} = \mathcal{A}_{N_{M1}} \cap \{1, 2, \ldots, M\},$$

this implies $\mathcal{A}_{N_{M1}}$ contains a k-term arithmetic progression, contradicting the initial choice of the sets $\mathcal{A}_i$. Similarly, suppose $\mathbf{Z}^+ \setminus \mathcal{L}$ contains a k-term arithmetic progression, all of whose terms do not exceed M. Then $\{1, 2, \ldots, M\} \setminus \mathcal{A}_{N_{M1}}$ contain the same k-term arithmetic progression. Since $N_{M1} \geq M$, in fact $\{1, 2, \ldots, N_{M1}\} \setminus \mathcal{A}_{N_{M1}}$ contains this progression as well, contradicting again the choice of the $\mathcal{A}_i$.

We now show that ii) $\Rightarrow$ iii) by induction on r, the case $r = 1$ being trivial and the case $r = 2$ being what is asserted in ii). Suppose now that $r \geq 3$, and that it is known that $W(k, r')$ exists for all k and all $r' < r$. We claim we may take $W(k, r) = W(W(k, 2), r-1)$. Indeed, let $\mathbf{Z}^+ = \mathcal{A}_1 \cup \mathcal{A}_2 \ldots \mathcal{A}_r$ be any partition of $\{1, 2, \ldots, W(W(k, 2), r-1)\}$. We can think of $\{1, 2, \ldots, W(W(k, 2), r-1)\}$ as partitioned into $r-1$ classes, simply by lumping together everything in $\mathcal{A}_1$ and $\mathcal{A}_2$ and keeping the remaining $r-2$ classes. Then by the definition of $W(W(k, 2), r-1)$, one of these new classes contains an arithmetic progression of length $W(k, 2)$. If it is one of $\mathcal{A}_3, \ldots, \mathcal{A}_r$, we're done, since clearly $W(k, 2) \geq k$. Otherwise, we have an arithmetic progression of $W(k, 2)$ integers all of which fall into either $\mathcal{A}_1$ or $\mathcal{A}_2$. Number (in order) the $W(k, 2)$ terms of this arithmetic progression as $1, 2, \ldots, W(k, 2)$ and think of them as divided into two classes (according as whether they are in $\mathcal{A}_1$ or $\mathcal{A}_2$); by the definition of $W(k, 2)$, there is an arithmetic progression of length k in one of these classes, and this in turn yields an arithmetic progression of length k in our original set. Since the implications iii) $\Rightarrow$ iv) and iv) $\Rightarrow$ i) are trivial, the theorem follows. $\square$

The reader familiar with point-set topology will recognize the construction of $\mathcal{L}$ that arose in proving the implication i) $\Rightarrow$ ii) as a proof of the sequential compactness of the powerset of $\mathbf{Z}^+$ (identified with $\{0,1\}^{\mathbf{Z}^+}$ given the usual product topology). This explains the name "compactness principle."

Henceforth we use the term *van der Waerden's theorem* to refer to any of the four equivalent assertion's above.

Theorem 6.2.2 (Rabung [Rab75], see also [Bro75]). *van der Waerden's theorem is equivalent to the following: Let $\mathcal{A} = \{a_1, a_2, \dots\}$ be an infinite set of positive integers with $a_1 < a_2 < \dots$. If $a_{i+1} - a_i$ is bounded, then $\mathcal{A}$ contains arbitrarily long arithmetic progressions.*

Proof. This result implies van der Waerden's theorem in the first form given above. Indeed, if the positive integers are partitioned into two sets $\mathcal{A}, \mathcal{B}$, then either the difference between consecutive elements of $\mathcal{A}$ remains bounded, in which case the given statement implies $\mathcal{A}$ contains arbitrarily long arithmetic progressions, or there are arbitrarily long blocks of consecutive integers not in $\mathcal{A}$. The latter possibility implies $\mathcal{B}$ contains these arbitrarily long blocks of consecutive integers, which means $\mathcal{B}$ contains arbitrarily long arithmetic progressions.

On the other hand, our statement follows from van der Waerden's theorem in its fourth form above. Let $\mathcal{A} = \{a_1 < a_2 < \dots\}$ be a set with $a_{i+1} - a_i \le r$ for some positive integer r. Then every $n > a_1$ is in one of the sets $\mathcal{A} + i$ for some $1 \le i \le r$; thus if we define

$$\mathcal{A}_i := \mathcal{A} - a_1 + i = \{a - a_1 + i : a \in \mathcal{A}\},$$

then $\mathbf{Z}^+ = \cup_{i=1}^r \mathcal{A}_i$. We would like to apply van der Waerden's theorem in its fourth form to deduce that one of the sets $\mathcal{A}_i$ (and hence $\mathcal{A}$ also) contains arbitrarily long arithmetic progressions, but the sets $\mathcal{A}_i$ may not be disjoint. To get around this, we apply a standard trick: define $\mathcal{B}_1 = \mathcal{A}_1$ and for $1 < i \le r$ define $\mathcal{B}_i = \mathcal{A}_i \setminus \cup_{j=1}^{i-1} \mathcal{B}_j$. Then $\cup_{i=1}^r \mathcal{B}_i = \mathbf{Z}^+$, and the $\mathcal{B}_i$ are disjoint. Now van der Waerden's theorem implies one of the $\mathcal{B}_i$ contains arithmetic progressions of arbitrary length, and since $\mathcal{B}_i \subset \mathcal{A}_i$, the result follows. □

6.2.2 A Proof of van der Waerden's Theorem

We now present a proof of van der Waerden's theorem due to Graham & Rothschild [GR74]. Our exposition follows very closely the treatment by Pomerance & Sárközy [PS95].

Let l, m be positive integers. We say two m-term sequences

$$x_1, \dots, x_m \quad \text{and} \quad x_1', x_2', \dots, x_m' \quad (x_i, x_i' \in \{0, 1, 2, \dots, l\})$$

are *l-equivalent* if they agree up to the last occurrence of l. More precisely, they are l-equivalent if neither sequence contains l or if there is a k, $1 \le k \le m$, with $x_i = x_i'$ for $1 \le i \le k$ and $x_i, x_i' < l$ for $k < i \le m$. (The first possibility can be considered a degenerate case of the second, provided we allow $k = 0$.)

The next two exercises are meant to help the reader internalize this rather complicated definition; only the second is required for the proof of van der Waerden's theorem.

Exercise 6.2.1. Check that l-equivalence is an equivalence relation on m-term sequences drawn from $\{0, 1, \dots, l\}$. What are the equivalence classes when $m = 1$?

Exercise 6.2.2. Show that if $x_1, \dots, x_{m+1}$ and $x'_1, \dots, x'_{m+1}$ are $(m+1)$-term l-equivalent sequences, and $x_{m+1}, x'_{m+1} < l$, then $x_1, \dots, x_m$ and $x'_1, \dots, x'_m$ are l-equivalent.

For every pair of positive integers l, m, we consider the following assertion:

Assertion $S(l, m)$**.** *For every positive integer r, there exists a positive integer $N(l, m, r)$ with the following property: if $\{1, 2, \dots, N(l, m, r)\}$ is partitioned into r parts $\mathcal{A}_1, \dots, \mathcal{A}_r$, then there are positive integers $a, d_1, \dots, d_m$ such that*

$$a + l(d_1 + \cdots + d_m) \le N(l, m, r) \tag{6.2}$$

and such that whenever $x_1, \dots, x_m$ and $x'_1, \dots, x'_m$ are l-equivalent m-term sequences from $\{0, 1, \dots, l\}$, $a + x_1 d_1 + \cdots + x_m d_m$ and $a + x'_1 d_1 + \cdots + x_m d'_m$ lie in the same $\mathcal{A}_j$.

Note that (6.2) ensures that both sums $a + \sum_{i=1}^m x_i d_i$ and $a + \sum_{i=1}^m x'_i d_i$ fall into $\{1, 2, \dots, N(l, m, r)\}$.

In order to elucidate the connection with van der Waerden's theorem, we now unravel the assertion $S(l, 1)$. Two elements $x_i, x'_i \in \{0, \dots, l\}$ (thought of as 1-term sequences) are l-equivalent if and only if either $x_i, x'_i < l$ or $x_i = x'_i$. Thus $S(l, 1)$ asserts (exactly) that for every positive integer r, there is a positive integer $N(l, 1, r)$ with the following property: if $\{1, 2, \dots, N(l, 1, r)\}$ is partitioned into r subsets, then there are positive integers a, d with $a + ld \le N$ and such that $a + 0d, a + d, \dots, a + (l+1)d$ fall into the same $\mathcal{A}_j$. (It also asserts that $a + xd$ and $a + x'd$ fall into the same class when $0 \le x = x' \le l$, but this is automatic.)

We see from the above that $S(l, 1)$ implies the existence of $W(l, r)$ for every r. Thus, van der Waerden's theorem is a corollary of the following result:

Theorem 6.2.3. *For all positive integers l and m, the assertion $S(l, m)$ is a theorem.*

The proof of Theorem 6.2.3 is divided into two parts. We first show that if $S(l, k)$ is a theorem for $k = 1, 2, \dots, m$, then so is $S(l, m+1)$. We then show that if $S(l, m)$ is a theorem for all m, then so is $S(l+1, 1)$. Since $S(1, 1)$ is a theorem (take $N(1, 1, r) = 2$, $a = d = 1$ and observe that x_i and x'_i are 1-equivalent 1-term sequences if and only if $x_i = x'_i$), the result follows by double induction.

Proof that $S(l, k)\ \forall k \le m \Rightarrow S(l, m+1)$. Assume that $S(l, k)$ is a theorem for $k = 1, 2, \dots, m$. For an arbitrary positive integer r, set $M := N(l, m, r)$ and $M' := N(l, 1, r^M)$. We claim we may take $N(l, m+1, r) = MM'$.

Suppose $\{1, 2, \ldots, MM'\}$ is partitioned into r sets $\mathcal{A}_1, \ldots, \mathcal{A}_r$. Let C be the function assigning to $i \in \{1, 2, \ldots, MM'\}$ the number $j \in \{1, 2, \ldots, r\}$ for which $i \in \mathcal{A}_j$. We consider the matrix

$$A := \begin{pmatrix} C(1) & C(2) & \ldots & C(M) \\ C(M+1) & C(M+2) & \ldots & C(2M) \\ \vdots & \vdots & \ddots & \vdots \\ C((M'-1)M+1) & C((M'-1)M+2) & \ldots & C(M'M) \end{pmatrix}.$$

We claim the matrix A contains an l-term arithmetic progression of identical rows. Each row is one of the r^M possible M-term sequences drawn from $\{1, 2, \ldots, r\}$. Now partition $\{1, 2, \ldots, M'\}$ into r^M classes, where i, j are in the same class if and only if the rows i, j are identical. By our earlier discussion of the meaning of $S(l, 1)$ and from the definition $M' = N(l, 1, r^M)$, there is an arithmetic of progression $b + id$, $i = 1, 2, \ldots, l-1$ of identical rows, where

$$b + ld \leq M'. \tag{6.3}$$

We now apply $S(l, m)$ to the $M = N(l, m, r)$ consecutive integers $\{(b-1)M + 1, (b-1)M + 2, \ldots, bM\}$ with the partition into r sets induced by our existing partition of $\{1, 2, \ldots, MM'\}$. (Think of applying $S(l, m)$ to the integers $\{1, 2, \ldots, M\}$ partitioned into r sets as follows: the integer $1 \leq i \leq M$ is in set j if and only if $i + (b-1)M$ is in set j with respect to our given partition.) We obtain the existence of integers $a, d_1, \ldots, d_m$ such that

(i) $a \geq (b-1)M + 1$, $a + l(d_1 + \cdots + d_m) \leq bM$.

(ii) if $x_1, \ldots, x_m$ and $x'_1, \ldots, x'_m$ are l-equivalent m-term sequences with entries from $\{0, \ldots, l\}$, then

$$C(a + \sum_{i=1}^{m} x_i d_i) = C(a + \sum_{i=1}^{m} x'_i d_i).$$

Set $d_{m+1} = dM$. We will prove $S(l, m+1)$ by showing:

(i') $a + l(d_1 + \cdots + d_{m+1}) \leq MM'$,

(ii') if $x_1, \ldots, x_{m+1}$ and $x'_1, \ldots, x'_{m+1}$ are l-equivalent $(m+1)$-term sequences from $\{0, \ldots, l\}$, then $C(a + \sum_{i=1}^{m+1} x_i d_i) = C(a + \sum_{i=1}^{m+1} x'_i d_i)$.

To prove (i'), it is sufficient to note

$$\begin{aligned} a + l(d_1 + \cdots + d_{m+1}) &= a + l(d_1 + \cdots + d_m) + ld_{m+1} \\ &\leq bM + ld_{m+1} = (b + ld)M \leq M'M, \end{aligned}$$

by (6.3).

We now prove (ii'). Let $x_1, \ldots, x_{m+1}$ and $x'_1, \ldots, x'_{m+1}$ be l-equivalent. If $x_{m+1} = l$ or $x'_{m+1} = l$, then $x_i = x'_i$ for each i, and (ii') certainly holds in this case. Thus we may assume $x_{m+1}, x'_{m+1} < l$. Define

$$j := a - (b-1)M + \sum_{i=1}^{m} x_i d_i, \quad j' := a - (b-1)M + \sum_{i=1}^{m} x'_i d_i. \tag{6.4}$$

Then from (i) above, we see $j, j' \in \{1, 2, \ldots, M\}$. We now look at how the j, j'th columns of A intersect the rows $b, b+d, \ldots, b+(l-1)d$. Since these rows are identical, the jth column is constant on these rows, as is the j'th. We now show that the jth column and j'th column agree on row b, so not only are they both constant, but they are both the *same* constant. For this, note that from (6.4),

$$(b-1)M + j = a + \sum_{i=1}^{m} x_i d_i, \quad (b-1)M + j' = a + \sum_{i=1}^{m} x'_i d_i.$$

But $x_1, \ldots, x_m$ and $x'_1, \ldots, x'_m$ are l-equivalent (Exercise 6.2.2), so that

$$C(a + \sum_{i=1}^{m} x_i d_i) = C(a + \sum_{i=1}^{m} x_i d'_i),$$

by (ii). The claim follows.

We now look at the jth and j'th columns on selected rows chosen from $b, b+d, \ldots, b+(l-1)d$. Since $x_m, x'_m < l$, both $b + x_{m+1}d$ and $b + x'_{m+1}d$ are among these rows. Now (6.4) implies the following expressions for the jth and j'th entries of rows $b + x_{m+1}d$, $b + x_{m+1}d'$ respectively:

$$(b + x_{m+1}d - 1)M + j = a + \sum_{i=1}^{m+1} x_i d_i, \quad (b + x'_{m+1}d - 1)M + j' = a + \sum_{i=1}^{m+1} x'_i d_i.$$

Hence $C(a + \sum_{i=1}^{m+1} x_i d_i) = C(a + \sum_{i=1}^{m+1} x'_i d_i)$. This proves (ii), and hence establishes $S(l, m+1)$. $\square$

Proof that $S(l,k)$ $\forall m \Rightarrow S(l+1,1)$. We wish to show the existence of $N(l+1, 1, r)$ for every positive integer r. It is easy to see we may take $N(l+1,1,1) = l+2$ (take $a = d = 1$ and notice the remaining conditions are trivially satisfied since there is only one set in the partition), so suppose that $r \geq 2$. We claim we may take $N(l+1,1,r) = 2N$, where $N := N(l,r,r)$.

Let $\{1, \ldots, 2N\}$ be partitioned into r subsets $\mathcal{A}_1, \ldots, \mathcal{A}_r$. In accordance with the definition of N, choose $a, d_1, \ldots, d_r$ such that $a + l(d_1 + \cdots + d_r) \leq N$ and such that if $x_1, \ldots, x_r$ and $x'_1, \ldots, x'_r$ are l-equivalent r-term sequences from $\{0, \ldots, l\}$, then $a + \sum_{i=1}^{r} x_i d_i$ and $a + \sum_{i=1}^{r} x'_i d_i$ fall in the same $\mathcal{A}_j$.

Since there are $\binom{r+1}{2}$ pairs u, v with $0 \leq u < v \leq r$ and since $\binom{r+1}{2} = (r+1)r/2 > r$ for $r \geq 2$, there exist $0 \leq u < v \leq r$ with

$$a + \sum_{i=1}^{u} l d_i, \quad a + \sum_{i=1}^{v} l d_i \tag{6.5}$$

in the same subset of the partition, say $\mathcal{A}_j$. Define

$$a' := a + \sum_{i=1}^{u} ld_i, \quad d' := \sum_{i=u+1}^{v} d_i.$$

We claim the $(l+1)$-term arithmetic progression $a', a'+d, \ldots a'+ld'$ lies entirely in $\mathcal{A}_j$ and that $a' + (l+1)d' \leq 2N$. This claim will imply $S(l+1, 1)$ holds, by our earlier interpretation.

Note that we already know that both $a', a'+ld'$ lie in $\mathcal{A}_j$ (these are just the two elements of (6.5)). It remains to show that the same holds for $a'+xd'$ when $x \in \{1, 2, \ldots, l-1\}$. But for these x, the r-term sequences

$$\underbrace{l, l, \ldots, l}_{u}, \underbrace{0, \ldots, 0}_{r-u} \quad \text{and} \quad \underbrace{l, \ldots, l}_{u}, \underbrace{x, \ldots, x}_{v-u}, \underbrace{0, \ldots, 0}_{r-v}$$

are l-equivalent. It follows that

$$a + \sum_{i=1}^{u} ld_i = a' \quad \text{and} \quad a + \sum_{i=1}^{u} ld_i + \sum_{i=u+1}^{v} xd_i = a' + xd$$

lie in the same set of the partition; since $a' \in \mathcal{A}_j$, so is $a' + xd'$.

Finally, to verify that $a' + (l+1)d' \leq 2N$, we note that

$$a' + ld' = a + \sum_{i=1}^{v} ld_i \leq a + \sum_{i=1}^{r} ld_i \leq N,$$

so that

$$a + (l+1)d' \leq 2(a + ld') \leq 2N. \qquad \square$$

6.2.3 Exercises

Exercise 6.2.3 *(Rabung [Rab75]).* Using the compactness principle, show that van der Waerden's theorem is equivalent to the following "finite version" of Rabung's result quoted above: For every pair of positive integers k, r, there exists a number $G(k, r)$ such that any set of g integers $\{a_1 < a_2 < \cdots < a_g\}$ with $a_{i+1} - a_i \leq r$ contains a k-term arithmetic progression.

Exercise 6.2.4. Use the second form of van der Waerden's theorem given in Theorem 6.2.1 to prove the following result (noticed by Schur): for every positive integer k, there is always a block of either k consecutive quadratic residues (mod p) or k consecutive quadratic nonresidues (mod p) whenever $p > p_0(k)$ is a sufficiently large prime.

If you are feeling particularly ingenious, prove the theorem of Brauer that for every k and all $p > p_1(k)$, there is always a block of k consecutive quadratic residues. (One can also prove this with residues replaced by nonresidues; for a survey of like results, see [Bra69]. For an analytic approach via character sums, see [Sch76, Theorem 5A]).

6.3 Roth's Theorem and Affine Properties

Recall that we defined $r_3(n)$ as the size of any largest subset of $\{1, 2, \ldots, n\}$ possessing no three terms in (nontrivial) arithmetic progression. In this chapter we will give two proofs of the following now-classic result:

Theorem 6.3.1 (Roth [Rot52]). *$r_3(n) = o(n)$.*

In this section we set the stage for Roth's result by placing it in the general setting of *affine properties*. This level of generality is excessive (but benign) as regards Szemerédi's combinatorial proof; however, it is a convenient framework to have in place when discussing Newman's analytic proof.

A property P applying to finite (possibly empty) sets of integers is said to be an *affine property* it if satisfies the following three conditions:

i. The empty set $\emptyset$ has P.

ii. If $\mathcal{A}$ has property P and $\mathcal{B} \subset \mathcal{A}$, then $\mathcal{B}$ has property P.

iii. For every pair of integers m, n with $m \neq 0$, the set $\mathcal{A}$ has property P if and only if $m \times \mathcal{A} + n := \{ma + n : a \in \mathcal{A}\}$ has property P.

Some Examples of Affine Properties.

i. The property P_0 of being a finite set of integers, with no restriction imposed, is an affine property.

ii. For any fixed (positive integer) k, the property of containing no k (distinct) terms in arithmetic progression is affine.

iii. Let $a_1, a_2, \ldots, a_k$ be $k \geq 3$ integers which sum to 0. The property of containing no solution to the equation $a_1x_1 + \cdots + a_kx_k$ is affine.

   Note that the property of having no 3 terms in arithmetic progression is equivalent to possessing no solution to $x_1 + x_2 = 2x_3$, which is a property of this type. Actually, Newman's proof of Roth's theorem has an analog for all properties in this family (see Exercise 6.7.2).

We define

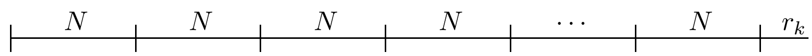
$$M(n, P) := \max_{\substack{\mathcal{A} \subset \{1,2,\ldots,n\} \\ A \text{ possesses } P}} |\mathcal{A}|;$$

i.e., $M(n, P)$ is the size of any largest subset of $\{1, 2, \ldots, n\}$ possessing P. For example, if P is the property of having no 3 terms in arithmetic progression, then $M(n, P)$ is what we have been calling $r_3(n)$. Condition (i) above ensures that $M(n, P)$ is well-defined for every affine property P. When $n = 0$, we adopt the convention that $\{1, 2, \ldots, n\}$ denotes the empty set, so that $M(0, P) = 0$.

Property iii) allows us to make the following observation:

$M(n, P)$ is the size of the largest subset possessing P of *every* set of n consecutive integers, or more generally of every n-term arithmetic progression of integers. (6.6)

Figure 6.1:



This will be very useful in the sequel. For example, in the proof of Roth's theorem the following situation often arises (with P the affine property of having no three terms in arithmetic progression): we have a set $\mathcal{A}$ free of three-term arithmetic progression and a set $\mathcal{J}$, which is a set of consecutive integers or a more general arithmetic progression. Then $\mathcal{A} \cap \mathcal{J}$ is AP-free, so we can estimate $|\mathcal{A} \cap \mathcal{J}|$ from above by $M(|\mathcal{J}|, P) = r_3(|\mathcal{J}|)$.

The *permission constant* associated to P is defined by

$$C_P := \lim_{n \to \infty} \frac{M(n, P)}{n}.$$

For example, $C_{P_0} = 1$, and Theorem 6.3.1 is the assertion that $C_P = 0$ for the affine property of possessing no 3 terms in arithmetic progression. This constant is well-defined for any affine property:

Lemma 6.3.2. *For any affine property P,*

$$\lim_{n \to \infty} \frac{M(n, P)}{n} = \inf_{n=1,2,\ldots} \frac{M(n, P)}{n}.$$

In particular, C_P always exists.

Proof of Lemma 6.3.2. Let c denote this infimum. Because $M(n, P)/n \geq c$ for every n, we need only show that for every $\epsilon > 0$, one has $M(n, P)/n < c + \epsilon$ whenever n is sufficiently large. Suppose otherwise; then we can choose $\epsilon > 0$ for which

$$\frac{M(n_k, P)}{n_k} \geq c + \epsilon \tag{6.7}$$

for a sequence of positive integers $\{n_k\}_{k \geq 1}$ with $n_k \to \infty$ as $k \to \infty$.

From the definition of c, we can choose N for which

$$\frac{M(N, P)}{P} \leq c + \frac{\epsilon}{2}. \tag{6.8}$$

For $k \geq 1$, write

$$n_k = N q_k + r_k,$$

where $0 \leq r_k < N$. We break up the integers in $\{1, 2, \ldots, n_k\}$ into blocks of N consecutive integers, as illustrated in Figure 6.1. If S is any subset of $\{1, 2, \ldots, n_k\}$ possessing P, then by the second defining condition of an affine property, the intersection of each of the blocks of N consecutive integers with

S also possesses P. Thus the size of this intersection is bounded by $M(N,P)$. Consequently,

$$|S| \leq q_k M(N,P) + r_k \leq q_k N\left(c + \frac{\epsilon}{2}\right) + r_k.$$

As this holds for all subsets of $\{1, 2, \ldots, n_k\}$ possessing P, it follows that

$$M(n_k, P) \leq q_k N\left(c + \frac{\epsilon}{2}\right) + r_k \leq n_k\left(c + \frac{\epsilon}{2}\right) + r_k = cn_k + n_k\frac{\epsilon}{2} + r_k. \quad (6.9)$$

On the other hand, by (6.7),

$$M(n_k, P) \geq n_k(c + \epsilon) = cn_k + \epsilon n_k. \quad (6.10)$$

Comparing the estimates (6.9) and (6.10) yields

$$\frac{\epsilon}{2}n_k \leq r_k < N,$$

which is false once k is sufficiently large. □

Naturally, for proving Roth's theorem, we are most concerned with the second family of examples of affine properties given above. Let P_k denote the property of possessing no k terms in arithmetic progression. We abbreviate the associated permission constant, C_{P_k}, by C_k. In this language, Roth's theorem is precisely the assertion that $C_3 = 0$.

In 1938, Behrend [Beh38] showed there were only two possibilities for the behavior of the C_k: either $C_k = 0$ for every k, or $C_k \nearrow 1$ as $k \to \infty$. It took nearly forty years before Szemerédi [Sze75] showed it was the second possibility that prevails.

We close this section with the amusing result that Szemerédi's theorem, ostensibly a statement only about the properties P_k, actually determines the permission constant of any affine property:

Theorem 6.3.3 (D.J. Newman). *We have*

$C_k = 0$ *for every* k *(Szemerédi)* $\Longleftrightarrow$ $C_P = 0$ *for every affine property* $P \neq P_0$.

Proof. The backward implication is obvious. For the forward implication, let P be an affine property different from the trivial property P_0.

Choose k for which $\{1, 2, \ldots, k\}$ does not possess P. Such a choice is possible for otherwise, by (ii) and (iii) in the definition of an affine property, every finite set would possess P, contradicting $P \neq P_0$. With this choice made it then follows from (iii) that any set possessing P contains no k terms in arithmetic-progression. Consequently, $M(n,P) \leq M(n, P_k)$, whence

$$C_P = \lim M(n,P)/n \leq \lim M(n,P_k)/n = 0. \quad \square$$

Exercise 6.3.1. Let $f\colon \mathbf{N} \to \mathbf{R}$ be a function satisfying $f(m+n) \leq f(m) + f(n)$ for all natural numbers m and n. Show that

$$\lim_{n\to\infty} \frac{f(n)}{n} = \inf_{n=1,2,\ldots} \frac{f(n)}{n}.$$

Exercise 6.3.2. Prove that $C_k \leq 1 - 1/k$ for each k.

6.4 Szemerédi's Combinatorial Proof of Roth's Theorem

We now present Szemerédi's proof of Roth's theorem. The proof is, for the most part, an expansion of the treatment in R.L. Graham, [Gra81, Chapter 5].

6.4.1 A Combinatorial Lemma

Lemma 6.4.1. *Let $\alpha = 2 + \sqrt{3}$. For $k \geq 0$, let $l_k := \alpha n^{1-1/2^k}$. If $\mathcal{A}$ is a subset of $\{1, 2, \ldots, n\}$ with at least l_k elements, then there exist $a \in \mathcal{A}$ and $x_0, \ldots, x_k > 0$ for which*

$$\mathcal{K}(a; x_0, \ldots, x_k) := \left\{ a + \sum_{i=0}^{k} \epsilon_i x_i : \epsilon_i = 0 \text{ or } 1 \right\} \subset \mathcal{A}.$$

Proof. We proceed by induction on k. Suppose first that $k = 0$. Since $l_0 = \alpha$ we have $|\mathcal{A}| \geq \alpha > 3$ in this case. So we can choose a as the smallest element of $\mathcal{A}$ and x_0 as the positive difference between a and any other element of $\mathcal{A}$.

Suppose the lemma holds for all values less than a certain $k \geq 1$, and that $\mathcal{A} = \{a_1 < \cdots < a_l\} \subset [1, n]$ with $|\mathcal{A}| = l \geq l_k$. We first show that among the $\binom{l_k}{2}$ differences $a_j - a_i$, $i < j$, at least l_{k-1} of them must be equal. But this follows since all the differences lie in $[1, n]$ while

$$\begin{aligned} \binom{l_k}{2} \geq \frac{1}{2} l_k (l_k - 1) \geq \frac{1}{2}(l_k - 1)^2 &= \frac{1}{2}(\alpha n^{1-1/2^k} - 1)^2 \\ &\geq \frac{1}{2}(\alpha - 1)^2 (n^{1-1/2^k})^2 = \alpha n^{2-1/2^{k-1}} = n l_{k-1}. \end{aligned}$$

Therefore we have

$$a_{i_1} - a_{j_1} = a_{i_2} - a_{j_2} = \cdots = a_{i_m} - a_{j_m} = d > 0,$$

for some $m \geq l_{k-1}$. By the induction hypothesis, we can find a and $x_0, \ldots, x_{k-1}$ with

$$\mathcal{K}(a; x_0, \ldots, x_{k-1}) \subset A' := \{a_{j_1}, \ldots, a_{j_m}\}.$$

But then with $x_k = d$, we have

$$\mathcal{K}(a; x_0, \ldots, x_{k-1}, x_k) \subset \mathcal{A},$$

since $a' \in \mathcal{A}'$ implies $a' + d \in \mathcal{A}$. ∎

Corollary 6.4.2. *Let c be a fixed positive constant. Suppose that $\mathcal{A}$ is contained in a set of n consecutive integers $\{m+1, 2, \ldots, m+n\}$ and that $|\mathcal{A}| \geq cn$. If the nonnegative integer k satisfies*

$$k < \frac{\log\log n - \log(\log\alpha - \log c)}{\log 2}, \tag{6.11}$$

then there exist $a \in \mathcal{A}$ and $x_0, \ldots, x_k > 0$ with

$$\mathcal{K}(a; x_0, \ldots, x_k) \subset \mathcal{A}.$$

(Here α is the constant of the preceding lemma.)

Proof. Multiplying both sides of (6.11) of by $\log 2$ and exponentiating, we find

$$2^k < \frac{\log n}{\log \alpha - \log c},$$

so that

$$\frac{1}{2^k} \log n > \log \alpha - \log c,$$

and

$$n^{1/2^k} > \alpha/c.$$

Then

$$|\mathcal{A}| \geq cn > \alpha n^{-1/2^k} n = \alpha n^{1-1/2^k},$$

so in the case $m = 0$ the result follows from the preceding lemma.

In general we obtain from the above integers $a', x_0, \ldots, x_k$, with $x_0, \ldots, x_k > 0$ and

$$\mathcal{K}(a'; c_0, \ldots, x_k) \subset \mathcal{A} - m := \{a - m : a \in \mathcal{A}\};$$

we can now take $x_0, \ldots, x_k$ as above and $a := a' + m$. □

6.4.2 Some Definitions

Let l be a positive integer.

We let $\mathcal{A} = \mathcal{A}(l)$ denote a subset of $\{1, 2, \ldots, l\}$ of size $r_3(l)$ with no three terms in arithmetic progression.

Partition $\{1, 2, \ldots, l\}$ into l' groups $\mathcal{J}_1, \mathcal{J}_2, \ldots, \mathcal{J}_{l'}$ of $\lfloor l^{1/2} \rfloor$ consecutive integers each, where the last group is perhaps incomplete. Note that

$$l' = l^{1/2} + O(1) \qquad (l \to \infty).$$

We assume for the sake of contradiction that that $C_3 > 0$.

Lemma 6.4.3. *It is possible to choose j as a function of l so that both of the following hold:*

i. $|\mathcal{J}_j \cap \mathcal{A}| > C_3 l^{1/2}/2$ *for all large l, and*

ii. $|j - l^{1/2}/2| = o(l^{1/2})$,

where the $o(l^{1/2})$ term describes the behavior as $l \to \infty$.

Let us sketch the proof: By definition of C_3, we have

$$r_3(n) = (C_3 + o(1))n \qquad (n \to \infty). \tag{6.12}$$

Applied with $n = l$, this implies the average size of $|\mathcal{J}_j \cap \mathcal{A}|$ is asymptotic to $C_3 l^{1/2}$. The falsity of the lemma would imply a deficiency in the number of terms of $\mathcal{A}$ in the sets $\mathcal{J}_j$ close to the center, and this would have to be compensated by an overabundance of terms of $\mathcal{A}$ among the outer sets. But (6.12) yields an upper bound on the contribution from any individual set, and this bound implies said compensation is impossible. We now make this precise:

Proof of Lemma 6.4.3. We prove the following, which suffices: Fix $\epsilon > 0$. Then for large l, say $l > l_0(\epsilon)$, there is a j satisfying (i) as well as

(ii') $|j - l^{1/2}| \le 2\epsilon l^{1/2}/C_3$.

To begin the proof, choose $n_0 = n_0(\epsilon)$ with

$$r_3(n) < (C_3 + \epsilon)n \quad (\text{for } n > n_0). \tag{6.13}$$

The length $\lfloor l^{1/2} \rfloor$ of the integer intervals $\mathcal{J}_j$, with $1 \le j < l'$, tends to infinity uniformly with l. So the size of the sets $\mathcal{J}_j$, $1 \le j < l'$, eventually all exceed n_0. Now (6.6) and (6.13) assure us that for large enough l,

$$|\mathcal{J}_j \cap \mathcal{A}| \le r_3(|\mathcal{J}_j|) \le |\mathcal{J}_j|(C_3 + \epsilon) \quad (1 \le j < l').$$

This bound also holds with $j = l'$, unless $|\mathcal{J}_{l'}| \le n_0$, i.e., unless $|\mathcal{J}_{l'}| = O(1)$. (Of course the implied constant here depends on ϵ.)

Write

$$\begin{aligned}|\mathcal{A}| = \left| \dot{\bigcup}_{j=1}^{l'} \mathcal{J}_j \cap \mathcal{A} \right| &= \sum_{j=1}^{l'} |\mathcal{J}_j \cap \mathcal{A}| \\ &= \sum_{|j - l^{1/2}/2| \le 2\epsilon l^{1/2}/C_3} + \sum_{|j - l^{1/2}/2| > 2\epsilon l^{1/2}/C_3} = \sum\nolimits_1 + \sum\nolimits_2,\end{aligned}$$

say. If there are no j for which (i) and (ii') hold simultaneously, then

$$\sum\nolimits_1 \le \left(\frac{C_3 l^{1/2}}{2}\right)\left(\frac{4\epsilon}{C_3} l^{1/2} + O(1)\right) = 2\epsilon l + o(l).$$

Also, by the above discussion,

$$\begin{aligned}\sum\nolimits_2 &\le \left(l - \left(\frac{4\epsilon}{C_3} l^{1/2} + O(1)\right)\left(l^{1/2} + O(1)\right)\right)(C_3 + \epsilon) + O(1) \\ &= l\left(1 - \frac{4\epsilon}{C_3}\right)(C_3 + \epsilon) + O(l^{1/2}) + O(1) \\ &= l\left(C_3 - 4\epsilon + \epsilon - \frac{4\epsilon^2}{C_3}\right) + o(l);\end{aligned}$$

adding, we obtain

$$\begin{aligned} r_3(l) = |\mathcal{A}| &= \sum\nolimits_1 + \sum\nolimits_2 \\ &\leq l\left(2\epsilon + C_3 - 4\epsilon + \epsilon - \frac{4\epsilon^2}{C_3}\right) + o(l) < l(C_3 - \epsilon) + o(l). \end{aligned}$$

But this contradicts (6.12) for all sufficiently large l. $\square$

Now choose j satisfying the conditions of the lemma and set $\mathcal{J} := \mathcal{J}_j$. Define $\mathcal{J}'$ and $\mathcal{J}''$ by

$$\mathcal{J}' := \{x \in \{1, 2, \ldots, l\} : x < \mathcal{J}\}, \quad \mathcal{J}'' := \{x \in \{1, 2, \ldots, l\} : x > \mathcal{J}\},$$

where we write $x < \mathcal{J}$ (resp. $x > \mathcal{J}$) to mean that x lies to the left (resp. right) of the entire integer interval $\mathcal{J}$. That is, $x < \mathcal{J}$ means $x \in \mathcal{J}_i$ for some $i < j$, and correspondingly for $x > \mathcal{J}$.

Condition (ii) of Lemma 6.4.3 permits us to estimate the sizes of J and J'':

Lemma 6.4.4. *We have*

$$|\mathcal{J}'|, |\mathcal{J}''| = l/2 + o(l). \tag{6.14}$$

Proof. Since $|\mathcal{J}| + |\mathcal{J}'| + |\mathcal{J}''| = l$ and $|\mathcal{J}| = O(l^{1/2})$, it suffices to prove the claim for $|\mathcal{J}'|$. We have

$$|\mathcal{J}'| = \lfloor l^{1/2} \rfloor (j-1) = (l^{1/2} + O(1))(j-1) = l^{1/2} j + O(j) + O(1).$$

The lemma follows since $j = l^{1/2}(1/2 + o(1))$. $\square$

Setting

$$k := \lfloor \log\log l \rfloor,$$

we have

$$k < \frac{\log\log l^{1/2} - \log(\log\alpha - \log\frac{C_3}{2})}{\log 2}$$

for large l. So by Corollary 6.4.2 (with $n = \lfloor l^{1/2} \rfloor$ and $c = C_3/2$) there exist a and $x_0, \ldots, x_k$ with

$$\mathcal{K}(a; x_0, \ldots, x_k) \subset \mathcal{J} \cap \mathcal{A},$$

where each x_i satisfies

$$0 < x_i \leq |\mathcal{J}| \leq l^{1/2}.$$

Define

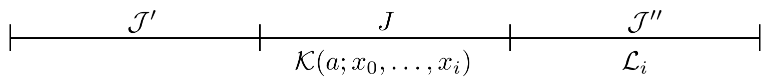
$$\mathcal{K}_i := \mathcal{K}(a; x_0, \ldots, x_i) \quad (-1 \leq i \leq k),$$

where we take $\mathcal{K}_{-1} = \{a\}$. Then take

$$\mathcal{L}_i := \{j'' \in \mathcal{J}'' : j', k_i, j'' \text{ form a three-term AP for some } j' \in A \cap J', k_i \in \mathcal{K}_i\} \quad (-1 \leq i \leq k).$$

Figure 6.2 shows the rough relative positions of $\mathcal{J}, \mathcal{J}', \mathcal{J}''$ as well as $\mathcal{K}_i$ and $\mathcal{L}_i$.

Figure 6.2:



6.4.3 Properties of the L_i

Because $\mathcal{K}_i \subset \mathcal{K}_{i+1}$, we also have $\mathcal{L}_i \subset \mathcal{L}_{i+1}$. The next lemma says that $\mathcal{L}_{i+1}$ is "almost" the sumset

$$\mathcal{L}_i + 2x_{i+1} := \{j'' + 2x_{i+1} : j'' \in \mathcal{L}_i\}.$$

Actually, we prove these two sets are equal apart from "edge effects" (the meaning of this will be explained by the proof below).

Lemma 6.4.5. *For $-1 \le i < k$, both set differences*

$$\mathcal{L}_{i+1} \setminus (\mathcal{L}_i \cup (\mathcal{L}_i + 2x_{i+1})) \qquad \textit{and} \;\; (\mathcal{L}_i \cup (\mathcal{L}_i + 2x_{i+1})) \setminus \mathcal{L}_{i+1}$$

have $O(l^{1/2})$ elements. Consequently,

$$|\mathcal{L}_{i+1}| = |\mathcal{L}_i \cup (\mathcal{L}_i + 2x_{i+1})| + O(l^{1/2}) \quad (-1 \le i \le k).$$

Proof. Suppose x lies in the first set difference above. Since $x \in \mathcal{L}_{i+1}$, there is a three-term progression

$$j', k, x \quad (j' \in \mathcal{J}', k \in \mathcal{K}_{i+1}).$$

Since $x \notin \mathcal{L}_i$, we must also have $k \notin \mathcal{K}_i$, so that $k = k' + x_{i+1}$ for some $k' \in \mathcal{K}_i$. Then $j', k', x - 2x_{i+1}$ is also a three-term arithmetic progression. By the definition of $\mathcal{L}_i$, either

$$x - 2x_{i+1} \in \mathcal{L}_i \, (\Longrightarrow x \in \mathcal{L}_i + 2x_{i+1})$$

or

$$x - 2x_{i+1} \notin \mathcal{J}''.$$

By hypothesis, $x \notin \mathcal{L}_i + 2x_{i+1}$, so the latter must hold. But x is in $\mathcal{L}_{i+1}$, so also in $\mathcal{J}''$, so that this possibility forces x to belong to the first $2x_{i+1}$ elements of $\mathcal{J}''$. Since $2x_{i+1} \le 2\lfloor l^{1/2} \rfloor$, this limits x to a set of at most $2\lfloor l^{1/2} \rfloor = O(l^{1/2})$ elements.

Because the $\mathcal{L}_i$ are increasing, the second set difference is just

$$(\mathcal{L}_i + 2x_{i+1}) \setminus \mathcal{L}_{i+1}. \tag{6.15}$$

Let $x' \in \mathcal{L}_i$ and choose an associated three-term progression

$$j', k, x' \quad (j' \in \mathcal{J}', k \in \mathcal{K}_i).$$

Then $x := x' + 2x_{i+1}$ belongs to the three-term progression

$$j', k + x_{i+1}, x \quad (k + x_{i+1} \in \mathcal{K}_{i+1}, x \in \mathcal{J}'').$$

So as long as $x' + 2x_{i+1}$ belongs to $\mathcal{J}''$, the element $x' + 2x_{i+1}$ belongs to $\mathcal{L}_{i+1}$, and we've proved

$$\text{if } x' \in \mathcal{L}_i, \text{ then either } x' + 2x_{i+1} \text{ either exceeds } l \text{ or lies in } \mathcal{L}_{i+1}. \tag{6.16}$$

It follows that an element x of (6.15) must be given by $x' + 2x_{i+1}$, where

$$l - 2x_{i+1} < x' \leq l.$$

Since $2x_{i+1} \leq 2\lfloor l^{1/2} \rfloor$, this limits x', and hence x, to a set of $O(l^{1/2})$ elements. □

Lemma 6.4.6. *We have*

$$|\mathcal{L}_{-1}| \geq |\mathcal{A} \cap \mathcal{J}'| + o(l).$$

Proof. Observe that $|\mathcal{L}_{-1}|$ counts the number of three-term progressions

$$j', a, j'', \quad (j' \in \mathcal{A} \cap \mathcal{J}', j'' \in \mathcal{J}'').$$

Since a is fixed, the choice of j' determines the common difference, so determines

$$j'' = (a - j') + a = 2a - j'. \tag{6.17}$$

Moreover, any choice of $j' \in \mathcal{A} \cap \mathcal{J}'$ determines an acceptable progression except those for which j'', as defined by (6.17), falls outside of $\mathcal{J}''$. There are thus two exceptional cases to worry about:

First, j'' could fall outside of $\mathcal{J}''$ because $j'' < \mathcal{J}''$. Since $\mathcal{J}$ has length $\lfloor l^{1/2} \rfloor$ and $a \in \mathcal{J}$, this is ruled out once the difference of the arithmetic progression j', a, j'' exceeds the length of $\mathcal{J}$; e.g.,

$$a - j' > l^{1/2}$$

suffices. This last inequality can only fail if (keeping in mind that $j' \in \mathcal{J}'$, so that $j' \leq a \in \mathcal{J}$)

$$a - l^{1/2} \leq j' \leq a.$$

So this case can hold for at most $O(l^{1/2}) = o(l)$ choices of $j' \in \mathcal{J}' \cap \mathcal{A}$.

The other way j'' could fall outside $\mathcal{J}''$ is if $j'' := 2a - j'$ is too big, i.e., $j'' > l$. Then

$$j' < 2a - l. \tag{6.18}$$

Let us estimate the right hand side here. Since $a \in \mathcal{J} = \mathcal{J}_j$, and $j = l^{1/2}(1/2 + o(1))$, we have

$$a \leq l^{1/2}(l^{1/2}/2 + o(l^{1/2})) = l/2 + o(l),$$

whence

$$2a - l \leq o(l).$$

By (6.18), we conclude this case can also hold for at most $o(l)$ values of $j' \in \mathcal{A} \cap \mathcal{J}'$. □

Lemma 6.4.7. *For each i, $-1 \le i \le k$, we have*

$$|\mathcal{L}_i| \ge \frac{lC_3}{2} + o(l) \tag{6.19}$$

Proof. Because the $\mathcal{L}_i$ are increasing, it suffices to obtain this bound for $|\mathcal{L}_{-1}|$. By the previous lemma, we have

$$\begin{aligned}|\mathcal{L}_{-1}| &\ge |\mathcal{A} \cap \mathcal{J}'| + o(l)\\ &= |\mathcal{A}| - |\mathcal{A} \cap J| - |A \cap \mathcal{J}''| + o(l)\\ &\ge (C_3 + o(1))l - (C_3 + o(1))(|J| + |J''|) + o(l).\end{aligned}$$

Here the bounds on the sizes of $\mathcal{A} \cap J$ and $\mathcal{A} \cap J''$ are deduced by appeals to (6.6) and (6.12). To complete the lower-bound estimate for the size of $\mathcal{L}_{-1}$, we note that

$$|\mathcal{J}| + |\mathcal{J}''| = O(l^{1/2}) + |\mathcal{J}''| = l/2 + o(l)$$

by Lemma 6.4.4, so that we finally obtain

$$|\mathcal{L}_i| \ge (C_3 + o(1))l - (C_3 + o(1))\,(l/2 + o(l)) + o(l) = \frac{lC_3}{2} + o(l). \qquad \square$$

6.4.4 Blocks and Gaps

Since

$$\mathcal{L}_{-1} \subset \mathcal{L}_0 \subset \cdots \subset \mathcal{L}_k$$

and $|\mathcal{L}_k| < l$, there is an index i, $-1 \le i < k$, with

$$|\mathcal{L}_{i+1} \setminus \mathcal{L}_i| \le \frac{l}{k+1} = \frac{l}{\lfloor \log\log l \rfloor + 1} < \frac{l}{\log\log l}. \tag{6.20}$$

Choose such an i and keep it fixed for the remainder of the proof.

Consider the set of elements of $\mathcal{J}''$ that lie in an arbitrary fixed residue class (mod $2x_{i+1}$), say

$$\mathcal{C}_j := \{x \in \mathcal{J}'' : x \equiv j \pmod{2x_{i+1}}\}.$$

By a *$\mathcal{C}_j$-block of $\mathcal{L}_i$*, we mean a subset $\mathcal{B} := \{b_1 < \cdots < b_t\}$ of $\mathcal{C}_j$ with the following three properties:

i. $b_{s+1} = b_s + 2x_{i+1}$ for $1 \le s < t$,

ii. $b_s \in \mathcal{L}_i$ for $1 \le s \le t$,

iii. $b_1 - 2x_{i+1} \notin \mathcal{L}_i$, $b_t + 2x_{i+1} \notin \mathcal{L}_i$.

By *a block of* $\mathcal{L}_i$, we mean a set that is a $\mathcal{C}_j$-block of $\mathcal{L}_i$ for some j.

Let $\mathcal{B}$ be a block with largest element b, and consider $b + 2x_{i+1}$. Because b was maximal, this element cannot belong to $\mathcal{L}_i$; on the other hand, (6.16) implies $b + 2x_{i+1}$ does belong to L_{i+1} provided only that it does not exceed l. Since $2x_{i+1} \le 2\lfloor l^{1/2} \rfloor$, we find that to all but $O(l^{1/2})$ blocks we may associate in this manner an element of $\mathcal{L}_{i+1} \setminus \mathcal{L}_i$. These elements are also distinct, since two distinct blocks cannot have the same largest element. So by (6.20), we find the number of blocks does not exceed

$$l/\log\log l + O(l^{1/2}).$$

Just as we divided into blocks the elements of $\mathcal{C}_j$ belonging to $\mathcal{L}_i$, we now divide into *gaps* those not belonging to $\mathcal{L}_i$. By a *$\mathcal{C}_j$-gap*, we mean a set of elements congruent to $j \pmod{2x_{i+1}}$ which either lies between two $\mathcal{C}_j$ blocks, precedes the initial $\mathcal{C}_j$-block or follows the final $\mathcal{C}_j$-block. Also, if there are no $\mathcal{C}_j$-blocks, we consider all of $\mathcal{C}_j$ a $\mathcal{C}_j$-gap. A *gap* is a set that is a $\mathcal{C}_j$-gap for some j.

The total number of gaps is bounded above by

$$\begin{aligned}\sum_{j \pmod{2x_{i+1}}} (\#\mathcal{C}_j\text{-blocks} + 1) &= \left(\frac{l}{\log\log l} + O(l^{1/2})\right) + O(2x_{i+1}) \\ &= \frac{l}{\log\log l} + O(l^{1/2}). \qquad (6.21)\end{aligned}$$

6.4.5 Denouement

We now have all the tools we need to complete the proof of Roth's theorem.

We will obtain a contradiction by showing that $|\mathcal{A} \cap \mathcal{J}'|$ is larger than allowed. More specifically, since $\mathcal{A} \cap \mathcal{J}'$ is AP-free, we must have

$$\begin{aligned}|\mathcal{A} \cap \mathcal{J}'| &\le (1 + o(1))C_3|\mathcal{J}'| \\ &\le (1 + o(1))C_3\left(\frac{l}{2} + o(l)\right) = (1 + o(1))\frac{lC_3}{2}. \qquad (6.22)\end{aligned}$$

On the other hand,

$$\begin{aligned}|\mathcal{A} \cap \mathcal{J}'| &= |\mathcal{A}| - |\mathcal{A} \cap \mathcal{J}''| - |\mathcal{A} \cap \mathcal{J}| \\ &= |\mathcal{A}| - |\mathcal{A} \cap \mathcal{J}''| + o(l) \\ &= l(C_3 + o(1)) - |\mathcal{A} \cap \mathcal{J}''| + o(l). \qquad (6.23)\end{aligned}$$

So a lower bound on $|\mathcal{A} \cap \mathcal{J}'|$ follows from an upper bound on $|\mathcal{A} \cap \mathcal{J}''|$. We certainly have the upper estimate

$$|\mathcal{A} \cap \mathcal{J}''| \le r_3(|\mathcal{J}''|) \le (C_3 + o(1))|\mathcal{J}''| = l\frac{C_3}{2} + o(l), \qquad (6.24)$$

using (6.14). If we substitute this into (6.23), we find

$$|\mathcal{A} \cap \mathcal{J}'| \ge (1 + o(1))\frac{lC_3}{2},$$

which is just a shade off from contradicting (6.22). In fact, any diminishment of the coefficient of l in the upper estimate (6.24) would give a lower bound contradicting (6.22). Below we will prove

$$|\mathcal{A} \cap \mathcal{J}''| \leq \frac{l}{2} C_3(1 - C_3) + o(l) \tag{6.25}$$

and thereby complete this demonstration of Roth's theorem.

The improvement on (6.24) is based on the observation that $\mathcal{A}$ being free of arithmetic progressions precludes it from intersecting $\mathcal{L}_i$, by the very way we defined $\mathcal{L}_i$. Consequently

$$|\mathcal{A} \cap \mathcal{J}''| = |\mathcal{A} \cap (\mathcal{J}'' \setminus \mathcal{L}_i)|.$$

We would like to estimate this last intersection from above by $(C_3+o(1))|\mathcal{J}''\setminus\mathcal{L}_i|$. However, it is not obvious how to justify this: $\mathcal{J}'' \setminus \mathcal{L}_i$ is not an arithmetic progression, so (6.6) is not immediately applicable.

But we are prepared! Above we partitioned $\mathcal{J}'' \setminus \mathcal{L}_i$ into so-called gaps. Our gaps *are* arithmetic progressions (with common difference $2x_{i+1}$) so (6.6) applies to them. We may therefore proceed by estimating the number of terms of $\mathcal{A}$ in each gap.

Say that a gap $\mathcal{G}$ is *small* if $|\mathcal{G}| < \log\log\log l$, otherwise *large*. For a large gap we have the expected estimate

$$|\mathcal{A} \cap \mathcal{G}| \leq r_3(|\mathcal{G}|) \leq (1 + o(1))|\mathcal{G}|;$$

this is because $|\mathcal{G}|$ is bounded below by a function tending to infinity with l, namely $\log\log\log l$. Moreover, the total number of elements (in $\mathcal{A}$ or otherwise) which belong to small gaps does not exceed

$$\left(\frac{l}{\log\log l} + O(l^{1/2})\right) \log\log\log l = o(l),$$

using (6.21). Putting this together, we find (using (6.14), (6.19))

$$\begin{aligned}
|\mathcal{A} \cap \mathcal{J}''| &= |\mathcal{A} \cap (\mathcal{J}'' \setminus \mathcal{L}_i)| \\
&\leq (C_3 + o(1))|\mathcal{J}'' \setminus \mathcal{L}_i| + o(l) \\
&= (C_3 + o(1))\left(|\mathcal{J}''| - |\mathcal{L}_i|\right) + o(l) \\
&\leq (C_3 + o(1))\left(\frac{l}{2} + o(l) - \frac{C_3 l}{2} + o(l)\right) + o(l) \\
&= \frac{l}{2} C_3(1 - C_3) + o(l).
\end{aligned}$$

6.5 More on Affine Properties

6.5.1 The Behavior of the Extremal Sets

Fix an affine property P. The underlying reason Roth's theorem holds is that the subsets of $\{1, 2, \ldots, n\}$ of size $M(n, P)$ which possess P (the *extremal sets*) behave like subsets of $\{1, 2, \ldots, n\}$ whose elements are chosen at random with probability C_P, at least with respect to containing three-term arithmetic progressions. Let us attempt to justify this here.

We begin with a proposition which is not needed for the proof of Roth's theorem (although that proof will require Lemma 6.5.2 below):

Proposition 6.5.1. *Let $f\colon \mathbf{Z}^+ \to \mathbf{N}$ be a nondecreasing function satisfying $0 \le f(n) \le n$ for every positive integer n. Suppose that $f(n)/n \to \varepsilon$ as $n \to \infty$. Let $E(n)$ denote the expected number of three-term arithmetic progressions in a randomly chosen subset of $\{1, 2, \ldots, n\}$ of size $f(n)$, where order counts and equality is allowed. Then as $n \to \infty$,*

$$E(n) = \varepsilon^3 \frac{n^2}{2} + o(n^2). \tag{6.26}$$

Lemma 6.5.2. *There are precisely $\lceil n^2/2 \rceil$ three-term arithmetic progressions consisting of elements of $\{1, 2, \ldots, n\}$. Here order counts and equality is allowed.*

Proof. The elements x, y, z are in arithmetic progression precisely when $z - y = y - x$, i.e., when $x + z = 2y$. It follows that the number of three-term arithmetic progressions of $\{1, 2, \ldots, n\}$ (counting degenerate progressions where all three terms are equal and taking order into account) is given by the number of ordered pairs (x, z) with $x, z \in \{1, 2, \ldots, n\}$ and $x \equiv z \pmod 2$. We can compute this by calculating the total number of pairs (x, z) and subtracting the number with x, z of opposite parity. If n is even, this comes out to

$$n^2 - 2(n/2)(n/2) = n^2 - \frac{n^2}{2} = \frac{n^2}{2} = \left\lceil \frac{n^2}{2} \right\rceil,$$

while if n is odd, it evaluates as

$$n^2 - 2\frac{n-1}{2}\frac{n+1}{2} = n^2 - \frac{n^2-1}{2} = \frac{n^2+1}{2} = \left\lceil \frac{n^2}{2} \right\rceil. \qquad \square$$

Proof of Proposition 6.5.1. If $f(n) < 3$ for all n, then $\varepsilon = 0$ and $E_n = 0$ for each n. Thus (6.26) asserts that $0 = o(n^2)$, which is certainly the case. Thus we can suppose that $f(n) \ge 3$ for large enough n.

Let B denote the set of 3-element arithmetic progressions of $\{1, 2, \ldots, n\}$, where order counts and trivial progressions are allowed; i.e.,

$$B := \{(x, y, z) : x, y, z \in \{1, 2, \ldots, n\}, x + y = 2z\}.$$

Then

$$E(n) = \frac{1}{\binom{n}{f(n)}} \sum_{\substack{S \subset \{1,2,\ldots,n\} \\ |S| = f(n)}} \sum_{\substack{b = (x,y,z) \in B \\ \{x,y,z\} \subset S}} 1.$$

We now reverse the order of summation. For n large enough that $f(n) \ge 3$, the number of $f(n)$-element subsets of $\{1, 2, \ldots, n\}$ containing a given $b = (x, y, z) \in$

B is $\binom{n-3}{f(n)-3}$. Thus

$$\begin{aligned}
E(n) &= \frac{1}{\binom{n}{f(n)}} \sum_{b=(x,y,z)\in B} \sum_{\substack{S\subset\{1,2,\ldots,n\}\\ |S|=f(n),\{x,y,z\}\subset S}} 1 \\
&= \frac{1}{\binom{n}{f(n)}} \sum_{b=(x,y,z)\in B} \binom{n-3}{f(n)-3} \\
&= \frac{\binom{n-3}{f(n)-3}}{\binom{n}{f(n)}} |B| = \frac{\binom{n-3}{f(n)-3}}{\binom{n}{f(n)}} \left(\frac{n^2}{2} + O(1)\right).
\end{aligned}$$

Now

$$\frac{\binom{n-3}{f(n)-3}}{\binom{n}{f(n)}} = \frac{f(n)/n}{1} \frac{f(n)/n - 1/n}{1-1/n} \frac{f(n)/n-2/n}{1-2/n} \to \varepsilon^3$$

as $n \to \infty$. Consequently,

$$E(n) = (\varepsilon + o(1))^3 \left(\frac{n^2}{2} + O(1)\right) = \varepsilon^3 \frac{n^2}{2} + o(n^2). \qquad \square$$

In §6.7 we will prove the following lemma:

Lemma 6.5.3. *Fix an affine property P, and for every positive integer n fix a subset $\mathcal{M}(n,P) \subset \{1,2,\ldots,n\}$ of size $M(n,P)$ possessing property P. Let $F(n)$ denote the number of three-term arithmetic progressions of $\mathcal{M}(n,P)$. Then as $n \to \infty$,*

$$F(n) = C_P^3 \frac{n^2}{2} + o(n^2).$$

Since $M(n,P)$ is nondecreasing in n for fixed P and $M(n,P)/n \to C_P$, this lemma paired with the result of Proposition 6.5.1 implies that the sets $\mathcal{M}(n,P)$ behave like randomly chosen subsets of $\{1,2,\ldots,N\}$ of size $M(n,P)$.

6.5.2 Roth's Theorem revisited

Let us see how to deduce Roth's theorem from Lemma 6.5.3:

Proof of Roth's Theorem (Theorem 6.3.1). Let P be the affine property of containing no nontrivial three-term arithmetic progression. For each n, fix a choice of an extremal set $\mathcal{M}(n,P)$, as in Lemma 6.5.3. The only three-term arithmetic progressions of $\mathcal{M}(n,P)$ are the trivial ones consisting of a single element repeated three times; thus

$$n \geq M(n,P) = F(n) = C_P^3 \frac{n^2}{2} + o(n^2).$$

It follows that $C_P = 0$, which is Roth's theorem. $\square$

The proof of Lemma 6.5.3 is based on on the following fundamental lemma:

Fundamental Lemma. *Fix an affine property P, and for each integer n fix a subset $\mathcal{M}(n,P) \subset \{1,2,\ldots,n\}$ of size $M(n,P)$ possessing property P. As $n \to \infty$, we have*

$$\sum_{a\in M(n,P)} e(a\theta) = C_P \sum_{a=1}^{n} e(a\theta) + o(n), \tag{6.27}$$

uniformly for $\theta \in [0,1]$.

The proof of the fundamental lemma and the succeeding deduction of Lemma 6.5.3 constitute an example of the *circle method*. We take some time out right now to discuss the basis for this method before returning to these proofs.

Exercise 6.5.1. The Fundamental Lemma can also be interpreted as asserting that the extremal sets display "average" behavior among subsets of the same size. Demonstrate this by showing the following:

Let $f\colon \mathbf{Z}^+ \to \mathbf{N}$ be a nondecreasing function satisfying $0 \le f(n) \le n$ for every positive integer n. Suppose that $f(n)/n \to \varepsilon$ as $n \to \infty$. Let $E(\theta,n)$ denote the expected value of the sum

$$\sum_{a\in\mathcal{A}} e(a\theta)$$

for a randomly chosen subset $\mathcal{A} \subset \{1,2,\ldots,n\}$ of size $f(n)$. Show that

$$E(\theta,n) = \varepsilon \sum_{a=1}^{n} e(a\theta) + o(n) \qquad (n\to\infty),$$

uniformly for $\theta \in [0,1]$.

6.6 Interlude: Remarks on the Circle Method

The proof of the fundamental lemma will be accomplished by a variant on a technique known as the *circle method* (aka the *Hardy-Littlewood method*). This method, based on analysis of exponential sums, has proven to be a powerful method in analytic number theory.

6.6.1 The Function $e(\theta)$

For real θ, we define

$$e(\theta) := e^{2\pi i\theta}.$$

Since $e^{2\pi i} = 1$, it follows that $e(\theta)$ is periodic mod 1. Moreover, as θ runs from 0 to 1, the values assumed by $e(\theta)$ trace out the unit circle $\{z : |z| = 1\}$. This is the essential reason for the term *circle method*.

6.6.2 Parseval's Formula

A *trigonometric polynomial* is a function $f\colon \mathbf{R} \to \mathbf{C}$ given by

$$f(\alpha) = \sum_n a_n e(n\alpha),$$

where $a_n = 0$ for all but finitely many integers n.

Theorem 6.6.1 (An Orthogonality Relation). *Let $m, n \in \mathbf{Z}$. Then*

$$\int_0^1 e(m\alpha)\overline{e(n\alpha)}\, d\alpha = \begin{cases} 1 & \text{if } m = n, \\ 0 & \text{otherwise.} \end{cases}$$

Proof. If $n = m$, then the integrand is identically 1 and the result is clear. Otherwise, the integrand has the form $e(h\alpha)$ where $h \neq 0$. Then

$$\begin{aligned} \int_0^1 e(h\alpha)\, d\alpha &= \int_0^1 \sin(2\pi h\alpha) + i\cos(2\pi h\alpha)\, d\alpha \\ &= \frac{-1}{2\pi h}\cos(2\pi h\alpha)\Big|_0^1 + \frac{i}{2\pi h}\sin(2\pi h\alpha)\Big|_0^1 = 0, \end{aligned}$$

since both sin and cos are periodic with period 2π. □

As an application of the orthogonality relations, we prove an important formula due to Parseval.

Theorem 6.6.2 (Parseval's Formula). *Let*

$$f(\theta) = \sum_n a_n e(n\theta), \quad g(\theta) = \sum_m b_m e(m\theta)$$

be trigonometric polynomials. Then

$$\int_0^1 f(\theta)\overline{g(\theta)}\, d\theta = \sum_n a_n \overline{b_n}.$$

In particular,

$$\int_0^1 |f(\theta)|^2\, d\theta = \sum_n |a_n|^2. \tag{6.28}$$

Proof. Observe

$$\begin{aligned} \int_0^1 f(\theta)\overline{g(\theta)}\, d\theta &= \int_0^1 \sum_{n,m} a_n \overline{b_m} e(n\theta)\overline{e(m\theta)}\, d\theta \\ &= \sum_{n,m} a_n \overline{b_m} \int_0^1 e(n\theta)\overline{e(m\theta)}\, d\theta = \sum_n a_n \overline{b_n}, \end{aligned}$$

where the final equality follows from the above orthogonality relation. □

6.6.3 Applications

Let $\mathcal{A}$ is a finite set of integers, and consider the trigonometric polynomial defined by

$$f(\theta) = \sum_{a \in \mathcal{A}} e(a\theta).$$

Then

$$f(\theta)^s = \sum_{a_1, a_2, \ldots, a_s \in \mathcal{A}} e((a_1 + a_2 + \cdots + a_s)\theta),$$

so that our orthogonality relation implies

$$R_s(n) = \int_0^1 f(\theta)^s e(-\theta n)\, d\theta,$$

where $R_s(n)$ counts the number of solutions to

$$n = a_1 + a_2 + \cdots + a_s.$$

For instance, we might take $\mathcal{A}$ as the set of (nonnegative) kth powers not exceeding N; then $R_s(n)$ is the number of ways of writing any $n \leq N$ as a sum of kth powers. Similarly, we might take $\mathcal{A}$ as the set of primes not exceeding N; then $R_s(n)$ is the number of (ordered) representations of any $n \leq N$ as a sum of s primes.

Because $R_s(n)$ can be represented as an integral, it is reasonable to hope that estimates for $R_s(n)$ could be obtained by studying the integrand. In many situations this is indeed the case; e.g., this is the modern method of attacking Waring's problem. It is also how Vinogradov showed that every sufficiently large odd integer is a sum of three primes.

More in line with our immediate purposes, we can use the circle method to count three-term arithmetic progressions from a finite set. If $\mathcal{A}$ is a finite set of integers, then the number of three-term arithmetic progressions from $\mathcal{A}$ (counting order and allowing equality) is just the number of solutions to $x + y = 2z$, and this is given by

$$\int_0^1 \left(\sum_{a \in \mathcal{A}} e(a\theta)^2 \right) \left(\sum_{a \in \mathcal{A}} e(-2a\theta) \right) d\theta.$$

In the next section, we shall prove the fundamental lemma by obtaining analytic information on the exponential sum $\sum_{a \in \mathcal{A}} e(a\theta)$.

6.6.4 Exercises

Exercise 6.6.1. Let $f(\theta) = (1 + e(\theta))^n$. Using the binomial theorem and Parseval's formula, show that

$$\sum_{k=0}^{n} \binom{n}{k}^2 = \int_0^1 |f(\theta)|^2\, d\theta.$$

Evaluate the integral and obtain the identity

$$\sum_{k=0}^{n} \binom{n}{k}^2 = \binom{2n}{n}.$$

Exercise 6.6.2. Show that Fermat's Last Theorem is equivalent to the assertion that for every positive integer N,

$$\int_0^1 \left(\sum_{n=1}^{N} e(n^k \theta) \right)^2 \left(\sum_{n=1}^{N} e(-n^k \theta) \right) d\theta = 0.$$

Exercise 6.6.3. Show that for any positive integer q, we have

$$\frac{1}{q} \sum_{j=1}^{q} e(mj/q)\overline{e(nj/q)} = \begin{cases} 1 & \text{if } m \equiv n \pmod q, \\ 0 & \text{otherwise.} \end{cases}$$

This is a discrete version of the orthogonality relation proved in this section; indeed, the left hand side is a Riemann sum for $\int_0^1 e(mt)\overline{e(nt)}\, dt$, and we recover the continuous version by letting q tend to infinity.

Exercise 6.6.4. Let V be the $\mathbf{C}$-vector space of continuous functions $f\colon [0,1] \to \mathbf{C}$ satisfying $f(0) = f(1)$. For $f, g \in V$, define

$$(f, g) := \int_0^1 f(\theta)\overline{g(\theta)}\, d\theta.$$

Show that (f, g) is a scalar product on V. Using the orthogonality relation proved in this section, show that the functions $\{e_m\}_{m \in \mathbf{Z}}$, where $e_m(\theta) := e^{2\pi i m\theta}$, form an orthonormal set. This explains the term "orthogonality relation."

Exercise 6.6.5 *(Bessel's Inequality).* Suppose f is a trigonometric polynomial, and let a_k (for $k \in \mathbf{Z}$) be defined by

$$a_k := \int_0^1 f(\theta) e(-k\theta)\, d\theta \tag{6.29}$$

By the orthogonality relations, a_k is the coefficient of $e(k\alpha)$ in $f(\alpha)$, so that Parseval implies (cf. (6.28)) $\int_0^1 |f|^2 = \sum_k |a_k|^2$.

Now suppose f is *any* piecewise continuous function on $[0, 1]$, and that a_k is defined by (6.29). Show that

$$\sum_k |a_k|^2 \le \int_0^1 |f|^2.$$

Hint: Reexpress the right hand side of the inequality

$$0 \le \int_0^1 \left| f(\theta) - \sum_{k=-N}^{N} a_k e(k\theta) \right|^2 d\theta.$$

6.7 Newman's Proof of Roth's Theorem

6.7.1 Further Preliminaries

We now fix an affine property P, and write C for C_P, $M(n)$ for $M(n,P)$ and $\mathcal{M}(n)$ for $\mathcal{M}(n,P)$. We extend $M(n)$ to be a function defined for all nonnegative real numbers by setting

$$M(x) = M(\lceil x \rceil).$$

Notice that because $M(0) = 0$ and $C = \inf_{n=1,2,\ldots} M(n)/n$, we always have

$$M(x) \geq C\lceil x \rceil \geq Cx.$$

Also, as $x \to \infty$,

$$M(x) = C\lceil x \rceil + o(\lceil x \rceil) = Cx + O(1) + o(\lceil x \rceil) = Cx + o(x);$$

that is, $M(x) - Cx = o(x)$.

For the proof, it will be useful to replace $M(x) - Cx$ with a monotone function majorizing it:

Lemma 6.7.1. *For $x > 0$, define*

$$R(x) := \sup_{0 \leq t \leq x} (M(t) - Ct).$$

Then $R(x)$ is a nondecreasing function satisfying $R(x) = o(x)$ as $x \to \infty$.

Proof. That $R(x)$ is nondecreasing is clear. Now let $\epsilon > 0$ be given. Choose a positive real number x_0 such that

$$M(x) - Cx < \epsilon x$$

whenever $x > x_0$. Now choose a positive real number x_1 such that $R(x_0)/x_1 < \epsilon$, and let $x_2 = \max\{x_0, x_1\}$. Now suppose $x > x_2$. If $0 \leq t \leq x_0$, then

$$M(t) - Ct \leq R(t) \leq R(x_0) < \epsilon x_1 < \epsilon x,$$

while if $x_0 < t \leq x$,

$$M(t) - Ct < \epsilon t \leq \epsilon x.$$

It follows that $(M(t) - Ct) \leq \epsilon x$ for all $0 \leq t \leq x$, so that $R(x) < \epsilon x$ when $x > x_2$. Since $\epsilon > 0$ was arbitrary and $R(x)$ is nonnegative, it follows that $R(x) = o(x)$. ∎

We need two more lemmas.

Lemma 6.7.2 (Dirichlet's Approximation Lemma). *Let $\alpha \in \mathbf{R}$. For every real $Q \geq 1$, there exists a rational number p/q with $1 \leq q \leq Q$, $\gcd(p,q) = 1$ and*

$$|p/q - \alpha| \leq \frac{1}{qQ}.$$

Proof. We give a proof utilizing the elementary theory of Farey fractions (see, e.g., [LeV96, §9.1]); a proof using nothing more than Dirichlet's box principle is outlined in Exercise 6.7.1. It suffices to prove the result when $0 \le \alpha < 1$.

Set $N = \lfloor Q \rfloor$. The number α lies between (inclusively) two terms of the Farey sequence $\mathfrak{F}_N$, say $p_1/q_1 \le \alpha \le p_2/q_2$. Consider the mediant $(p_1+p_2)/(q_1+q_2)$; because this lies between $p_1/q_1, p_2/q_2$ and does not appear in $\mathfrak{F}_N$, we must have $q_1 + q_2 \ge N + 1$. Now α lies in (at least) one of the intervals $[p_1/q_1, (p_1 + p_2)/(q_1+q_2)], [(p_1+p_2)/(q_1+q_2), p_2/q_2]$. If it lies in the first then,

$$\left|\alpha - \frac{p_1}{q_1}\right| \le \frac{p_1+p_2}{q_1+q_2} - \frac{p_1}{q_1} = \frac{1}{q_1(q_1+q_2)} \le \frac{1}{q_1(N+1)} \le \frac{1}{qQ},$$

and we may take $p = p_1, q = q_1$. Similarly, if it lies in the second, then

$$\left|\alpha - \frac{p_2}{q_2}\right| \le \frac{p_2}{q_2} - \frac{p_1+p_2}{q_1+q_2} = \frac{1}{q_2(q_1+q_2)} \le \frac{1}{q_2(N+1)} \le \frac{1}{qQ},$$

and we may take $p = p_2, q = q_2$. □

Lemma 6.7.3. *Let F be a trigonometric polynomial of the form*

$$F(\theta) = \sum_{k=0}^{n} c_k e(k\theta).$$

For $j = 1, \ldots, n+1$, define

$$F_j(\theta) := \sum_{k=n-j+1}^{n} c_k e(k\theta).$$

That is, F_j is the jth partial sum, beginning from the tail. For any $\alpha, \beta \in \mathbf{R}$, we have

$$|F(\alpha) - F(\beta)| \le 2\pi|\alpha - \beta| \sum_{j=1}^{n} |F_j(\beta)|.$$

Proof. If $e(\alpha) = e(\beta)$ then $F(\alpha) = F(\beta)$, and the result holds. So assume otherwise and write

$$F(\alpha) - F(\beta) = \sum_{k=1}^{n} c_k(e(k\alpha) - e(k\beta)).$$

Set $z = e(\alpha), w = e(\beta)$ in the identity

$$\frac{z^k - w^k}{z - w} = z^{k-1} + z^{k-2}w + \cdots + w^{k-1}$$

to obtain

$$\begin{aligned}\frac{1}{e(\alpha)-e(\beta)}\sum_{k=1}^{n}c_k(e(k\alpha)-e(k\beta)) &= \sum_{k=1}^{n}c_k\sum_{j=0}^{k-1}e(j\alpha)e((k-1-j)\beta)\\ &= \sum_{j=0}^{n-1}e(j\alpha)\sum_{k=j+1}^{n}c_k e((k-1-j)\beta)\\ &= \sum_{j=0}^{n-1}e(j\alpha)e((-1-j)\beta)\sum_{k=j+1}^{n}c_k e(k\beta).\end{aligned}$$

The triangle inequality now implies

$$\begin{aligned}|F(\alpha)-F(\beta)| &\le |e(\alpha)-e(\beta)|\sum_{j=0}^{n-1}\left|\sum_{k=j+1}^{n}c_k e(k\beta)\right|\\ &= |e(\alpha)-e(\beta)|\sum_{j=0}^{n-1}|F_{n-j}(\beta)| = |e(\alpha)-e(\beta)|\sum_{j=1}^{n}|F_j(\beta)|.\end{aligned}$$

The result follows, since

$$e(\alpha)-e(\beta)=\int_{\alpha}^{\beta}e'(t)dt = 2\pi i\int_{\alpha}^{\beta}e(t)dt,$$

so that

$$|e(\alpha)-e(\beta)| \le 2\pi|\beta-\alpha|$$

by the M-L inequality. □

6.7.2 Proof of The Fundamental Lemma

Proof of the Fundamental Lemma. For n a positive integer, define the trigonometric polynomial

$$g(\theta) := \sum_{a\in\mathcal{M}(n,P)}e(a\theta) - C\sum_{a=1}^{n}e(a\theta).$$

Note that g depends on n; however, in what follows we suppress this dependence to avoid notational clutter.

We wish to show that $g(\theta)=o(n)$ uniformly for $\theta\in[0,1]$ as $n\to\infty$. Equivalently, we want to show that given any $\epsilon>0$, there exists $N_0(\epsilon)$ such that

$$\sup_{\theta\in[0,1]}\left|\frac{g(\theta)}{n}\right| < \epsilon \tag{6.30}$$

provided $n > N_0(\epsilon)$.

Given $\epsilon > 0$, choose positive integers K_0, K_1 such that

$$x > K_0 \Longrightarrow R(x)/x < \frac{\epsilon}{22}$$
$$x > K_1 \Longrightarrow R(x)/x < \frac{\epsilon}{22K_0}.$$

We now prove that we may take $N_0(\epsilon) = \max\{K_0, K_1\}$ in the above.

For $n > N_0$, set $Q = n/K_0$. Our strategy is to estimate $g(\theta)$ and the partial sums

$$g_j(\theta) := \sum_{\substack{a\in\mathcal{M}(n)\\ n-j+1\le a\le N}} e(a\theta) - C \sum_{n-j+1\le a\le n} e(a\theta) \quad (1 \le j \le n),$$

at those rational numbers p/q with $q \le Q$. By Dirichlet's approximation lemma, corresponding to every $\theta \in [0,1]$ is a rational number p/q with $|p/q - \theta| \le (qQ)^{-1}$. A bound for $g(\theta)$ will then follow from Lemma 6.7.3.

To start things off, write, for $1 \le j \le n$,

$$\begin{aligned} g_j\left(p/q\right) &= \sum_{\substack{a\in\mathcal{M}(n)\\ n-j+1\le a\le n}} e\left(ap/q\right) - C \sum_{n-j+1\le a\le n} e\left(ap/q\right) \\ &= \sum_{k=1}^{q} e\left(kp/q\right) \left(\sum_{\substack{a\in\mathcal{M}(n)\\ n-j+1\le a\le n\\ a\equiv k \pmod q}} 1 - C \sum_{\substack{n-j+1\le a\le n\\ a\equiv k \pmod q}} 1 \right). \end{aligned}$$

The first inner sum counts the size of

$$\{a \in \mathcal{M}(n) : n-j+1 \le a \le n, a \equiv k \pmod q\}$$
$$\subset \{n-j+1 \le a \le n : a \equiv k \pmod q\}.$$

Here the right hand side set is affine equivalent to a subset $\mathcal{S}$ of at most $\lceil j/q \rceil$ consecutive integers. The left hand side is the image under an affine map of a subset of S; since the left hand side set possesses P and P is an affine property, it follows that the cardinality of the left hand side is bounded above by $M(\lceil j/q \rceil) = M(j/q)$; this motivates us to reexpress

$$\begin{aligned} g_j(p/q) = -\sum_{k=1}^{q} e(kp/q) &\left(M(j/q) - \sum_{\substack{a\in\mathcal{M}(n)\\ n-j+1\le a\le n\\ a\equiv k \pmod q}} 1 \right) \\ &+ \sum_{k=1}^{q} e(kp/q) \left(M(j/q) - C \sum_{\substack{n-j+1\le a\le n\\ a\equiv k \pmod q}} 1 \right). \quad (6.31) \end{aligned}$$

Both parenthesized expressions are in fact nonnegative; the first of these was proven to be nonnegative above, and the latter is nonnegative owing to the relation

$$M(j/q) = M(\lceil j/q \rceil) \geq C\lceil j/q \rceil \geq C \sum_{\substack{n-j+1 \leq a \leq n \\ a \equiv k \pmod q}} 1.$$

Consequently, the triangle inequality applied to (6.31) yields

$$\begin{aligned}
|g_j(p/q)| &\leq \sum_{k=1}^{q} \left(M(j/q) - \sum_{\substack{a \in \mathcal{M}(n) \\ n-j+1 \leq a \leq n \\ a \equiv k \pmod q}} 1 \right) + \sum_{k=1}^{q} \left(M(j/q) - C \sum_{\substack{n-j+1 \leq a \leq n \\ a \equiv k \pmod q}} 1 \right) \\
&= 2qM(j/q) - Cj - \sum_{\substack{a \in \mathcal{M}(n) \\ n-j+1 \leq a \leq n}} 1.
\end{aligned}$$

Now substitute the estimate

$$\sum_{\substack{a \in \mathcal{M}(n) \\ n-j+1 \leq n \leq n}} 1 = M(n) - |\{a \in \mathcal{M}(n) : a \leq n-j\}| \geq M(n) - M(n-j)$$

and note that $M(n) \geq Cn$ to obtain

$$\begin{aligned}
|g_j(p/q)| &\leq 2qM(j/q) - Cj - Cn + M(n-j) \\
&= 2q(M(j/q) - Cj/q) + (M(n-j) - C(n-j)) \\
&\leq 2qR(j/q) + R(n-j) \\
&\leq 2qR(n/q) + R(n)
\end{aligned}$$

for $1 \leq j \leq n$. Since g has no constant term, $g = g_n$, so that the same estimate holds for $|g(p/q)|$.

Now apply Dirichlet's approximation theorem. Given $\theta \in [0,1]$, choose p/q with $q \leq Q$ such that $|\theta - p/q| \leq (qQ)^{-1}$. Taking $\beta = p/q$ in Lemma 6.7.3 implies

$$\begin{aligned}
|g(\theta)| &\leq |g(p/q)| + 2\pi \frac{1}{qQ} \sum_{j=1}^{n} |g_j(p/q)| \\
&\leq \left(1 + 2\pi \frac{1}{qQ} n\right) (2qR(n/q) + R(n)) \\
&= \left(1 + 2\pi \frac{K_0}{q}\right) (2qR(n/q) + R(n)).
\end{aligned}$$

We now consider two cases, depending on the size of q.

Case I: $q \le K_0$**:** Use the estimate $R(n/q) \le R(n)$ to obtain

$$\begin{aligned}|g(\theta)| &\le \left(1 + \frac{2\pi K_0}{q}\right)(2q+1)R(n) \le 3q\left(1 + \frac{2\pi K_0}{q}\right)R(n)\\ &= (3q + 6\pi K_0)R(n) \le K_0(3+6\pi)R(n) \le K_0(3+6\pi)\frac{\epsilon n}{22K_0}\\ &= \frac{3+6\pi}{22}\epsilon n < \epsilon n.\end{aligned}$$

Case II: $q > K_0$**:** Here we use the fact that $q \le Q = n/K_0$, so that $n \ge n/q \ge K_0$. This implies

$$\begin{aligned}|g(\theta)| &\le \left(1 + 2\pi\frac{K_0}{q}\right)(2qR(n/q) + R(n))\\ &\le (1+2\pi)(2qR(n/q) + R(n)) \le (1+2\pi)\left(2q\frac{\epsilon}{22}\frac{n}{q} + \frac{\epsilon}{22}n\right)\\ &= \frac{3+6\pi}{22}\epsilon n < \epsilon n.\end{aligned}$$

Thus the estimate (6.30) holds. As $\epsilon > 0$ was arbitrary, the result follows. □

Remark. The distinction made in the two cases above is an example of a phenomenon that occurs quite frequently in applications of the circle method. The behavior of a trigonometric sum near a number α often depends on how well-approximated α is by a rational number with "small" denominator. In applications, this often necessitates breaking up the range of integration $[0,1]$ (or some other appropriately chosen unit interval) into *major* and *minor* arcs, the former corresponding to the numbers well-approximated by rationals with small denominator, and the latter taken to be the complement of this set.

6.7.3 Proof of Lemma 6.5.3

We now deduce Lemma 6.5.3 as a corollary of the Fundamental Lemma.

Proof of Lemma 6.5.3. Let

$$f(\theta) = \sum_{a \in \mathcal{M}(n,P)} e(a\theta). \tag{6.32}$$

The number of three-term arithmetic progressions from $\mathcal{M}(n,P)$, counting order and allowing equality, is given by

$$\int_0^1 f(\theta)^2 f(-2\theta)\, d\theta. \tag{6.33}$$

Write

$$f(\theta) = Ch(\theta) + g(\theta), \tag{6.34}$$

where

$$h(\theta) := \sum_{a=1}^{n} e(a\theta). \tag{6.35}$$

Thus

$$\begin{aligned} f(\theta)^2 f(-2\theta) &= (Ch(\theta) + g(\theta))^2 (Ch(-2\theta) + g(-2\theta)) \\ &= (C^2 h(\theta)^2 + 2Ch(\theta)g(\theta) + g(\theta)^2)(Ch(-2\theta) + g(-2\theta)), \end{aligned}$$

so that the integral (6.33) can be written as $I_1 + I_2 + \cdots + I_6$, where

$$\begin{aligned} I_1 &= \int_0^1 C^3 h(\theta)^2 h(-2\theta)\, d\theta, & I_2 &= \int_0^1 C^2 h(\theta)^2 g(-2\theta)\, d\theta, \\ I_3 &= \int_0^1 2C^2 h(\theta) h(-2\theta) g(\theta)\, d\theta, & I_4 &= \int_0^1 2Ch(\theta) g(\theta) g(-2\theta)\, d\theta, \\ I_5 &= \int_0^1 Cg(\theta)^2 h(-2\theta)\, d\theta, & I_6 &= \int_0^1 g(\theta)^2 g(-2\theta)\, d\theta. \end{aligned}$$

But $\int_0^1 h(\theta)^2 h(-2\theta)\, d\theta$ counts the total number of three-term arithmetic progressions from $\{1, 2, \ldots n\}$, which we already determined in Lemma 6.5.2 to be $\lceil n^2/2 \rceil$. Thus

$$I_1 = C\left(\frac{n^2}{2} + O(1)\right) = C\frac{n^2}{2} + o(n^2).$$

It suffices now to show that $I_2, I_3, \ldots, I_6$ are all $o(n^2)$. We can write each of the integrands appearing in any of these, up to a constant, in the form $f_1(\theta)f_2(\theta)f_3(\theta)$, where $f_1(\theta)$ and $f_2(\theta)$ are among the functions $h(\theta)$, $h(-2\theta)$, $g(\theta)$, and $g(-2\theta)$, and where $f_3(\theta)$ is either $g(\theta)$ or $g(-2\theta)$.

Note that both $h(\theta)$ and $g(\theta)$ can be written as $\sum_j c_j e(j\theta)$, where j runs over a set of at most n integers and $|c_j| \le 1$ for each j. Replacing θ by by -2θ if necessary, we see that each f_i admits such a representation. It follows now by Parseval's inequality that for each i,

$$\int_0^1 |f_i|^2\, d\theta = \sum_j |c_j|^2 \le n. \tag{6.36}$$

Now write

$$\begin{aligned} \int_0^1 |f_1 f_2 f_3| &\le \sup_{\theta\in[0,1]} |f_3(\theta)| \int_0^1 |f_1(\theta)||f_2(\theta)|\, d\theta \\ &\le \sup_{\theta\in[0,1]} |f_3(\theta)| \left(\int_0^1 |f_1(\theta)|^2\, d\theta\right)^{1/2} \left(\int_0^1 |f_2(\theta)|^2\, d\theta\right)^{1/2}. \end{aligned}$$

Estimating the integrals by (6.36), we see that for $2 \le j \le 6$,

$$I_j \ll \left(\sup_{\theta\in[0,1]} |f_3(\theta)|\right) n^{1/2} n^{1/2} = o(n^2),$$

since, by the fundamental lemma,

$$\sup_{\theta\in[0,1]} |f_3(\theta)| = \sup_{\theta\in[0,1]} |g(\theta)| = o(n). \qquad \square$$

6.7.4 Exercises

Exercise 6.7.1 *(Another Proof of Dirichlet's Approximation Lemma).* Show that if $\alpha_0, \dots, \alpha_N$ are $N+1$ real numbers, then $\|\alpha_j - \alpha_i\| \le (N+1)^{-1}$ for some $i \ne j$. Use this result to give another proof of Lemma 6.7.2, by taking $\alpha_i = i\alpha$ for $i = 0, \dots, Q$.

Suggestion: One way to proceed is by observing that $2\pi\|\alpha_j - \alpha_i\|$ is the shortest distance along the unit circle between $e(\alpha_j)$ and $e(\alpha_i)$.

Exercise 6.7.2 *(A Generalization of Roth's Theorem, †).* Let $a_1, \dots, a_k$ be $k \ge 3$ integers which sum to zero. Let P denote the property of possessing no nontrivial solution to the equation

$$a_1x_1 + a_2x_2 + \cdots + a_3x_3 = 0.$$

Show that P is an affine property and that $C_P = 0$. (Roth's theorem is the case $k = 3$ and $a_1 = a_2 = 1, a_3 = -2$.)

6.8 The Number of Three Term Progressions

Let $0 < \delta \le 1$. By Roth's theorem, any subset of $[1, n]$ with at least δn elements contains a three-term arithmetic progression provided n is large enough. We now derive from the same theorem a lower bound for the number of such progressions:

Theorem 6.8.1 (P. Varnavides [Var59]). *Let $0 < \delta \le 1$. If $n > n_0(\delta)$, and $S \subset [1, n]$ is a set of more than δn positive integers, then there are at least $C(\delta)n^2$ three term arithmetic progressions of elements of S. Here trivial arithmetic progressions (of three equal terms) are* not *allowed and progressions are counted without regard to order.*

Remark. The theorem is best possible, for it follows from Lemma 6.5.2 that the total number of three-term arithmetic progressions drawn from all of $[1, n]$ is, counted as above and with lower order terms discarded, simply $n^2/4$.

If k is large, any subset of $[1, k]$ with more than $\frac{\delta}{2}k$ integers contains a nontrivial three-term arithmetic progression; indeed, this is nothing more than the assertion that $r_3(k) \le \frac{\delta}{2}k$ for all large k, which is immediate from Roth's theorem. Fix $k = k(\delta)$ to be the smallest positive integer exceeding 8 with this property.

We now consider all k-term arithmetic progressions of the form

$$1 \le u < u + d < \cdots < u + (k-1)d \le n. \tag{6.37}$$

Such an arithmetic progression will be called "good" if it contains at least $\frac{\delta}{2}k$ terms of S. The proof of Theorem 6.8.1 rests on the following estimate for the number of good progressions:

Lemma 6.8.2. *There are constants $n_1 = n_1(\delta, k) = n_1(\delta)$ and $C_1 = C_1(\delta, k) = C_1(\delta)$ for which the number G of "good progressions" satisfies $G > C_1 n^2$ whenever $n > n_1$.*

Proof. We wlll show that for small d, namely

$$1 \le d < \frac{\delta}{k^2} n, \tag{6.38}$$

there are $\gg_\delta n$ good progressions with common difference d. Adding up the contribution from each such d gives the result.

Fix a value of d satisfying (6.38). Let $S' = S \cap [kd, n - kd]$. If the element $a \in S'$ occurs in the progression (6.37), then

$$u \equiv a \pmod{d}, \quad u \le a \le u + (k-1)d; \tag{6.39}$$

conversely, if u satisfies these conditions, then

$$u \ge a - (k-1)d \ge d \ge 1, \quad u + (k-1)d \le a + (k-1)d \le n - d \le n,$$

and the progression (6.37) contains a. So the number of times a given element $a \in S'$ occurs in a progression of the form (6.37) is precisely the number of integers satisfying the conditions (6.39). There are exactly k of these: namely, $u = a$ and the $k-1$ integers $u \equiv a \pmod{d}$ in the range $a - (k-1)d \le u < a$. Since S' contains at least

$$|S| - 2kd > \delta n - 2\frac{\delta}{k} n = \delta(1 - 2/k)n$$

elements, we have, with $S_{u,d}$ denoting the number of terms of S in the progression (6.37),

$$\sum_{u: 1 \le u + (k-1)d \le n} S_{u,d} \ge k|S'| > k\delta(1 - 2/k)n > \frac{3}{4}\delta k n.$$

(Recall $k > 8$.) As d is fixed, there are no more than N progressions of the form (6.37), whence the upper bound

$$\sum_{u: 1 \le u + (k-1)d \le n} S_{u,d} \le \frac{1}{2}\delta k \cdot n + k \cdot G_d;$$

here we have written G_d for the number of good progressions of difference d. Combining this with the previous lower bound for the same sum shows

$$G_d \ge \frac{3}{4}\delta n - \frac{1}{2}\delta n = \frac{1}{4}\delta n.$$

Consequently,

$$G = \sum G_d \ge \sum_{1 \le d < \frac{\delta}{k^2} n} \frac{1}{4}\delta n \ge \frac{1}{4}\delta n \cdot \frac{\delta}{2k^2} n = \frac{\delta^2}{8k^2} n^2,$$

provided $n > n_1(\delta, k) = n_1(\delta)$. The lemma follows with this n_1 and $C_1 = \frac{\delta^2}{8k^2}$. □

Proof of Theorem 6.8.1. Let P be the set of elements of a progression of the form (6.37). We map P bijectively onto $\{1, 2, \ldots, k\}$ by the map

$$\phi_P(a) = 1 + (a - u)/d.$$

As an affine map, this preserves arithmetic progressions, and

$a_1 < a_2 < a_3$ form a 3-term AP $\Longleftrightarrow \phi(a_1) < \phi(a_2) < \phi(a_3)$ form a 3-term AP.

If P is a good progression, then

$$\phi(P \cap S) \subset \{1, 2, \ldots, k\}, \quad |\phi(P \cap S)| = |P \cap S| > \frac{\delta}{2}k;$$

by the choice of k, $\phi(P \cap S)$ contains a nontrivial three-term arithmetic progression, and the same must then hold for $P \cap S$. So from each good progression P, we can choose a corresponding nontrivial three-term arithmetic subprogression of elements of S. By the preceding lemma, more than $C_1(\delta)n^2$ (not necessarily distinct!) subprogressions are produced by this process, provided $n > n_1(\delta)$.

Let P' be such a subprogression, so that its common difference d' satisfies

$$d' \leq (k - 1)d/2 < kd/2. \tag{6.40}$$

If P is a k-term progression given by (6.37) which contains P', then its difference d must divide d', say $d = d'/t$; by (6.40), we have $t < k/2$. So there are at most $k/2$ possibilities for the common difference of P. For any fixed common difference d, there are most at most $k - 2$ k-term progressions P of the form (6.37) which contain P'.

Consequently, P' can be contained in no more than

$$(k/2)(k - 2) =: C_2(k, \delta) = C_2(\delta)$$

such arithmetic progressions. Hence, with $C := C_1/C_2$ and $n_0 := n_1$, there are more than Cn^2 three-term progressions consisting of elements of S, provided $n > n_0$. ☐

6.9 The Higher-Dimensional Situation

In this section we digress from the main matter to discuss (without proofs) higher dimensional analogs of of the theorems of van der Waerden and Szemerédi.

This next result of Gallai (appearing in [Rad33]) generalizes van der Waerden's theorem to several dimensions (see Exercise 6.9.1). A short elementary proof can be found in [And76]:

Theorem (Gallai). *Let F be a finite subset of $\mathbf{N}^d$. For any r-coloring of $\mathbf{N}^d$, there is a positive integer a and a point $\mathbf{v} \in \mathbf{N}^d$ for which the set $a \times F + \mathbf{v} := \{a\mathbf{f} + \mathbf{v} : \mathbf{f} \in F\}$ is monochromatic. Moreover, the dilation factor a and the coordinates of $\mathbf{v}$ are bounded by a function depending only on F and r (and not on the particular k-coloring).*

By contrast, there is no elementary proof known for the corresponding analog of Szemerédi's theorem in higher dimensions:

Theorem (Furstenberg & Katznelson [FK78]). *Let $S \subset \mathbf{Z}^d$. Suppose that for some sequence of parallelepipeds $\Pi_n = [a_n^{(1)}, b_n^{(1)}] \times \cdots \times [a_n^{(d)}, b_n^{(d)}] \subset \mathbf{Z}^d$, with $b_n^{(i)} - a_n^{(i)} \to \infty$, $i = 1, 2, \ldots, d$, we have*

$$|S \cap \Pi_n| / |\Pi_n| > \epsilon$$

for some $\epsilon > 0$. If F is a finite subset of $\mathbf{Z}^d$, then there exists a positive integer a and a point $\mathbf{v} \in \mathbf{Z}^d$ for which $a \times F + \mathbf{v} \subset S$.

So far everything generalizes as well as one could hope. But now consider the form of van der Waerden's theorem put forth by Rabung: a sequence with bounded gaps contains arbitrarily long arithmetic progressions. This next result of Dekking shows this need not be true in higher dimensions:

Theorem (Dekking [Dek79]). *There is an infinite sequence $\mathbf{v}_0, \mathbf{v}_1, \dots \in \mathbf{Z}^2$ with each $\mathbf{v}_{i+1} - \mathbf{v}_i = (0,1)$ or $(1,0)$, and such that no five of the $\mathbf{v}_i$ form an arithmetic progression of vectors.*

That is, one can walk to infinity in $\mathbf{Z}^2$ taking only unit steps right and up in such a way that one assumes no five positions in arithmetic progression.

It is easy to describe the sequence Dekking constructs, though not easy to prove it has the stated property: Think of a and b as formal symbols, and consider the set of all "words" on a and b (finite or countably infinite nonempty concatenations of these symbols). We define a function θ from the set of words to itself, given by taking $\theta(a) = abb$, $\theta(b) = aaab$, and extending to all words by concatenation. Because $\theta(a)$ begins with a, $\theta^k(a)$ will begin with $\theta^{(k-1)}(a)$ for every positive integer k. This implies that for any given n, the first n symbols of $\theta^k(a)$ are eventually constant (i.e., constant for $k > k_0(n)$); this allows us to define an infinite "limit word" $\theta^\infty(a)$. Then the sequence of moves is given by interpreting every 0 in the limit word as a unit step in the direction $(0,1)$, and every 1 as a unit step in the direction $(1,0)$.

In contrast to Dekking's result, we have the following theorem, proposed by T.C. Brown as an advanced problem in the *Monthly*:

Theorem 6.9.1. *Let $\mathbf{v}_0, \mathbf{v}_1, \mathbf{v}_2, \ldots$ be an infinite sequence of points in $\mathbf{Z}^2$ with $\mathbf{v}_{i+1} - \mathbf{v}_i = (1,0)$ or $(0,1)$ for each i. Then for every positive integer k, there is a line containing k of the $\mathbf{v}_i$.*

Proof (P.L. Montgomery [BM72]). Without loss of generality we may assume $\mathbf{v}_0 = \mathbf{0}$. Let $k > 1$ be given. For each i, write $\mathbf{v}_i = (x_i, y_i)$. An easy induction shows $x_i + y_i = i$ for each i. Let

$$s_i = x_i / i \quad (i = 1, 2, \ldots).$$

To begin with, consider the special case where some s shows up infinitely often in the sequence of s_i; then the line

$$\frac{x}{x+y} = s$$

contains *infinitely many* of the $\mathbf{v}_i$. We are not guaranteed such an s exists, but compactness tells us there is always some limit point l of the s_i in $[0, 1]$.

We consider two cases, according as l is or is not the limit of the entire sequence of s_i. If it is not, then for some $\epsilon > 0$, $|s_i - l| \geq \epsilon$ for infinitely many i. It follows that we may choose a rational number $r = p/q$ for which

$$s_i - p/q = \frac{qx_i - pi}{qi}$$

changes sign infinitely often. Indeed, if $s_i - l \geq \epsilon$ infinitely often, let r be any rational from the interval $(l, l + \epsilon)$, otherwise choose r from $(l - \epsilon, l)$.

It follows that $qx_i - pi$ is integer valued and changes sign infinitely often, but

$$|(qx_{i+1} - p(i + 1)) - (qx_i - pi)| = |q(x_{i+1} - x_i) - p| \leq q + p$$

is bounded. Consequently, $qx_i - pi$ assumes the same value t infinitely often. Then the line

$$qx - p(x + y) = t \tag{6.41}$$

contains infinitely many of the $\mathbf{v}_i$.

Now suppose l is the limit of the s_i. Let $r = p/q$ be a rational approximation to l, *not* necessarily in lowest terms, satisfying

$$|p/q - l| < \frac{1}{2kq}, \quad q > k.$$

(This is clearly possible if $l \in \mathbf{Q}$, otherwise invoke Lemma 6.7.2.) We claim that the line (6.41) contains at least k of the $\mathbf{v}_i$. For this, it suffices to show $qx_i - pi$ assumes the same value at least k times, as i ranges over $0, 1, 2, \ldots$. If this is false, then for each $j \geq 1$ one can choose i, $0 \leq i \leq (k - 1)(2j + 1)$, with

$$|qx_i - pi| \geq j + 1. \tag{6.42}$$

But for any such (necessarily nonzero) i,

$$\left|\frac{x_i}{i} - \frac{p}{q}\right| \geq \frac{j + 1}{qi} \geq \frac{j + 1}{q(k - 1)(2j + 1)} \geq \frac{1}{2kq}. \tag{6.43}$$

Letting

$$\epsilon := \frac{1}{2kq} - |p/q - l| > 0,$$

we see

$$|s_i - l| = \left|\frac{x_i}{i} - l\right| \geq \epsilon. \tag{6.44}$$

Moreover, since we can take j arbitrarily large, (6.42) shows there are infinitely many i satisfying (6.43) and hence (6.44). But this contradicts our supposition that $s_i \to l$. $\square$

This theorem suggests that when studying sequences of bounded gaps in $\mathbf{Z}^2$, we should look for large configurations of collinear points, not necessarily in arithmetic progression (equally spaced). Thus the proper generalization of Rabung's form of van der Waerden's theorem is as follows:

Theorem (Ramsey & Gerver [GR79]). *Let $B > 0$ and let k be a positive integer. There exists a number $N = N(B, k)$ with the property that if $n \geq N(B, k)$ and $\mathbf{v}_0, \mathbf{v}_1, \dots, \mathbf{v}_n$ is any sequence of $n+1$ points of $\mathbf{Z}^2$ with*

$$\|\mathbf{v}_{i+1} - \mathbf{v}_i\| \leq B \qquad (i = 0, 1, \dots, n-1), \tag{6.45}$$

then k of the $\mathbf{v}_i$ are in arithmetic progression.

Pomerance [Pom80] has succeeded in proving the proper analog of Szemeredi's theorem in this situation: it suffices the gap be bounded on average, in that the last theorem remains true if the condition (6.45) is replaced by

$$\frac{1}{n} \sum_{i=0}^{n-1} \|\mathbf{v}_{i+1} - \mathbf{v}_i\| \leq B.$$

What about dimensions 3 and higher? Ramsey & Gerver, in the already-cited paper, construct a walk in 3-space showing that even the analog of Theorem 6.9.1 fails.

6.9.1 Exercises

Exercise 6.9.1. Show that Gallai's theorem implies van der Waerden's in the third form listed in Theorem 6.2.1.

Exercise 6.9.2. Show that Dekking's theorem would be false if 5 were replaced by 4. Suggestion: Show that every infinite word on the two symbols a, b contains three consecutive blocks that are permutations of each other.

Exercise 6.9.3. Prove that the assumptions of Theorem 6.9.1 do not imply the existence of a line containing infinitely many of the $\mathbf{v}_i$.

Exercise 6.9.4 *(Pomerance [Pom79, Theorem 4.1], †).* The quoted theorem of Pomerance implies that if $a_1 < a_2 < \dots$ is an increasing sequence with positive lower density then there are k collinear points on the graph $\{(n, a_n)\}$ for every k. In this exercise we prove the same when $a_n = p_n$, the nth prime.

It suffices to produce collinear points on the inverse of the prime number graph, $\{(p_n, n) : n = 1, 2, \dots\}$.

a) Let k be a positive integer. Let $u = e^k$, $v = 2e^k$, and let T be the (closed) parallelogram bounded by the vertical lines $x = u$, $x = v$ and the diagonal lines with slope $1/k$ through $(u, \mathrm{li}(u) + u/\log^3 u)$ and $(v, \mathrm{li}(u) - u/\log^3 u)$.

   Prove that there are $\ll ku/\log^3 u$ lines of slope $1/k$ passing through lattice points contained in T (as $k \to \infty$).

b) Using the prime number theorem in the strong form

$$\pi(x) = \mathrm{li}(x) + O(x/\log^5 x) \quad (x \to \infty),$$

prove that every point (p_n, n) with $u \le p_n \le v$ lies in T once k is sufficiently large.

c) Show that as $k \to \infty$, there are $\gg u/\log u$ points (p_n, n) with $u \le p_n \le v$. Conclude from a) and b) that there is a line of slope $1/k$ passing through $\gg \frac{1}{k}\log^2 u = k$ of these points.

6.10 Behrend's Lower Bound for $r_3(n)$

The aim of this section is to describe Behrend's construction of dense sets free of arithmetic progressions. As a consequence, we shall obtain the best known lower bound for $r_3(N)$ (up to the value of the constant c):

Theorem 6.10.1. *As $N \to \infty$,*

$$r_3(N) > N \exp((-c + o(1))\sqrt{\log N}),$$

where $c = 2\sqrt{2\log 2}$.

We first introduce a multi-dimensional analog of our problem. For integers $n, d \ge 2$, let $R(n, d)$ be the set of lattice points in the n-dimensional interval $[0, d)^n$, i.e.,

$$R(n, d) := \{\mathbf{x} = (x_0, x_1, \ldots, x_{n-1}) \in \mathbf{N}^n : x_i < d \text{ for } i = 1, \ldots, n\}.$$

We set about to estimate from below the largest subset of $R(n, d)$ free of arithmetic progressions. We shall obtain the required bound through clever application of the following geometric lemma:

Lemma 6.10.2. *Let S be a sphere in $\mathbf{R}^n$, and let L be a line. Then L intersects S in at most two distinct points.*

Proof. Without loss of generality, we may assume the sphere is centered at the origin. Let S be defined by $\|\mathbf{x}\|^2 = R$, and let the line L be parametrized by $\mathbf{u} + t\mathbf{v}$, $t \in \mathbf{R}$. The points of intersection correspond to those real t with

$$R = \|\mathbf{u} + t\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + 2(\mathbf{u} \cdot \mathbf{v})t + t^2\|v\|^2.$$

Since a quadratic equation has at most 2 real roots, the result follows. □

This lemma suggests partitioning the points of $R(n, d)$ according to the their distance from the origin (absolute value). Thus, we define

$$S_k(n, d) := \{\mathbf{x} \in R(n, d) : \|x\|^2 = k\}.$$

Since

$$\|x\|^2 = x_0^2 + \cdots + x_{n-1}^2 \leq n(d-1)^2,$$

we have

$$R(n,d) = \dot{\bigcup}_{k=0}^{n(d-1)^2} S_k(n,d).$$

Since $|R(n,d)| = d^n$, it follows that for some $0 \leq k \leq d(n-1)^2$,

$$|S_k(n,d)| \geq \frac{d^n}{n(d-1)^2+1} \geq \frac{d^n}{nd^2} = \frac{d^{n-2}}{n}.$$

This set $S_k(n,d)$ is free of any nontrivial arithmetic progression. Indeed, if $\mathbf{x}, \mathbf{y}, \mathbf{z} \in S_k(n,d)$ were in arithmetic progression with $\mathbf{x} \neq \mathbf{z}$, then the line through $\mathbf{x}$ and $\mathbf{z}$ would intersect the sphere $\|\mathbf{x}\|^2 = k$ in the three distinct points $\mathbf{x}, \mathbf{z}, (\mathbf{x}+\mathbf{z})/2 = \mathbf{y}$, contradicting Lemma 6.10.2. Thus we have shown:

Lemma 6.10.3. *For integers $n, d \geq 2$, there exists an arithmetic-progression free subset of $R(n,d)$ with at least d^{n-2}/n elements.*

This is an interesting result in itself. The relevance for the original problem comes from the existence of the "nice" map $\psi_{n,d} \colon R_{n,d} \to \mathbf{N}$, given by

$$(x_0, x_1, \ldots, x_{n-1}) \mapsto x_0 + x_1(2d-1) + \cdots + x_{n-1}(2d-1)^{n-1}.$$

What makes this map so nice? First, it is injective, as follows from uniqueness of radix $2d-1$ representation. Second, if $\psi(\mathbf{x}) + \psi(\mathbf{y}) = 2\psi(\mathbf{z})$, then $\mathbf{x}+\mathbf{y} = 2\mathbf{z}$. Indeed, the given equation translates into the condition that $\sum_{i=0}^{n-1}(x_i+y_i)(2d-1)^i = \sum_{i=0}^{n-1} 2z_i(2d-1)^i$. Since each of x_i, y_i, z_i are at most $d-1$, uniqueness of the base $2d-1$ expansion forces $x_i + y_i = 2z_i$ for each i, so that $\mathbf{x}+\mathbf{y} = 2\mathbf{z}$ as claimed.

Together, these two properties imply that if $S \subset R(n,d)$ is A.P. free, then $\psi(S) \subset \mathbf{N}$ is an A.P. free subset with the same size. If we note that the image of any point of $R_{n,d}$ under ψ is always smaller than

$$(d-1)(2d-1)^{n-1}\sum_{i=0}^{\infty}(2d-1)^{-i} = \frac{1}{2}(2d-1)^n < (2d-1)^n,$$

we see that there is always a subset of the nonnegative integers less than $(2d-1)^n$, of size at least d^{n-2}/n, with no three terms in arithmetic progression. Shifting everything by 1 now implies

$$r_3((2d-1)^n) \geq d^{n-2}/n.$$

Our strategy for proving Theorem 6.10.1 will be to choose n as a slow-growing function of N, defined whenever N is sufficiently large. We will then determine the integer $d \geq 2$ (as a function of N) by the inequalities

$$(2d-1)^n \leq N < (2d+1)^n. \tag{6.46}$$

The choice of n will be optimized so that the right hand side of the estimate

$$r_3(N) \geq r_3((2d-1)^n) \geq \frac{d^{n-2}}{n} > \frac{(N^{1/n}-1)^{n-2}}{n2^{n-2}} = \frac{N^{1-2/n}}{n2^{n-2}}(1-N^{-1/n})^{n-2} \tag{6.47}$$

is essentially as large as possible.

Note that for this plan to be carried out, for every large N there should be an integer $d \geq 2$ for which (6.46) is satisfied. This will be the case as long as $3^n \leq N$, so as long as $n \leq \log N/\log 3$; in particular, it will hold if $n = o(\log N)$.

We begin by maximizing the term

$$\frac{N^{1-2/n}}{n2^{n-2}}$$

appearing on the right hand side of (6.47). Write

$$\frac{N^{1-2/n}}{n2^{n-2}} = \frac{N^{1-2/n}}{n2^{n-1}} = N^{1-\frac{2}{n}-\frac{\log n}{\log N}-\frac{(n-2)\log 2}{\log N}} = N^{1-g}, \tag{6.48}$$

say. Now fix N, and think of g as a function only of n. As long as N is large,

$$g'(n) = \frac{-2}{n^2} + \frac{1}{n\log N} + \frac{\log 2}{\log N} = \frac{1}{n^2}\left(-2 + \frac{1}{\log N}n + \frac{\log 2}{\log N}n^2\right)$$

has a unique positive zero at

$$n_0 = \frac{-\frac{1}{\log N} + \sqrt{\frac{1}{\log^2 N} + 8\frac{\log 2}{\log N}}}{2\log 2/\log N} = \sqrt{\frac{2\log N}{\log 2}} + O(1).$$

It can be shown (again, assuming N is large) that this is actually the global minimum of g on the interval $[2,\infty)$. This motivates choosing

$$n = \left\lfloor \sqrt{\frac{2\log N}{\log 2}} \right\rfloor.$$

Notice that the hypothesis $n = o(\log N)$ is satisfied for this choice of n.

It straightforward to establish each of the estimates

$$\begin{aligned}
\frac{2}{n} &= \frac{\sqrt{2}\log 2 + O(1/\sqrt{\log N})}{\sqrt{\log N}} = \frac{\sqrt{2}\log 2 + o(1)}{\sqrt{\log N}}, \\
\frac{\log n}{\log N} &= \frac{O(\log\log N/\sqrt{\log N})}{\sqrt{\log N}} = \frac{o(1)}{\sqrt{\log N}}, \\
\frac{(n-1)\log 2}{\log N} &= \frac{\sqrt{2\log 2} + O(1/\sqrt{\log N})}{\sqrt{\log N}} = \frac{\sqrt{2}\log 2 + o(1)}{\sqrt{\log N}}.
\end{aligned}$$

Thus with this choice of n,

$$g = \frac{2\sqrt{2}\log 2 + o(1)}{\sqrt{\log N}},$$

so that

$$\frac{N^{1-2/n}}{n2^{n-2}} = N^{1-g} = N\exp((-c+o(1))\sqrt{\log N}).$$

Referring to (6.47), we see that to estimate $r_3(N)$ from below, it remains only to obtain a lower bound for $(1-N^{-1/n})^{n-2}$. But for our choice of n, we in fact have $(1-N^{-1/n})^{n-2} \to 1$; indeed, this holds whenever $n = o(\log N/\log\log N)$. It follows that for large N,

$$r_3(N) > \frac{1}{2}\frac{N^{1-2/n}}{n2^{n-2}} \geq N\exp((-c+o(1))\sqrt{\log N}).$$

6.10.1 Exercises

Exercise 6.10.1. Verify the unproved claims made in the course of our choice of the function $n = n(N)$. In particular:

a) Check that if we choose $n = o(\log N/\log\log N)$, then $(1-N^{-1/n})^{n-2} \to 1$ as $N \to \infty$.

b) Show that if N is sufficiently large the global minimum of g on $[2,\infty]$ occurs at the point n_0.

Exercise 6.10.2. Show directly from (6.47) that if we choose $n \geq 2$ as a constant function of N, then we obtain the estimate $r_3(N) \gg_n N^{1-2/n}$ as $N \to \infty$. Thus if we only wish to show $r_3(N) \gg_\epsilon N^{1-\epsilon}$ for each $\epsilon > 0$, we can forego the detailed analysis given above.

Exercise 6.10.3 *(Salem & Spencer).* Let n, d be positive integers with $d \mid n$ and $d > 1$. Let $S_{n,d}$ be the set of nonnegative integers with base $2d-1$ expansion of the form $a_{n-1}a_{n-2}\ldots a_0$, where exactly n/d of the a_i are equal to each of $0, 1, \ldots, d-1$. Then $S_{n,d}$ contains no nontrivial three-term arithmetic progression.

Salem & Spencer [SS42] used the result of this exercise to obtain the bound

$$r_3(N) > N\exp\left(-c\frac{\log N}{\log\log N}\right),$$

for any constant $c > \log 2$ and all $N > N_0(c)$.

References

[And76] Peter G. Anderson, *A generalization of Baudet's conjecture (van der Waerden's theorem)*, Amer. Math. Monthly **83** (1976), no. 5, 359–361. MR 53 #160

[Beh38] F. A. Behrend, *On sequences of integers containing no arithmetic progression*, Časopis Pěst. Mat. **67** (1938), 235–239.

[Beh46] ———, *On sets of integers which contain no three terms in arithmetical progression*, Proc. Nat. Acad. Sci. U. S. A. **32** (1946), 331–332. MR 8,317d

[BM72] T.C. Brown and P.L. Montgomery, *Problem 5811*, Amer. Math. Monthly **79** (1972), no. 10, 1143–1144.

[Bou99] J. Bourgain, *On triples in arithmetic progression*, Geom. Funct. Anal. **9** (1999), no. 5, 968–984. MR 2001h:11132

[Bra69] Alfred Brauer, *Combinatorial methods in the distribution of kth power residues*, Combinatorial Mathematics and its Applications (Proc. Conf., Univ. North Carolina, Chapel Hill, N.C., 1967), Univ. North Carolina Press, Chapel Hill, N.C., 1969, pp. 14–37. MR 40 #1330

[Bro75] T. C. Brown, *Variations on van der Waerden's and Ramsey's theorems*, Amer. Math. Monthly **82** (1975), no. 10, 993–995. MR 52 #2915

[Dek79] F. M. Dekking, *Strongly non-repetitive sequences and progression-free sets*, J. Combin. Theory Ser. A **27** (1979), no. 2, 181–185. MR 81b:05027

[ES03] Paul Erdős and János Surányi, *Topics in the theory of numbers*, Undergraduate Texts in Mathematics, Springer-Verlag, New York, 2003, Translated from the second Hungarian edition by Barry Guiduli. MR 2003j:11001

[ET36] P. Erdős and P. Turán, *On some sequences of integers*, J. London Math. Soc. **11** (1936), 261–264.

[FK78] H. Furstenberg and Y. Katznelson, *An ergodic Szemerédi theorem for commuting transformations*, J. Analyse Math. **34** (1978), 275–291 (1979). MR 82c:28032

[GR74] R. L. Graham and B. L. Rothschild, *A short proof of van der Waerden's theorem on arithmetic progressions*, Proc. Amer. Math. Soc. **42** (1974), 385–386. MR 48 #8257

[GR79] Joseph L. Gerver and L. Thomas Ramsey, *On certain sequences of lattice points*, Pacific J. Math. **83** (1979), no. 2, 357–363. MR 81c:10039

[Gra81] Ronald L. Graham, *Rudiments of Ramsey theory*, CBMS Regional Conference Series in Mathematics, vol. 45, American Mathematical Society, Providence, R.I., 1981. MR 82j:05018

[GT04] Ben Green and Terence Tao, *The primes contain arbitrarily long arithmetic progressions*, Preprint, April 2004.

[LeV96] William J. LeVeque, *Fundamentals of number theory*, Dover Publications Inc., Mineola, NY, 1996, Reprint of the 1977 original. MR 97a:11002

[New81] Donald J. Newman, *Sequences without arithmetic progressions*, Analytic number theory (Philadelphia, Pa., 1980), Lecture Notes in Math., vol. 899, Springer, Berlin, 1981, pp. 311–314. MR 83h:10102

[New98] ______, *Analytic number theory*, Graduate Texts in Mathematics, vol. 177, Springer-Verlag, New York, 1998. MR 98m:11001

[Pom79] Carl Pomerance, *The prime number graph*, Math. Comp. **33** (1979), no. 145, 399–408. MR 80d:10013

[Pom80] ______, *Collinear subsets of lattice point sequences—an analog of Szemerédi's theorem*, J. Combin. Theory Ser. A **28** (1980), no. 2, 140–149. MR 81m:10104

[PS95] Carl Pomerance and András Sárközy, *Combinatorial number theory*, Handbook of combinatorics, Vol. 1, 2, Elsevier, Amsterdam, 1995, pp. 967–1018. MR 97e:11032

[Rab75] John R. Rabung, *On applications of van der Waerden's theorem*, Math. Mag. **48** (1975), 142–148. MR 51 #3092

[Rad33] Richard Rado, *Verallgemeinerung Eines Satzes von van der Waerden mit Anwendungen auf ein Problem der Zahlentheorie*, Sonderausg. Sitzunsber. Preuss. Akad. Wiss. Phys. Math. Klasse **17** (1933), 1–10.

[Rot52] Klaus Roth, *Sur quelques ensembles d'entiers*, C. R. Acad. Sci. Paris **234** (1952), 388–390. MR 13,724d

[Ruz99] Imre Z. Ruzsa, *Erdős and the integers*, J. Number Theory **79** (1999), no. 1, 115–163. MR 2002e:11002

[Sch76] Wolfgang M. Schmidt, *Equations over finite fields. An elementary approach*, Springer-Verlag, Berlin, 1976, Lecture Notes in Mathematics, Vol. 536. MR 55 #2744

[Soi95] Alexander Soifer, *Issai Schur: Ramsey theory before Ramsey*, Geombinatorics **5** (1995), no. 1, 6–23. MR 96j:01034

[SS42] R. Salem and D. C. Spencer, *On sets of integers which contain no three terms in arithmetical progression*, Proc. Nat. Acad. Sci. U. S. A. **28** (1942), 561–563. MR 4,131e

[Sze70] E. Szemerédi, *On sets of integers containing no four elements in arithmetic progression*, Number Theory (Colloq., János Bolyai Math. Soc., Debrecen, 1968), North-Holland, Amsterdam, 1970, pp. 197–204. MR 42 #7622

[Sze75] ______, *On sets of integers containing no k elements in arithmetic progression*, Acta Arith. **27** (1975), 199–245, Collection of articles in memory of Juriĭ Vladimirovič Linnik. MR 51 #5547

[Sze90] ———, *Integer sets containing no arithmetic progressions*, Acta Math. Hungar. **56** (1990), no. 1-2, 155–158. MR 92c:11100

[Var59] P. Varnavides, *On certain sets of positive density*, J. London Math. Soc. **34** (1959), 358–360. MR 21 #5595

[vdW27] B. L. van der Waerden, *Beweis einer Baudetschen Vermutung*, Nieuw Arch. Wiskd. **15** (1927), 212–216.

[vdW71] ———, *How the proof of Baudet's conjecture was found*, Studies in Pure Mathematics (Presented to Richard Rado), Academic Press, London, 1971, pp. 251–260. MR 42 #5764

Chapter 7

A Short Proof of Waring's Conjecture

> Every integer is a square or the sum of two, three, or four squares; every integer is a cube or the sum of at most nine cubes; every integer is also the square of a square, or the sum of up to nineteen such, and so forth. – Edward Waring, *Meditationes Algebraicae* (1770)

7.1 Introduction

In 1770, the same year Lagrange proved that every nonnegative integer could be written as a sum of four squares, Waring put forward analogous statements for higher powers, asserting that every positive integer was a sum of nine nonnegative cubes, nineteen nonnegative fourth powers, "and so forth." It has become usual to interpret this final clause as a statement of the following conjecture:

Waring's Conjecture. *For each $k \geq 1$, there exists $s = s(k) \geq 1$ for which*

$$n = \sum_{i=1}^{s} x_i^k, \qquad x_i \geq 0,$$

is solvable for each $n \geq 0$.

Expressed in the language of Chapter 5, this is the assertion that the set of nonnegative kth powers is a basis (of $\mathbf{N}$) of finite order for each $k = 1, 2, \ldots$.

Lagrange's four squares theorem tells us we may take $s(2) = 4$ in Waring's conjecture. If we define $g(k)$ as the least permissible value of $s(k)$ (the order of the basis), then we also have $g(2) = 4$, since every integer $\equiv 7 \pmod 8$ actually requires 4 squares in its representation. Keeping this notation, it is reasonable to read Waring's other assertions as claims that $g(3) = 9$ and $g(4) = 19$.

The first to prove the existence of $g(k)$ for all k was Hilbert [Hil09]; his proof depended on the four squares theorem and the existence of certain complicated algebraic identities; see Exercise 7.1.2 for an example of this sort of reasoning. A more fruitful line of attack was introduced by Hardy & Littlewood [HL20] and later refined by Vinogradov. The starting point is the realization that

$$r_{k,s}(n) := \sum_{\substack{x_1^k+\cdots+x_s^k=n \\ x_i \geq 0}} 1$$

can be expressed as an integral. Namely, with

$$f(\alpha) := \sum_{m=0}^{N} e(\alpha m^k),$$

where $N = \lfloor n^{1/k} \rfloor$, we have

$$r_{k,s}(n) = \int_0^1 f(\alpha)^s e(-\alpha n)\, d\alpha.$$

(See the discussion in §6.6.2 if this is unfamiliar.) The behavior of $r_{k,s}(n)$ for fixed k and large s can then be analyzed in terms of the behavior of f. Roughly speaking, it turns out that f is small away from numbers well-approximable by rationals while it can be conveniently estimated near such numbers. Carrying forward this (hard) analysis, Hardy & Littlewood proved that for fixed k and large s, say $s \geq s_1(k)$,

$$r_{k,s}(n) = \frac{\Gamma(1+1/k)^s}{\Gamma(s/k)} n^{s/k-1} \mathfrak{G}(n) + o(n^{s/k-1}) \quad (n \to \infty), \tag{7.1}$$

where Γ is the usual Gamma-function and $\mathfrak{G}$ is a complicated arithmetic function bounded between two positive constants.

It follows immediately from this asymptotic formula that every *sufficiently large integer* is a sum of at most s_1 kth powers, so *every* nonnegative integer is a sum of at most s_2 such powers for some s_2. The same analytic method, with considerable refinements over the years, has been used to settle numerous related problems; for example, Waring's claims about $g(3)$ and $g(4)$ have since been vindicated (the latter resisting attack until 1986!), and it is now known that $g(k) = 2^k + \lfloor (3/2)^k \rfloor - 2$ with at most finitely many exceptions (compare with Exercise 7.1.1). Complete treatments of the Hardy-Littlewood method can be found in Davenport's "blue book" [Dav63] and Vaughan's monograph [Vau97]. An excellent survey of developments up to 1971 is [Ell71].

In this chapter we content ourselves with a proof that $g(k)$ exists for every k. Our simple and short argument is due to D.J. Newman ([New60], [New98, Chapter V]). The key lemma is that for fixed k and large s,

$$r_{k,s}(n) \ll n^{s/k-1} \quad (n = 1, 2, \ldots). \tag{7.2}$$

As we shall shortly prove, this upper bound implies the set of integers expressible as a sum of s kth powers has positive lower density, so the proof of Waring's conjecture can be completed by an appeal to the results of Schnirelmann discussed in Chapter 5. Whereas Linnik [Lin43] proved (7.2) by elementary means, Newman proves (7.2) by stripping down the Hardy-Littlewood approach to the asymptotic formula. The resulting argument is one of the shortest known paths to Waring's conjecture.

7.1.1 Exercises

Exercise 7.1.1. Show that $2^k\lfloor(3/2)^k\rfloor-1$ is a sum of $2^k+\lfloor(3/2)^k\rfloor-2$ nonnegative kth powers and no fewer.

This result was first noted by Johannes Albert Euler (son of Leonhard) in 1772. Hint: be greedy!

Exercise 7.1.2 *(Liouville).* Verify the identity

$$\sum_{1\le i<j\le 4}(y_i+y_j)^4+\sum_{1\le i<j\le 4}(y_i-y_j)^4=6(y_1^2+y_2^2+y_3^2+y_4^2)^2.$$

a) Using Lagrange's four squares theorem, prove that every number of the form $6m^2$ is a sum of 12 nonnegative fourth powers.

b) By applying the four squares theorem again, show that every nonnegative multiple of 6 is a sum of 48 nonnegative fourth powers.

c) Conclude that every nonnegative integer is a sum of 53 nonnegative fourth powers.

Exercise 7.1.3. Assume Waring's conjecture. Show that if we define $h(k)$ as the smallest number of nonnegative kth powers needed to represent every nonnegative *rational* number, then $h(k)$ exists and satisfies $h(k)\le g(k)$.

The next few exercise present an analog of Waring's Problem for the polynomial ring $\mathbf{C}[t]$. We use the symbol Δ_j to denote the jth iterate of the forward difference operator, so that

$$\Delta_1(f;h_1)(t):=f(t+h_1)-f(t),$$
$$\Delta_{j+1}(f;h_1,\dots,h_{j+1}):=\Delta_1(\Delta_j(f;h_1,\dots,h_j),h_{j+1}).$$

Exercise 7.1.4. Let R be a commutative ring with identity and let k be a positive integer. Let $\mathcal{A}=\{\pm g(T)^k : g(T)\in R[T]\}$.

a) Show that if $f(T)$ can be represented as the sum of finitely many elements of $\mathcal{A}$, then the same is true of $\Delta_1(f;h_1)(T)$, for any $h_1\in R$.

b) Show that if T can be written as the sum of at most $s=s(k)$ elements of $\mathcal{A}$, then $\mathcal{A}$ is a basis of $R[T]$ of order at most s.

Exercise 7.1.5 *(continuation).*

a) Iterating part a) of the preceding exercise, show that for some $C\in R$, the polynomial $k!T+C$ can be written as a finite sum of elements of $\mathcal{A}$.

b) From part b) of the preceding exercise, deduce that if $R = K$ is a field of characteristic 0, then $\mathcal{A}$ is a basis of finite order. On the other hand, show that $\mathcal{A}$ is never a basis of finite order if $R = \mathbf{Z}$.

c) Assume that $R = K$ is an algebraically closed field of characteristic 0. Show that $\mathcal{A}$ is a basis of order at most k. Also show that $\mathcal{A}$ coincides with the set of kth powers in $K[T]$. Thus the analog of Waring's conjecture holds for $K[T]$.

The argument of the preceding exercise (in the special case $K = \mathbf{C}$) is due to S. Hurwitz. If we use $g(k, \mathbf{C}[T])$ to denote the order of the corresponding basis, then Hurwitz conjectures that $g(k, \mathbf{C}[T]) = k$ for each positive integer k. Newman & Slater [NS79] have shown $g(k, \mathbf{C}[T]) > k^{1/2}$ for each $k > 1$.

Exercise 7.1.6 *(Waring's Problem for Formal Power Series).* Let K be an algebraically closed field of characteristic 0. Show that for each $k > 1$, the set of kth powers is an additive basis of order 2 for $K[[T]]$.

Suggestion: First characterize the kth powers in $K[[T]]$.

7.2 The Linnik-Newman Approach

Waring's conjecture will be deduced from the following result, whose proof is deferred to the next section:

Fundamental Lemma. *Let $k \geq 2$. For all large s (depending only on k),*

$$r_{k,s}(n) \ll n^{s/k-1} \quad (n = 1, 2, \dots).$$

In particular, for s sufficiently large (depending on k),

$$r_{k,s}(n) \ll x^{s/k-1} \quad (1 \leq n \leq x). \tag{7.3}$$

The implied constants here depend only on s and k.

Remark. For fixed k and s, we can easily see that the *average* of $r_{k,s}(n)$ up to x is $O(x^{s/k-1})$:

$$\sum_{n \leq x} r_{k,s}(n) = |\{(x_1, \dots, x_s) \in \mathbf{N}^k : x_1^k + \dots + x_s^k \leq x\}| \leq (x^{1/k} + 1)^s.$$

The central difficulty is in proving that if s is large, none of the terms $r_{k,s}(n)$ are much bigger than average.

To see how the fundamental lemma is applied, suppose $\mathcal{A}$ is a set of natural numbers containing 0 with $\underline{d}(\mathcal{A}) > 0$. Define

$$r_{k,s}^{\mathcal{A}}(n) := \sum_{\substack{x_1^k + \dots + x_s^k = n \\ x_i \in \mathcal{A}}} 1,$$

so that $r^{\mathcal{A}}_{k,s} \le r_{k,s}$, with equality when $\mathcal{A}$ is the set of all natural numbers.

Let $k \ge 2$ and $s = s(k)$ be fixed integers for which (7.3) holds, and fix a positive number $d < \underline{d}(\mathcal{A})$. Then for large x,

$$\sum_{n \le x} r^{\mathcal{A}}_{k,s}(n) = |\{(x_1, \ldots, x_s) \in \mathcal{A}^k : x_1^k + \cdots + x_s^k \le x\}|$$
$$= |\{(x_1, \ldots, x_s) \in \mathcal{A}^k : x_i \le (x/s)^{1/k}\}| \ge (d(x/s)^{1/k})^s = \frac{d^s}{s^{s/k}} x^{s/k}.$$

But for $n \le x$, one has

$$r^{\mathcal{A}}_{k,s}(n) \le r_{k,s}(n) \ll x^{s/k-1},$$

so that there have to be

$$\gg \frac{d^s}{s^{s/k}} x^{s/k} \Big/ x^{s/k-1} \gg x$$

values of $n \le x$ with $r^{\mathcal{A}}_{k,s}(n) > 0$. In other words, if $\mathcal{B}$ denotes the set of kth powers of elements of $\mathcal{A}$, then the s-fold sumset $s\mathcal{B}$ has positive lower density. If $\mathcal{A}$ is primitive, in the sense that there is no $d > 1$ dividing all the elements of $\mathcal{A}$, then the same holds for $\mathcal{B}$ and hence for $s\mathcal{B}$. By Theorem 5.3.9, $s\mathcal{B}$ is an asymptotic basis of finite order s', say. But then $\mathcal{B}$ is an asymptotic basis of order $\le s's$.

We codify what we have proven in the following theorem:

Theorem 7.2.1 (Waring's Conjecture for Sets of Positive Density). *Let $\mathcal{A}$ be a set of natural numbers containing* 0 *with positive lower density and with* $\gcd(\mathcal{A}) = 1$. *Then for each positive integer k, the set $\mathcal{B}$ of kth powers of $\mathcal{A}$ is an asymptotic basis of finite order.*

Moreover, there is an s depending only on k (and not on $\mathcal{A}$) with the property that the sumset $s\mathcal{B}$ has positive lower density.

Strictly speaking, the argument above applies only to $k > 1$, but the case $k = 1$ is already contained in Theorem 5.3.9.

Examples.

i. We apply the first half of Theorem 7.2.1 with $\mathcal{A} = \mathbf{N}$ and find that the set of kth powers is an asymptotic basis of finite order for each $k = 1, 2, \ldots$. Since 0 and 1 are in $\mathcal{A}$, the set of nonnegative kth powers is an actual basis of finite order as well. We have therefore proved Waring's original conjecture.

ii. Reasoning similarly with $\mathcal{A}$ as the set of odd numbers, we find every positive integer is the sum of a bounded number of kth powers of odd integers.

iii. For result of a slightly different flavor, we take $\mathcal{A}$ as the set of positive integers expessible as a sum of two primes (as is permissible by Theorem 5.4.5); we find, e.g., that all large natural numbers are the sum of a bounded number of terms $(p + q)^7$, with p and q prime.

7.3 Proof of the Fundamental Lemma

Let N be a positive integer and let $f(\alpha) = \sum_{m=0}^{N} e(\alpha m^k)$. Then we noted in the introduction that

$$r_{k,s}(n) = \int_0^1 f(\alpha)^s e(-\alpha n)\, d\alpha$$

for the choice $N = \lfloor n^{1/k} \rfloor$.

Consequently, to prove the fundamental lemma it suffices to show that for fixed k and large s (depending on k),

$$\int_0^1 |f(\alpha)|^s \, d\alpha \ll_{s,k} N^{s-k}$$

for $N = 1, 2, \ldots$.

The rough idea in proving this is as follows: cover $[0,1]$ with small intervals centered at rational numbers and estimate the above integral over each of these. We'll show f is small on those intervals centered at rationals with large denominator; the key tool here is Weyl's estimates for certain exponential sums, a simplified form of which we present in the following subsection. The intervals centered at rationals with small denominators are divided up further, and a suitable approximation to f on these intervals is introduced. Crude estimates give what is required to complete the proof.

7.3.1 A Simplified Estimate of the Weyl Sums

Theorem 7.3.1 (Newman). *Let I be a set of at most $N \geq 1$ consecutive integers, and let $\epsilon > 0$. Let f be a polynomial of degree $k \geq 2$ with real coefficients and first coefficient a integral and prime to the positive integer $b \leq N$. Then*

$$S = S(f, b, I) := \sum_{n \in I} e\left(\frac{f(n)}{b}\right) \ll_{k,\epsilon} N^{1+\epsilon} b^{-1/2^{k-1}}, \tag{7.4}$$

where the implied constant depends only on k and ϵ.

We need a few lemmas. The first is an elementary bound on the rate of growth of the number of divisors function.

Lemma 7.3.2. *For every $\epsilon > 0$, we have $\tau(n) \ll_\epsilon n^\epsilon$ for $n = 1, 2, 3, \ldots$.*

Proof. Define an arithmetic function $f(n)$ by $f(n) := \tau(n)/n^\epsilon$. When $q = p^k$ is a prime power,

$$f(q) = f(p^k) = \frac{k+1}{p^{k\epsilon}} \leq \frac{\log 2q/\log 2}{q^\epsilon} < 1$$

as soon as $q > q_0 = q_0(\epsilon)$. It follows that the set S of prime powers q for which $f(q) > 1$ is finite. Since f is multiplicative,

$$\frac{\tau(n)}{n^\epsilon} = f(n) = \prod_{p^k \| n} f(p^k) \leq C := \prod_{q \in S} f(q) < \infty$$

for every $n = 1, 2, 3, \ldots$. □

The next result is a well-known inequality whose proof is included for completeness' sake:

Lemma 7.3.3. *For $-\pi/2 \le \theta \le \pi/2$, we have*

$$|\sin(\theta)| \ge 2|\theta|/\pi. \tag{7.5}$$

Proof. We may assume $\theta \ge 0$, since both sides of (7.5) represent even functions. Equality holds in (7.5) for $\theta = 0$ and $\theta = \pi/2$, so we may further assume $0 < \theta < \pi/2$. For such θ, the inequality $\tan(\theta) \ge \theta$ holds. Indeed, $f(\theta) := \tan(\theta) - \theta$ satisfies $f(0) = 0$, and $f'(\theta) = \sec^2(\theta) - 1 = \tan^2(\theta) \ge 0$ for $\theta \in [0, \pi/2)$, so that f is nondecreasing on this interval.

Now consider the function $g(\theta) = \sin(\theta)/\theta$. Then $g(\pi/2) = 2/\pi$, and

$$g'(\theta) = \frac{\theta\cos(\theta) - \sin(\theta)}{\theta^2}.$$

Since $\tan(\theta) \ge \theta$ for $0 < \theta < \pi/2$, the numerator here is nonpositive for $0 < \theta < \pi/2$. Hence g is nonincreasing on the interval $(0, \pi/2)$. Since g is continuous at $\theta = \pi/2$, it follows that for any $0 < \theta < \pi/2$,

$$g(\theta) = \sin(\theta)/\theta \ge g(\pi/2) = 2/\pi,$$

as was to be shown. □

We need the next lemma to jumpstart the inductive proof of Theorem 7.3.1:

Lemma 7.3.4. *Let α, β be real numbers. For integers r, Q with $Q \ge 0$,*

$$\left|\sum\nolimits_{r+1\le m\le r+Q} e(\alpha m + \beta)\right| \le \min\{Q, (2\|\alpha\|)^{-1}\} \le \min\{Q, \|\alpha\|^{-1}\}. \tag{7.6}$$

Proof. The case $Q = 0$ is clear, so we suppose that Q is positive. Since $e(\beta + (r+1)\alpha)$ is on the unit circle, we may also assume $\beta = 0$ and $r = -1$, so that the sum runs from 0 to $Q - 1$. Finally, shifting α by an integer we may assume $|\alpha| \le 1/2$, so that $\|\alpha\| = |\alpha|$.

By the triangle inequality, the left hand side is at most Q in absolute value; moreover, equality holds when $\|\alpha\| = 0$. When α is non-integral, summing the geometric series shows the left hand side is

$$\left|\frac{e(Q\alpha) - 1}{e(\alpha) - 1}\right| \le \frac{2}{|e(\alpha) - 1|} = \frac{2}{|e(\alpha/2) - e(-\alpha/2)|} = \frac{2}{|\sin(\pi\alpha)|}.$$

Since $|\pi\alpha| \le \pi/2$, Lemma 7.3.3 implies

$$|\sin(\pi\alpha)| \ge 2|\alpha| = 2\|\alpha\|.$$

This gives the first inequality in (7.6); the second is immediate from the observation that the function $\min\{X, Y\}$ is nondecreasing in both arguments. □

Proof of Theorem 7.3.1. The proof is by induction on k and is based on the identity

$$\begin{aligned}|S|^2 &= \sum_{n,m\in I} e\left(\frac{f(n)-f(m)}{b}\right) \\ &= |I| + 2\Re \sum_{\substack{n,m\in I\\ n>m}} e\left(\frac{f(n)-f(m)}{b}\right).\end{aligned}$$

Reorganizing the sum according to the value of $n-m \leq N$ and recalling that $|I| \leq N$, we find

$$|S|^2 \ll N + \sum_{d=1}^{N} \left| \sum_{m\in I\cap I-d} e\left(\frac{f(m+d)-f(m)}{b}\right)\right|. \tag{7.7}$$

What is of interest here is that the inner sum is of the same form as S, with f replaced by $\Delta(f;d)$ of degree one less than f, and I replaced by $I\cap(I-d)$, which is again a set of at most N consecutive integers. This allows for an inductive argument.

Our base case is when $k=2$. Then

$$\frac{f(m+d)-f(m)}{b}$$

is a linear polynomial with leading coefficient $\alpha := 2ad/b$, so that (7.6) and (7.7) together imply

$$|S|^2 \ll N + \sum_{d=1}^{N} \max\{N, \|2ad/b|^{-1}\} \ll N + \sum_{d=1}^{2N} \max\{N, \|ad/b\|^{-1}\}.$$

To estimate this final sum we break it into blocks of b consecutive integers, allowing the final block to perhaps be incomplete. Since $\gcd(a,b)=1$, the sum over each block is

$$\ll N + \sum_{d'=1}^{b} \|d'/b\|^{-1} \ll N + \sum_{d'\leq b/2} \frac{b}{d'} \ll N + b\log b.$$

Now $b \leq N$, so that $\log b \leq \log N \ll_\epsilon N^{2\epsilon}$; since there are $\ll 2N/b + 1 \ll N/b$ such blocks, we get (keeping in mind that $b \leq N$)

$$\begin{aligned}|S|^2 &\ll N + Nb^{-1}\left(N + bN^{2\epsilon}\right) \\ &\ll N + N^2b^{-1} + N^{1+2\epsilon} \ll N^2b^{-1} + N^{1+2\epsilon} \ll N^{2+2\epsilon}b^{-1},\end{aligned}$$

Taking the square root yields (7.4) for $k=2$.

Now suppose $k \geq 3$ and that the lemma is known to hold for $k-1$. We apply the induction hypothesis to the inner sum in (7.7). To circumvent the difficulty that the leading coefficient kad of $f(m+d)-f(m)$ might not be prime to b, we write

$$\frac{kad}{b} = \frac{a'}{b'},$$

in lowest terms. Then $b' = b'(d)$ divides b (so is $\leq N$).

Moreover, the number of $d \in [1, N]$ with a given value of b' is $\ll_k Nb/b'$. To prove this last claim, suppose b' is the denominator of kad/b in reduced terms. Then b/b' divides kad, and as $\gcd(a,b) = 1$, we must have b/b' divides kd. But there are only

$$\ll \frac{kN}{b/b'} + 1 \ll_k Nb'/b + 1 \ll Nb'/b$$

integers $\leq kN$ divisible by b/b', and the assertion follows.

Therefore, by the induction hypothesis,

$$\begin{aligned} |S|^2 &\ll N + \sum_{b'|b} N\frac{b'}{b}N^{1+\epsilon}b'^{-1/2^{k-2}} \\ &\ll N + \frac{1}{b}N^{2+\epsilon}\sum_{b'|b} b'^{1-1/2^{k-2}} \ll N + \frac{1}{b}N^{2+\epsilon}b^{1-1/2^{k-2}}\tau(b). \end{aligned}$$

Since $\tau(b) \ll_\epsilon b^\epsilon \leq N^\epsilon$, we get the estimate

$$|S|^2 \ll N + N^{2+2\epsilon}b^{-1/2^{k-2}} \ll N^{2+2\epsilon}b^{-1/2^{k-2}}.$$

Again, taking square roots gives the result. □

Remark. Newman includes the case $k = 1$ in his statement of the lemma, using it as his "obvious" base case. But the lemma is false for $k = 1$ (exercise!), and this necessitates a more complicated argument than that which he presents in [New98, Chapter V].

7.3.2 Completion of the Proof

Let $k \geq 2$ be fixed. Set $\nu = 1/3$.

For $0 \leq a \leq q \leq N^{k-\nu}$, $\gcd(a,q) = 1$ and $j \geq 0$, define

$$\mathcal{I}_j(q,a) := \left\{\alpha : \left|\alpha - \frac{a}{q}\right| \leq \frac{1}{qN^{k-\nu}}; \frac{j}{N^k} \leq \left|\alpha - \frac{a}{q}\right| < \frac{j+1}{N^k}\right\}.$$

By Dirichlet's approximation lemma (Lemma 6.7.2),

$$[0,1] \subset \bigcup_{a,q} \cup_j \mathcal{I}_j(q,a). \tag{7.8}$$

Note that for fixed a, q,

$$\mathcal{I}_j(q,a) = \emptyset \quad (j > N^\nu); \tag{7.9}$$

hence only finitely many of the $\mathcal{I}_j(q,a)$ are nonempty.

Lemma 7.3.5. *For $\alpha \in \mathcal{I}_j(q,a)$,*

$$f(\alpha) \ll \frac{N}{(q+j)^{1/2^k}},$$

where the implied constant depends only on k.

Proof. We distinguish two cases according as q is large or small. Suppose first q is large, say $q > N^{2\nu}$. Then if $\alpha \in I_j(q,a)$, we have (applying Theorem 7.3.1 with $\epsilon = \nu/2^k$)

$$\begin{aligned} f(a/q) = 1 + \sum_{m=1}^{N} e(am^k/q) &\ll 1 + N^{1+\nu/2^k} q^{-1/2^{k-1}} \ll N \cdot N^{\nu/2^k} q^{-1/2^{k-1}} \\ &\ll N \cdot q^{1/2^{k+1}} q^{-1/2^{k-1}} \ll N q^{-1/2^k}. \end{aligned} \tag{7.10}$$

Also, since $|f'|$ is bounded by $2\pi N^{k+1}$,

$$\begin{aligned} |f(\alpha) - f(a/q)| &= \left| \int_{a/q}^{\alpha} f'(t)\, dt \right| \le 2\pi N^{k+1} |\alpha - a/q| \\ &\ll N^{1+\nu} q^{-1} = N \cdot N^{\nu} q^{-1} \ll N q^{-1/2} \ll N q^{-1/2^k}. \end{aligned} \tag{7.11}$$

Combining (7.10) and (7.11) gives the result in this case, since $q > N^{2\nu} \ge N^{\nu}$ implies $j = 0$.

Now suppose q is small, i.e., $q \le N^{2\nu}$. Let $A := q^{-1} \sum_{m=1}^{q} e(am^k/q)$, so that $|A| \le 1$ trivially, and

$$A \ll q^{-1} q^{1+1/2^k} q^{-1/2^{k-1}} = q^{-1/2^k}. \tag{7.12}$$

by Theorem 7.3.1 (applied with $\epsilon = 1/2^k$). Write

$$f = 1 + \sum_{m=1}^{N} e(am^k/q) = 1 + S_1 + AS_2,$$

where

$$\begin{aligned} S_1 &:= \sum\nolimits_{m=1}^{N} \bigl(e(am^k/q) - A\bigr)\, e((\alpha - a/q)m^k), \\ S_2 &:= \sum\nolimits_{m=1}^{N} e((\alpha - a/q)m^k). \end{aligned}$$

S_1 takes the form $\sum a_m g(m)$, where

$$a_m = e(am^k/q) - A, \quad g(m) = e((\alpha - a/q)m^k).$$

Now A was selected so that the sum of the a_m vanishes when taken over any block of q consecutive integers; since each a_m is bounded by 2, we obtain a uniform bound on the partial sums:

$$\left| \sum\nolimits_{m=1}^{N} a_m \right| = \left| \sum\nolimits_{m=q[N/q]+1}^{N} \bigl(e(am^k/q) - A\bigr) \right| \le 2q.$$

We also have the bound on the variation

$$V := \int_0^N |g'(t)|\,dt = \int_0^N 2\pi|\alpha - a/q|kt^{k-1}\,dt = 2\pi|\alpha - a/q|N^k \leq 2\pi N^\nu/q$$

and we know that the maximum of $|g|$ on $[0, N]$ is 1. By Appendix A, Corollary A.2.4 we deduce

$$|S_1| \leq (2q)(2\pi N^\nu/q + 1) \ll N^\nu + q \ll N^{2\nu}$$

while by Appendix A, Corollary A.2.2,

$$S_2 = \int_0^N g(t)\,dt + O(V) = v(\alpha - a/q) + O(N^\nu/q),$$

where

$$v(\beta) := \int_0^N e(\beta u^k)\,du.$$

Consequently,

$$\begin{aligned} f(\alpha) &= AS_2 + S_1 + 1 \\ &= Av(\alpha - a/q) + O(|A|N^\nu/q) + O(N^{2\nu}) + O(1) \\ &= Av(\alpha - a/q) + O(N^{2\nu}). \end{aligned}$$

In particular, referring to (7.12) shows

$$\begin{aligned} f(\alpha) &\ll |A||v(\alpha - a/q)| + N^{2\nu} \\ &\ll q^{-1/2^k}|v(\alpha - a/q)| + N^{2\nu}. \end{aligned}$$

To complete the proof we need to estimate $v(\alpha - a/q)$. Trivially $v(\alpha - a/q) \leq N$. To get a sharper estimate when $|\alpha - a/q|$ is large, we change variables to find

$$\begin{aligned} v(\alpha - a/q) &= |\alpha - a/q|^{-1/k} \int_0^{N/|\alpha - a/q|^{1/k}} e(u^k)\,du \\ &\ll |\alpha - a/q|^{-1/k} = N(N^k|\alpha - a/q|)^{-1/k} \ll Nj^{-1/k}, \end{aligned}$$

where we have used the convergence of $\int_0^\infty e(u^k)\,du$. Of course this last estimate is only useful (or sensible!) if $j \geq 1$, but if we combine it with the trivial estimate we find

$$|v(\alpha - a/q)| \ll N(j+1)^{-1/k} \ll N(j+1)^{-1/2^k}.$$

Therefore,

$$\begin{aligned} f(\alpha) &\ll Nq^{-1/2^k}(j+1)^{-1/2^k} + N^{2\nu} \\ &\ll N(q+j)^{-1/2^k} + N(q+j)^{-1/2} \ll N(q+j)^{-1/2^k}; \end{aligned}$$

in transitioning between lines we have used that $q, j \ll N^{2\nu}$ (by (7.9)) and $1 - \nu \geq 2\nu$. $\square$

Proof of the Fundamental Lemma. Let $s \geq 2^{k+2}$. Since $\mathcal{I}_j(q,a)$ is contained in a union of two intervals of total length $2N^{-k}$,

$$\int_{\mathcal{I}_j(q,a)} |f(\alpha)|^s \, d\alpha \ll \frac{N^s}{(q+j)^4} N^{-k} \ll \frac{N^{s-k}}{(q+j)^4}.$$

By (7.8) and (7.9),

$$\begin{aligned}\int_0^1 |f(\alpha)|^s \, d\alpha &\leq \sum_{q \leq N^{k-\nu}} \sum_{\substack{0 \leq a \leq q \\ \gcd(a,q)=1}} \sum_{0 \leq j \leq N^\nu} \int_{\mathcal{I}_j(q,a)} |f(\alpha)|^s \, d\alpha \\ &\ll N^{s-k} \sum_{q \leq N^{k-\nu}} \sum_{0 \leq j \leq N^\nu} \frac{q}{(q+j)^4} \ll N^{s-k} \sum_{\substack{q \geq 1 \\ j \geq 0}} \frac{1}{(q+j)^3} \ll N^{s-k},\end{aligned}$$

since $\sum_{q \geq 1} \sum_{j \geq 0} (q+j)^{-3}$ converges (the inner sum being $O(q^{-2})$). The remarks at the beginning of this section suffice to complete the proof. □

7.3.3 Exercises

The following two exercises are based on observations appearing in [New98].

In this section we proved that for $k \geq 2$ and $s > s(k)$,

$$\int_0^1 |f(\alpha)|^s \; d\alpha \ll_{s,k} N^{s-k} \quad (N = 1, 2, \ldots). \tag{7.13}$$

Exercise 7.3.1.

a) Noting that $|f'| \leq 2\pi N^{k+1}$, deduce from $f(0) = N+1$ that $|f(\alpha)| \geq N/2$ for $0 \leq \alpha \leq (4\pi N^k)^{-1}$.

b) Prove that for any $k, s \geq 1$,

$$\int_0^1 |f(\alpha)|^s \, d\alpha \geq (2^{s+2}\pi)^{-1} N^{s-k},$$

so that the estimate of this section is best possible.

Exercise 7.3.2. Suppose that for some $k, s \geq 1$, one has

$$\int_0^1 |f(\alpha)|^s \; d\alpha \leq C N^{s-k} \quad (N = 1, 2, \ldots).$$

Prove that the same inequality holds (with the same C) with s replaced by any $s' \geq s$.

References

[Dav63] H. Davenport, *Analytic methods for Diophantine equations and Diophantine inequalities*, The University of Michigan, Fall Semester, vol. 1962, Ann Arbor Publishers, Ann Arbor, Mich., 1963. MR 28 #3002

[Ell71] W. J. Ellison, *Waring's problem*, Amer. Math. Monthly **78** (1971), no. 1, 10–36. MR 54 #2611

[Guz03] Alberto Guzman, *Derivatives and integrals of multivariable functions*, Birkhäuser Boston Inc., Boston, MA, 2003. MR 1 997 439

[Hil09] D. Hilbert, *Beweis für die Darstellbarkeit der ganzen Zahlen durch eine feste Anzahl n-ter Potenzen (Waringsches Problem). Dem Andenken an Hermann Minkowski gewidmet*, Math. Annalen **67** (1909), 281–300.

[HL20] G.H. Hardy and J.E. Littlewood, *Some problems of 'Partitio Numerorum'; I: A new solution of Waring's problem.*, Gött. Nachr. (1920), 33–54.

[Lin43] U. V. Linnik, *An elementary solution of the problem of Waring by Schnirelman's method*, Rec. Math. [Mat. Sbornik] N.S. **12(54)** (1943), 225–230. MR 5,200b

[New60] Donald J. Newman, *A simplified proof of Waring's conjecture*, Michigan Math. J **7** (1960), 291–295. MR 22 #10967

[New98] ———, *Analytic number theory*, Graduate Texts in Mathematics, vol. 177, Springer-Verlag, New York, 1998. MR 98m:11001

[NS79] D. J. Newman and Morton Slater, *Waring's problem for the ring of polynomials*, J. Number Theory **11** (1979), no. 4, 477–487. MR 80m:10016

[Vau97] R. C. Vaughan, *The Hardy-Littlewood method*, second ed., Cambridge Tracts in Mathematics, vol. 125, Cambridge University Press, Cambridge, 1997. MR 98a:11133

Part III

Appendices

Appendix A

Some Notions from Asymptotics

In this chapter we survey results from asymptotics needed in the text. The material of this chapter (especially of the first two sections) is standard and can be found in the opening chapters of any analytic number theory text (e.g., see [Ten95, Chapter 0], [Mur01, Chapter 2] or [Apo76, Chapters 3, 4]).

Another helpful but unpublished reference is [Hil02].

A.1 Big Oh, little oh, and their relatives

A.1.1 Big-Oh notation

Suppose f and g are complex-valued functions of one or several variables. We write $f = O(g)$ if $|f| \leq C|g|$ for some constant C and all specified values of the variables; the constant C is referred to as the *implied constant*. For example, $x = O(x^2)$ for $x \geq 1$ (we may take $C = 1$), but *not* in the larger range $x \geq 0$ (consider values of x approaching 0).[1]

The notation $f(x) = O(g(x))$ as $x \to a$ means that $f = O(g)$ on some deleted neighborhood U of a, with respect to the topology of the extended real line $\mathbf{R} \cup \{\pm\infty\}$. In particular, when $a = \infty$, this condition reduces to the statement that $f = O(g)$ on some set U of the form (x_0, ∞). Similar conventions apply (with left and right deleted neighborhoods, respectively) if we replace "$x \to a$" by "$x \uparrow a$" (x approaches a from below) or "$x \downarrow a$" (x approaches a from above).

Instead of letting x tend to a through all values, as was the case in the previous paragraph, we sometimes restrict the values of x to a prescribed subset

[1] The following quote, drawn from C.A. Roger's preface to [Dav77], illustrates the formidable utility of this notation: "Davenport, Erdős, Ko, Mahler and Žilinskas found time to play regular bridge. Mahler, who had only recently taken up the game, was prone to miss the best play, and the others and Mahler himself were soon describing poor play as being O.M. (or more correctly $O(M)$). Mahler long remained unaware that the Landau notation was in use and that this stood for 'Order of Mahler."'

S. In order for x to tend to a along values of S, it must be that a is in the closure of S. In this case, we define $f(x) = O(g(x))$ as $x \to a$ along values $x \in S$ to mean that $f(x) = O(g(x))$ for some set U, where U is the intersection of S with a deleted neighborhood of a. A common such situation is when S is an infinite set of positive integers and $a = \infty$; in this case, the assertion $f = O(g)$ unwinds to the assertion that there exist (finite) constants C, N_0 for which $|f| \leq C|g|$ for $N \geq N_0, N \in S$.

Finally, if f and g depend on a parameter λ, then we write $f = O_\lambda(g)$ to mean that C (as well as the neighborhood U, in the situations of the last two paragraphs) is allowed to depend on λ. If C (and the neighborhood U) can be chosen independently of λ, then we say that the estimate holds *uniformly in* λ. For example, the function $f_n(x, y) = n(x^2 + y^2)$ is $O_n(1)$ for $x, y \in [0, 1]$, while the function $g_n(x, y) = \sin(n(x^2 + y^2))$ is $O(1)$ uniformly in (real) n for x, y in the same range.

If f and g depend on λ, but in stating a big-Oh estimate we do not specify either dependence on λ or uniformity in λ, then the former is assumed by default. That is, we always assume the weaker statement is being made.

The notation $f \ll g$ is sometimes used in place of $f = O(g)$. We write $f \gg g$ to mean $g \ll f$, and $f \asymp g$ to mean $f \ll g$ and $g \ll f$; in the latter case we say that g gives the order of magnitude of f, or that f and g have the same order of magnitude. Subscripts used with any of these symbols denote dependence on parameters, as above.

A.1.2 Little-oh notation

We say that $f = o(g)$ as $x \to a$ if

$$\lim_{x \to a} f(x)/g(x) = 0,$$

i.e., if for every $\epsilon > 0$ there is a deleted neighborhood U of a such that $|f(x)/g(x)| < \epsilon$ for every $x \in U$. If f, g depend on a parameter λ, then we say $f = o(g)$ *uniformly in* λ if for every ϵ, the set U can be chosen independently of λ. For example, $x = o(x^2)$ as $x \to \infty$, and $\sin(nx) = o(x)$ as $x \to \infty$, uniformly in n.

If the values of x tending to a are restricted to lie in the set S (where we assume a belongs to the closure of S), then the definitions are the same as above, except that U is now the intersection with S of a deleted neighborhood of a.

We say "f is asymptotic to g as x tends to a," and write $f \sim g$ as $x \to a$, if $\lim_{x \to a} f(x)/g(x) = 1$. Most commonly $a = \infty$, which we assume if no a is specified. Note that $f \sim g$ is equivalent to $f = g(1 + o(1))$, provided g is nonvanishing in a deleted neighborhood of a.

A.2 Estimation of Sums

A.2.1 Comparison of a Sum and an Integral

If f does not oscillate too wildly, it is reasonable to expect the Riemann sum $\sum_{n=a+1}^{b} f(n)$ to approximate the integral $\int_a^b f(t)\,dt$. Our next theorem gives us an expression for the error in this approximation.

Theorem A.2.1 (Euler's Summation Formula). *Let $a < b$ be integers and suppose $f\colon [a,b] \to \mathbf{C}$ is continuously differentiable. Then*

$$\sum_{n=a+1}^{b} f(n) = \int_a^b f(t)\,dt + \int_a^b \{t\} f'(t)\,dt.$$

Proof. We begin with the identity

$$g(1) = \int_0^1 g(t)\,dt + \int_0^1 t g'(t)\,dt,$$

which (integrating by parts) holds whenever g is continuously differentiable on $[0,1]$. Applying this identity with $g(t) = f(n-1+t)$ shows

$$\begin{aligned} f(n) &= \int_0^1 f(n-1+t)\,dt + \int_0^1 t f'(n-1+t)\,dt \\ &= \int_{n-1}^{n} f(t)\,dt + \int_{n-1}^{n} \{t\} f'(t)\,dt. \end{aligned}$$

Summing over $n = a+1, \dots, b$ yields the result. $\square$

Example. We take $f(n) = 1/n$, $a = 1$, and $b = N$ in Theorem A.2.1 to obtain

$$\begin{aligned} \sum_{n \le N} \frac{1}{n} &= 1 + \int_1^N \frac{dt}{t} + \int_1^N -\frac{\{t\}}{t^2}\,dt \\ &= \log N + 1 - \int_1^N \frac{\{t\}}{t^2}\,dt, \end{aligned}$$

the computation being valid for positive integers $N \ge 2$. Because $\{t\}/t^2 \ll t^{-2}$, this last integral converges when extended to infinity, so that

$$\sum_{n \le N} \frac{1}{n} = \log N + 1 - C + O\left(\int_N^\infty \frac{dt}{t^2}\right) = \log N + \gamma + O(1/N),$$

where $\gamma := 1 - C$. The same estimate holds if the limit of summation N is replaced by an arbitrary real number $x \ge 2$; this follows from what we've just done by taking $N = \lfloor x \rfloor$ and noticing that

$$\log\lfloor x \rfloor = \log x + \log\left(1 - \frac{\{x\}}{x}\right) = \log x + O(1/x).$$

The following crude form of Euler's summation formula often suffices in applications:

Corollary A.2.2. *Let $a < b$ be integers and suppose $f\colon [a,b] \to \mathbf{C}$ is continuously differentiable. Then*

$$\sum_{n=a+1}^{b} f(n) = \int_a^b f(t)\,dt + O\left(\int_a^b |f'(t)|\,dt\right).$$

A.2.2 Partial Summation

In number theory one often needs estimates for sums of the form $\sum_{y<n\le x} a_n b_n$ in situations where one already has estimates for $S(t) := \sum_{y<n\le t} a_n$.

As was observed by Abel, if y, x are integers (with $y \le x$) then we can rewrite the desired sum as

$$\begin{aligned}\sum_{n=y+1}^{x} a_n b_n &= \sum_{n=y+1}^{x} (S(n) - S(n-1))b_n \\ &= \sum_{n=y+1}^{x} S(n)b_n - \sum_{n=y+1}^{x-1} S(n)b_{n+1} \\ &= \sum_{n=y+1}^{x-1} S(n)(b_n - b_{n+1}) + S(x)b_x. \end{aligned} \tag{A.1}$$

When the values of b_n are given by a smooth function f, there is a more telling way of reexpressing this sum completely analogous to the usual formula for integration by parts. Before stating this result, we introduce a bit of further notation: If f and c are complex-valued functions defined on $[a,b]$, with f' piecewise continuous here, we define

$$\int_a^b f(t)\,dc(t) = c(b)f(b) - c(a)f(a) - \int_a^b f'(t)c(t)\,dt, \tag{A.2}$$

whenever the right-hand integral exists. This last condition is certainly satisfied if c is piecewise continuous on the interval or (a weaker condition) if c is of bounded variation. Note that if $c(t) = t$, then the right hand side of (A.2) is the same as that appearing in the formula for integration by parts, so that this agrees with the usual definition of $\int f(t)\,dt$.

Just as with ordinary integrals, it is convenient to extend the notation to allow $b = \infty$ by setting

$$\int_a^\infty f(t)\,dc(t) = \lim_{b\to\infty} \int_a^b f(t)\,dc(t).$$

Theorem A.2.3 (Partial summation). *Suppose $y \le x$ are real numbers and that complex numbers a_n are defined for all $n \in (y, x]$. Define $S(t) = \sum_{y<n\le t} a_n$. If f' is piecewise continuous on $[y, x]$, then*

$$\sum_{y<n\le x} a_n f(n) = \int_x^y f(t)\, dS(t).$$

Proof. We have

$$\begin{aligned}
\int_y^x f(t)\, dS(t) &= S(x)f(x) - \int_y^x \left(\sum_{y<n\le t} a_n \right) f'(t)\, dt \\
&= S(x)f(x) - \sum_{y<n\le x} a_n \int_n^x f'(t)\, dt \\
&= S(x)f(x) - \left(f(x) \sum_{y<n\le x} a_n - \sum_{y<n\le x} a_n f(n) \right) \\
&= \sum_{y<n\le x} a_n f(n). \qquad \square
\end{aligned}$$

Example. Let $\mathcal{A}$ be a set of positive integers with density 1/2, so that $A(x) = x(1/2 + o(1))$, and suppose we wish to estimate $\sum_{n\le x, n\in\mathcal{A}} 1/a$. Take a_n to be the characteristic function of $\mathcal{A}$; then $S(t) = A(t)$, and

$$\begin{aligned}
\sum_{a\le x, a\in\mathcal{A}} \frac{1}{a} = \sum_{n\le x} a_n \frac{1}{n} &= \int_{1/2}^x \frac{1}{t}\, dA(t) \\
&= \frac{A(t)}{t} - \frac{A(1/2)}{1/2} - \int_{1/2}^x \frac{-A(t)}{t^2}\, dt \\
&= (1/2 + o(1)) + \int_1^x \left(\frac{1}{2t} + o\left(\frac{1}{t}\right) \right) dt \qquad \text{(A.3)} \\
&= O(1) + \frac{1}{2}\log x + o(\log x) = (1/2 + o(1)) \log x. \qquad \text{(A.4)}
\end{aligned}$$

The reader should offer a careful justification of the transition from (A.3) to (A.4), keeping in mind that one cannot in general pull o's outside an integral.

The following corollary of Theorem A.2.3 will be required in Chapter 7:

Corollary A.2.4 (Newman [New60]). *Let N be a positive integer, f a continuously differentiable complex-valued function on $[0, N]$ and $a_1, \ldots, a_N$ any complex numbers. Let*

$$M := \max_{m=1,2,\ldots,N} \left| \sum_{n=1}^m a_n \right|, \quad M' := \max_{x\in[0,N]} |f(t)|, \quad V := \int_0^N |f'(t)|\, dt.$$

Then

$$\left|\sum_{n=1}^{N} a_n f(n)\right| \le M(V + M').$$

Proof. Let $S(t) = \sum_{n \le t} a_n$. Then

$$\sum_{n=1}^{N} a_n f(n) = \int_{1/2}^{N} f(t)\, dS(t) = S(N)f(N) - \int_{1}^{N} f'(t)S(t)\, dt$$
$$\ll MM' + MV = M(V + M'). \qquad \square$$

Those with some background in analysis will recognize that V coincides (under our hypotheses on f) with the "total variation" of f on $[0, N]$.

A.3 Counting Lattice Points

A.3.1 . . . in Homothetically Expanding Regions

Let X be a bounded subset of $\mathbf{R}^n$. Then the number $N(X)$ of lattice points contained in X is a Riemann sum for the (Jordan)-volume

$$\mathrm{vol}(X) := \int_{\mathbf{R}^n} \chi_X \, dV = \int_X 1\, dV.$$

Thus, letting $a \times X$ denote the dilation of X by a, it is reasonable to expect the count of lattice points contained in $a \times X$ to be approximately $\mathrm{vol}(a \times X) = a^n \mathrm{vol}(X)$. The next theorem says that for fixed X, this is true asymptotically as $a \to \infty$.

Theorem A.3.1. *Let X be a bounded subset of $\mathbf{R}^n$ which possesses a Jordan-volume $\mathrm{vol}(X)$, and let $N(a \times X)$ denote the number of lattice points contained in the dilation $a \times X$. Then $N(a \times X)/a^n \to \mathrm{vol}(X)$ as $a \to \infty$.*

The same holds if $\mathrm{vol}(X)$ is interpreted as n-dimensional Lebesgue measure, provided the boundary of X has measure zero.

Proof for the Lebesgue measure (sketch). It is enough to show $N(a_j \times X)/a_j^n \to \mathrm{vol}(X)$ holds for every given sequence $\{a_j\}_{j \ge 1}$ of positive numbers with $a_j \to \infty$.

Instead of letting X expand by the factor a_j, we think of contracting the usual lattice by the same factor, so that $\mathbf{R}^n$ is partitioned into n-dimensional cubes of volume a_j^{-n}. We define $f_j = 1$ on those cubes where the corner with smallest coordinates (the analog of the southwest corner in $\mathbf{R}^2$) belongs to X and $f_j = 0$ on the other cubes. More precisely (to fix what we do about the boundaries of the cubes), using the notation $\lfloor \mathbf{x} \rfloor = (\lfloor x_1 \rfloor, \ldots, \lfloor x_n \rfloor)$ for $\mathbf{x} \in \mathbf{R}^n$, we define

$$f_j(\mathbf{x}) := \begin{cases} 1 & \text{if } \lfloor a_j \mathbf{x} \rfloor \in a_j X, \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$N(a_j \times X)/a_j^n = \int_{\mathbf{R}^n} f_j.$$

Moreover, $f_j(\mathbf{x}) \to \chi_X(\mathbf{x})$ for every $\mathbf{x}$ not on the boundary of X, so for almost every $\mathbf{x} \in \mathbf{R}^n$. By the bounded convergence theorem,

$$\lim_{j\to\infty} N(a_j \times X)/a_j^n = \lim_{j\to\infty} \int_{\mathbf{R}^n} f_j = \int_{\mathbf{R}^n} \chi_X = \mathrm{vol}(X). \qquad \square$$

Proof for the Riemann integral (sketch). Proceeding as above, we interpret $N(a\times X)/a^n$ as a Riemann sum for $\mathrm{vol}(X)$, where the corresponding partition is into n-dimensional cubes of side length a^{-1}. As $a \to \infty$, the mesh of the partition tends to 0, so the Riemann sum tends to the integral (by [Guz03, §4.2, Theorem 4]). □

For example, taking X as the unit circle implies there are $\pi r^2 + o(r^2)$ lattice points in the circle of radius r centered at the origin. For regions like this, which are bounded by simple closed rectifiable plane curves, better results are provided by our next theorem.

A.3.2 . . . enclosed by a Jordan curve

Theorem A.3.2. *Let J be a simple, rectifiable closed plane curve of length $L \geq 1$ enclosing a region of area (Lebesgue measure) A. Then the number N of lattice points inside or on J satisfies $N = A + O(L)$, with an absolute implied constant.*

Remark. Jarnik showed that under these conditions $|N - A| < L$, and a short proof was later published by Steinhaus [Ste47]; his argument is reproduced in [Hua82, Chapter 6, Theorem 9.2]. The proof given below of the weaker Theorem A.3.2 is taken from [Lan69, Teil 8, Einleitung].

Proof. Let R denote the region enclosed by J, including the boundary. Let M be the number of (closed) squares (of the usual lattice) containing a point of the boundary curve J. We may interpret N, as defined in the theorem statement, as the number of lattice squares whose southwest corner belongs to R.

We first claim $N - A \leq M$: to see this, consider the squares whose southwest corner belongs to R. These are of two types: those which lie entirely in R and those which contain at least one point outside of R. There are at most A of the former, since each square has area 1, and at most M of the latter, since any square intersecting both R and its complement must intersect J.

On the other hand, we also have $A - N \leq M$: if we think of covering R with lattice squares, it is obvious that A is bounded by the number of squares that intersect R. Such a square either has southwest corner belonging to R or outside of R; in the former case it is counted by N and in the latter by M.

Consequently,

$$|N - A| \leq M. \tag{A.5}$$

It remains to estimate M. For this, note that a curve of length < 1 can intersect at most 4 squares, since among any five squares there are two at least distance 1 apart. Breaking J into $\lfloor L \rfloor + 1$ pieces of length $L/(\lfloor L \rfloor + 1) < 1$, we see

$$M \leq 4(\lfloor L \rfloor + 1) \leq 4(L+1) \leq 8L. \tag{A.6}$$

The result now follows from (A.5) and (A.6). $\square$

We immediately obtain an improvement on our earlier estimate:

Corollary A.3.3. *The number of lattice points inside or on any circle of radius $r \geq 1$ in the plane is $\pi r^2 + O(r)$, where the implied constant is absolute.*

References

[Apo76] Tom M. Apostol, *Introduction to analytic number theory*, Springer-Verlag, New York, 1976, Undergraduate Texts in Mathematics. MR 55 #7892

[Dav77] Harold Davenport, *The collected works of Harold Davenport. Vol. I*, Academic Press [Harcourt Brace Jovanovich Publishers], London, 1977, Edited by B. J. Birch, H. Halberstam and C. A. Rogers. MR 58 #27141a

[Guz03] Alberto Guzman, *Derivatives and integrals of multivariable functions*, Birkhäuser Boston Inc., Boston, MA, 2003. MR 1 997 439

[Hil02] A.J. Hildebrand, *Short course on asymptotics*, Lecture Notes from REU Program in Number Theory (unpublished), 2002.

[Hua82] Loo Keng Hua, *Introduction to number theory*, Springer-Verlag, Berlin, 1982, Translated from the Chinese by Peter Shiu. MR 83f:10001

[Lan69] Edmund Landau, *Vorlesungen über Zahlentheorie. Erster Band, zweiter Teil; zweiter Band; dritter Band*, Chelsea Publishing Co., New York, 1969. MR 40 #4076

[Mur01] M. Ram Murty, *Problems in analytic number theory*, Graduate Texts in Mathematics, vol. 206, Springer-Verlag, New York, 2001, Readings in Mathematics. MR 2001k:11002

[New60] Donald J. Newman, *A simplified proof of Waring's conjecture*, Michigan Math. J **7** (1960), 291–295. MR 22 #10967

[Ste47] H. Steinhaus, *Sur un théorème de M. V. Jarník*, Colloquium Math. **1** (1947), 1–5. MR 9,335d

[Ten95] Gérald Tenenbaum, *Introduction to analytic and probabilistic number theory*, Cambridge Studies in Advanced Mathematics, vol. 46, Cambridge University Press, Cambridge, 1995, Translated from the second French edition (1995) by C. B. Thomas. MR 97e:11005b

Appendix B

Finitely Generated Abelian Groups

In this brief appendix we develop the basic theorems on finitely generated $\mathbf{Z}$-modules (aka finitely generated abelian groups).

B.1 The Fundamental Theorem

Theorem B.1.1 (Fundamental Theorem of Finitely Generated Abelian Groups, sans uniqueness). *Let M be a finitely generated $\mathbf{Z}$-module, which we suppose can be generated by k elements and no fewer. Then there are elements $\overline{\mathbf{e_1}}, \ldots, \overline{\mathbf{e_k}}$ of M with*

$$M = \mathbf{Z}\overline{\mathbf{e_1}} \oplus \cdots \oplus \mathbf{Z}\overline{\mathbf{e_k}},$$

and such that the order of $\overline{\mathbf{e_i}}$ divides the order of $\overline{\mathbf{e_i} + \mathbf{1}}$ (for $1 \leq i \leq r-1$), where r is such that the order of $\overline{\mathbf{e_i}}$ is finite for $i \leq r$ and infinite for $r < i \leq k$.

Remark. Naturally, this internal direct sum decomposition implies an external direct sum decomposition. Namely, if r (respectively s) is the number of generators $\overline{\mathbf{e_i}}$ of finite (respectively infinite) order, then letting m_i denote the order of $\overline{\mathbf{e_i}}$, we have

$$M \cong \mathbf{Z}/m_1\mathbf{Z} \oplus \cdots \oplus \mathbf{Z}/m_r\mathbf{Z} \oplus \mathbf{Z}^s,$$

where $m_1 \mid m_2 \mid \cdots \mid m_r$.

Proof (E. Artin). We proceed by induction on k, the smallest number of elements required to generate M. The case $k = 1$ being trivial, we assume $k > 1$. We consider all minimal generating sets $\mathbf{e_1}, \ldots, \mathbf{e_k}$ for M and all nontrivial relations

$$m_1\mathbf{e_1} + \cdots + m_k\mathbf{e_k} = 0, \tag{B.1}$$

where "nontrivial" means not all the $m_i = 0$. From all nontrivial relations (B.1) among all minimal generating systems, we fix a relation (B.1) containing the smallest positive coefficient. (If there are no such relations, then any minimal generating system is a $\mathbf{Z}$-basis for M, so the theorem holds.) By reordering the $\mathbf{e_i}$, we can assume m_1 is this minimal positive coefficient.

We claim:

i) In any relation
$$n_1\mathbf{e_1} + \cdots + n_k\mathbf{e_k} = 0$$
among the $\mathbf{e_i}$, we have $m_1 \mid n_1$.

ii) For every $i = 1, \ldots, k$, we have $m_1 \mid m_k$.

To prove the first claim, note that if $m_1 \nmid n_1$, then with q chosen so that $0 < n_1 - m_1q < m_1$, we have
$$(n_1 - m_1q)\mathbf{e_1} + \cdots + (n_k - m_kq)\mathbf{e_k} = 0;$$
this relation has a smaller positive coefficient (namely $n_1 - m_1q$) than m_1, contradicting minimality.

To prove the second claim, suppose for example that $m_1 \nmid m_2$. Then writing $m_2 = m_1q + r$, we obtain
$$m_1(\mathbf{e_1} + q\mathbf{e_2}) + r\mathbf{e_2} + \cdots + m_k\mathbf{e_k} = 0. \tag{B.2}$$
But $\mathbf{e_1} + q\mathbf{e_2}, \ldots, \mathbf{e_k}$ is also minimal generating system. Since the relation (B.2) possesses a smaller positive coefficient than m_1, we again obtain a contradiction to the minimality of m_1.

By the second claim we have $m_i = m_1q_i$ for integers $q_1, \ldots, q_k$, and therefore
$$\begin{aligned} 0 &= m_1\mathbf{e_1} + m_1q_2\mathbf{e_2} + \cdots + m_1q_k\mathbf{e_k} \\ &= m_1(\mathbf{e_1} + q_2\mathbf{e_2} + \cdots + q_k\mathbf{e_k}). \end{aligned}$$
Define
$$\overline{\mathbf{e_1}} := \mathbf{e_1} + q_2\mathbf{e_2} \cdots + q_k\mathbf{e_k}. \tag{B.3}$$
It follows from claim i) above that set of integers m with $m\overline{\mathbf{e_1}} = 0$ is exactly the principal ideal generated by m_1; hence $\mathbf{Z}\overline{\mathbf{e_1}} \cong \mathbf{Z}/m_1\mathbf{Z}$.

Now consider the (minimal) generating system $\overline{\mathbf{e_1}}, \mathbf{e_2}, \ldots, \mathbf{e_k}$. We claim
$$M = \mathbf{Z}g_1 \oplus (\mathbf{Z}g_2 + \cdots + \mathbf{Z}g_k).$$
That this holds with $\oplus$ replaced by $+$ is clear; to see that the sum is direct, first note that if
$$n_1\overline{\mathbf{e_1}} + n_2\mathbf{e_2} + \cdots + n_k\mathbf{e_k} = 0$$
for some $n_1, \ldots, n_k \in \mathbf{Z}$, then (by (B.3)) we can find integers $N_2, \ldots, N_k \in \mathbf{Z}$ for which
$$n_1\mathbf{e_1} + N_2\mathbf{e_2} + \cdots + N_k\mathbf{e_k} = 0;$$
by our first claim, we then have $m_1 \mid n_1$. Since $m_1\overline{\mathbf{e_1}} = 0$, in this case both $n_1\overline{\mathbf{e_1}}$ and $n_2\mathbf{e_2} + \cdots + n_k\mathbf{e_k} = 0$. Therefore the sum is direct as claimed.

Now the submodule generated by $\mathbf{e_2}, \ldots, \mathbf{e_k}$ has a minimal generating system of size $k-1$. (The size of a minimal generating set clearly does not exceed $k-1$, and if we could pick a generating set of fewer than $k-1$ elements then appending $\mathbf{e_1}$ we would find a generating set of M with fewer than k elements.) By induction, choose $\overline{\mathbf{e_2}}, \ldots, \overline{\mathbf{e_k}}$ for which

$$\mathbf{Z}\mathbf{e_2} + \cdots + \mathbf{Z}\mathbf{e_k} = \mathbf{Z}\overline{\mathbf{e_2}} \oplus \cdots \oplus \mathbf{Z}\overline{\mathbf{e_k}}$$

and arranged so as to satisfy the divisibility conditions of the theorem. Let m_i denote the order of $\overline{\mathbf{e_i}}$, for $i = 2, \ldots, k$, so that $m_2 \mid \cdots \mid m_k$.

To complete the proof we need only check that m_1 divides m_2, provided m_2 is defined. To see this, write down the relation

$$m_1\overline{\mathbf{e_1}} + m_2\overline{\mathbf{e_2}} + 0\overline{\mathbf{e_3}} + 0\overline{\mathbf{e_4}} + \cdots + 0\overline{\mathbf{e_k}} = 0.$$

If m_1 does not divide m_2, then we can write $m_2 = m_1 Q + R$ with $0 < R < m_1$ to obtain

$$m_1(\overline{\mathbf{e_1}} + Q\mathbf{e_2}) + R\overline{\mathbf{e_2}} + \cdots + m_k\overline{\mathbf{e_k}} = 0;$$

this is a relation among a minimal system of generators with a positive coefficient smaller than m_1. □

B.2 Free Z-modules of Finite Rank

A **Z**-module M is said to be free of finite rank if there are elements $\mathbf{e_1}, \mathbf{e_2}, \ldots, \mathbf{e_n}$ in M with the property that every $\mathbf{v} \in M$ admits a unique expression in the form

$$\mathbf{v} = c_1\mathbf{e_1} + \cdots + c_n\mathbf{e_n} \quad (c_i \in \mathbf{Z});$$

in this case $\mathbf{e_1}, \ldots, \mathbf{e_n}$ are said to be a *basis* for M. We define the *rank* of a such a module as the number of elements in a basis; the next lemma assures us this is well-defined.

Lemma B.2.1. *Suppose M is a **Z**-module with a basis of n elements, where n is a positive integer. Then every basis of M possesses n elements.*

Proof. Because a **Z**-module with a basis of n elements is isomorphic to $\mathbf{Z}^n$, we may assume $M = \mathbf{Z}^n$. Let $\mathbf{e_1}, \ldots, \mathbf{e_m}$ be any basis of $\mathbf{Z}^n$. Then $\mathbf{e_1}, \ldots, \mathbf{e_m}$ is also a basis of $\mathbf{Q}^n$. Indeed, any linear dependence over **Q** implies a linear dependence over **Z** upon clearing denominators. Also, if $v \in \mathbf{Q}^n$, then $rv \in \mathbf{Q}^n$, for some positive integer r, so rv is in the **Z**-span of $\mathbf{e_1}, \ldots, \mathbf{e_m}$, whence v is the **Q**-span of $\mathbf{e_1}, \ldots, \mathbf{e_m}$. The equality $m = n$ follows now from uniqueness of dimension. □

Theorem B.2.2. *Let M be a free **Z**-module of rank n. Then every submodule N of M is free of rank at most n.*

Proof. We may assume $M = \mathbf{Z}^n$. For each $1 \leq k \leq n$, let N_k be the submodule of N consisting of those points $(x_1, \ldots, x_n) \in N$ whose first $k-1$ coordinates $x_1, \ldots, x_{k-1}$ vanish, with the convention that $N_1 = N$. Let I_k be the set of kth coordinates of elements of N_k. Then I_k is an ideal of $\mathbf{Z}$, so that $I_k = \mathbf{Z}g_k$ for some integer g_k. If $g_k \neq 0$ then let $\mathbf{e_k}$ be any element of N_k with kth coordinate g_k; when $g_k = 0$, let $\mathbf{e_k} = 0$.

We claim $\mathbf{e_1}, \ldots, \mathbf{e_n}$ generate N. For let $\mathbf{v} \in \mathbf{N}$ be given; we can choose, by definition of g_1, an integer c_1 such that $\mathbf{v} - c_1\mathbf{e_1}$ has vanishing first coordinate. We can then choose, by definition of g_2, an integer c_2 such that

$$\mathbf{v} - c_1\mathbf{e_1} - c_2\mathbf{e_2}$$

has vanishing first and second coordinates, and so on. In this way we find

$$\mathbf{v} = c_1\mathbf{e_1} + \cdots + c_n\mathbf{e_n}.$$

It follows that the $\mathbf{e_i}$ span N, and this remains true if we throw away those $\mathbf{e_i} = 0$.

We claim that the remaining elements $\mathbf{e_{i_1}}, \mathbf{e_{i_2}}, \ldots, \mathbf{e_{i_k}}$, with $i_1 < i_2 < \cdots < i_k$, form a basis for N. It only remains to check linear independence. So let

$$\mathbf{v} := c_1\mathbf{e_{i_1}} + \cdots + c_k\mathbf{e_{i_k}}$$

be a linear combination of these vectors, with not all the c_j vanishing, and let c_j be the first nonvanishing coefficient. Then $\mathbf{v}$ has i_jth coordinate $c_j g_{i_j} \neq 0$, so $\mathbf{v} \neq 0$. $\square$

Index