Structure and Randomness in Arithmetic Settings

by

Erick Bryce Wong

B.Sc., Simon Fraser University, 1994
M.Sc., Simon Fraser University, 1997

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

The Faculty of Graduate Studies

(Mathematics)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

August 2012

© Erick Bryce Wong 2012

Abstract

We study questions in three arithmetic settings, each of which carries aspects of random-like behaviour. In the setting of arithmetic functions, we establish mild conditions under which the tuple of multiplicative functions $[f_1, f_2, \ldots, f_d]$, evaluated at d consecutive integers $n+1, \ldots, n+d$, closely approximates points in $\mathbb{R}^d$ for a positive proportion of n; we obtain a further generalization which allows these functions to be composed with various arithmetic progressions.

Secondly, we examine the eigenvalues of random integer matrices, showing that most matrices have no rational eigenvalues; we also identify the precise distributions of both real and rational eigenvalues in the 2×2 case.

Finally, we consider the set $S(k)$ of numbers represented by the quadratic form $x^2 + ky^2$, showing that it contains infinitely many strings of five consecutive integers under many choices of k; we also characterize exactly which numbers can appear as the difference of two consecutive values in $S(k)$.

Preface

A version of Chapter 2 was published as:

- Wong, E. B. (2008) Simultaneous approximation of reals by values of arithmetic functions. *Anatomy of Integers*, volume 46 of *CRM Proc. Lecture Notes*, pp. 289–297. Amer. Math. Soc., Providence, RI. [86]

Chapter 3 is based on two published papers:

- Martin, G. and Wong, E. (2008) The number of 2 × 2 integer matrices having a prescribed integer eigenvalue. *Algebra Number Theory*, 2(8):979–1000. [55]

- Martin, G. and Wong, E. B. (2009) Almost all integer matrices have no integer eigenvalues. *Amer. Math. Monthly*, 116(7):588–597. [54]

The earliest manuscripts were drafted by Dr. Martin, based on my original research; thereafter, we shared equal responsibility for both writing and subsequent research.

In addition, a version of Proposition 3.6 in Section 3.2 “Basic bounds on singularity” was published as Lemma 1.1 in:

- Pataki, G., Tural M. and Wong, E. B. (2010) Basis reduction and the complexity of branch-and-bound. In *Proceedings of the Twenty-First Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1254–1261, Philadelphia, PA. SIAM. [66]

My primary contribution to this paper was the formulation and proof of this lemma. A version of this proof was submitted to *SIAM J. Computing* in an extended form of the above paper.

Table of Contents

List of Tables

List of Figures

Acknowledgements

My immediate thanks go to my supervisor Greg Martin, for his careful readings of many drafts, and helpful responses to my panicked e-mails. Through many different turns he has been a wise mentor, an inspiring teacher, an affable colleague, and a tireless supporter.

I also thank my committee members, Izabella Łaba and József Solymosi, and my examiners, Igor Shparlinski, Mark Greenstreet, Michael Bennett and Gordon Semenoff, for the valuable contribution of their time and insights.

It has been an unusually long journey from my Master's degree to this point, shaped by countless individuals. I am truly grateful to all, though space permits me to mention but a few.

I thank my family for their ever-present love and support. I feel quite fortunate to have always had the freedom to pursue my interests, however esoteric (let me also thank my nephew Paris for sharing in those interests).

I recognize the generous assistance of the Natural Sciences and Engineering Research Council of Canada for a portion of this work. I am grateful for the support of the MS/MRI Research Group, especially Andrew Riddehough for his personal encouragement.

I owe many thanks to all the staff of the mathematics department, particularly Lee Yupitun for her boundless knowledge of administrative vagaries.

I am fortunate to have met people here who fuelled my mathematical curiosity in memorable ways; I am doubly fortunate to count them among my friends. In order of appearance: Vishaal Kapoor, Mike Tsiang, Florina Halasan, Nishant Chandgotia, Vasu Tewari and Paul Pollack.

Finally, I wish to thank my wife Eva Koo for companionship, support and encouragement through good times and bad; for believing in me when I could not; for putting some structure in a random life.

Dedication

To my father, who almost made it to 90 years.

Happy Birthday, Dad.

Chapter 1

Introduction

1.1 Setting the stage

Within mathematics, the field of number theory is notorious for its ability to generate problems which are eminently plausible and well-supported by evidence, and yet for which a rigorous proof appears far beyond the reach of current technology. We would like to begin this chapter by attempting to convey to the reader why such a disparity should exist.

One often finds in analytic number theory that the basic *statistics* of a particular object of interest can be understood quite well. At the same time, such an object seems to possess almost no tractable *structure* which serves to distinguish it from any other random process with the same statistics. Roughly speaking, so many of the things we are interested in behave as though they are *random* even though they are decidedly not so (for a concrete analogy, the digits of π seem to vary randomly, but the 763rd digit of π will always be 9 no matter how often one repeats the experiment). While it is entirely reasonable to make predictions about such objects, and these predictions are easily seen to be true for the vast majority of their random counterparts, this lack of structure leaves us with no easy way to certify that our particular object is not one of the unlucky few.

A prime example to illustrate this dilemma is given by the familiar sequence of *prime numbers*: 2, 3, 5, 7, 11, 13, 17, 19, and so on. The primes are in some sense defined by their lack of structure, and it is indeed a non-trivial task to predict the spacings between each prime number and the next. Yet the statistics of this sequence have been well-understood for more than a century. If we use the conventional notation $\pi(n)$ to denote the number of prime numbers between 1 and n, then the Prime Number Theorem (a triumph of late 19th century mathematics) states that $\pi(n)$ is *asymptotic* to $n/\log n$, that is

$$\lim_{n\to\infty} \pi(n) \Big/ \frac{n}{\log n} = 1.$$

As n gets progressively larger, $n/\log n$ becomes a better approximation to $\pi(n)$, in the sense of relative error. Thus, a number chosen at random from the interval $[1, n]$ bears a roughly 1 in $\log n$ chance of being prime (here, as in the entirety of this thesis, $\log n$ denotes the natural logarithm of n).

1.2 Modelling the primes

The preceding observation gives rise naturally to the *Cramér model*, which has been rather successful at making quantitative predictions about primes (though far from perfectly so: see [28, 52]). We describe a slightly simplified version of this model: fix a number $n > 2$, and choose a random subset $P(n)$ of the numbers $\{1, 2, \ldots, n\}$ by selecting each element independently with probability $1/\log n$, so that $P(n)$ has $n/\log n$ elements on average. The Cramér model predicts that if a given pattern is highly likely to occur within $P(n)$, then the same pattern should also occur with similar frequency in the primes up to n. The advantage of such a simple random model lies in the ease with which one can analyze the properties of $P(n)$.

The simplest type of pattern one could reasonably apply this model to is that of *twin primes*: pairs of primes like 71 and 73 which differ by exactly 2. In other words, we wish to understand how often it is that both m and $m+2$ are simultaneously prime. In the random model, the probability that m and $m+2$ are both included in $P(n)$ is exactly $1/\log^2 n$ for any $1 \le m \le n-2$, and so the expected number of "twins" in $P(n)$ is $(n-2)/\log^2 n$, or roughly $n/\log^2 n$. According to the Cramér model, we should expect the true number of twin primes up to n to be roughly $n/\log^2 n$.

This prediction turns out to be quite good, with one significant caveat: notice that $P(n)$ also contains consecutive pairs of the form $\{m, m+1\}$ with the same frequency $n/\log^2 n$. On the other hand, this pattern is nigh impossible for primes, because primes are always odd (with the sole exception of the prime 2). The primes do possess a structure after all, at least in a negative sense: their lack of divisibility by 2 makes the pattern $\{m, m+1\}$ impossible in the primes (aside from $\{2, 3\}$), however common it is in the random model. The same reasoning has a more subtle influence on the pattern $\{m, m+2\}$: the frequency should be *better* than the random model predicts, because if m is prime then $m+2$ is certainly odd, and this makes it more likely to be prime than some random number which may equally be odd or even.

This simple idea of examining a problem through the lens of "even" and "odd" (more generally, the remainders modulo a particular prime p or prime power p^r) is common in number theory, and is referred to as a *local* argument. By adjusting the naïve estimate $n/\log^2 n$ with a certain correction factor for each prime p (called a local factor), and combining all such factors into one constant, Hardy and Littlewood (in 1923, long before much computational evidence was available) obtained an asymptotic conjecture (called a *heuristic*) for the number of twin primes up to n:

$$\#\{m \leq n : m, m+2 \text{ are prime}\} \sim C_2 \cdot \frac{n}{\log^2 n}, \tag{1.1}$$

where C_2 has the approximate value 1.32032363.

The Hardy–Littlewood conjecture for twin primes has shown strikingly accurate agreement[1] with computational evidence. In [63], Nicely computed the exact number of twin primes up to 10^{14}, and found the prediction to match to within one part in 2 million! (Incidentally, it was this computation which led to the discovery of the infamous "FDIV" division bug in the Intel Pentium processor.)

However, we are very far from understanding how to prove that this estimate remains accurate for all larger values of n. In fact, it remains an open problem whether there are *infinitely* many twin primes, let alone that they occur with the exact frequency predicted by Hardy and Littlewood.

[1] To achieve this striking accuracy, one should first replace $n/\log^2 n$ by the almost identical integral $\int_2^n (\log t)^{-2} dt$; this relates to the simplification we referred to earlier.

1.3 Other arithmetic structures

The twin primes problem of the previous section can be viewed as a special case of the study of the *gaps* between consecutive primes. If we denote the sequence of primes in increasing order by $p_1, p_2, p_3, \ldots$, we may restate the twin primes conjecture as "how often is $p_{n+1} - p_n = 2$?". More generally, one could naturally ask what is known about the spacing $d_n = p_{n+1} - p_n$? (The desire to understand gaps between primes was the motivation for Cramér introducing his random model.) The Prime Number Theorem can easily be used to deduce that, on average, d_n has size roughly equal to $\log n$. While the state of our knowledge is insufficient to conclude that $d_n = 2$ infinitely often, it is a remarkable achievement of Goldston, Pintz and Yıldırım [26] that d_n can be arbitrarily small compared to its average value: for any $\epsilon > 0$, there are infinitely many n for which $d_n < \epsilon \log n$.

One might also ask for longer patterns than twin primes. Note that it is unreasonable to expect $\{m, m+2, m+4\}$ to all be prime for local reasons modulo 3 (there is a sole exception at $m = 3$). But there is no obvious obstruction that would prevent $\{m, m+6, m+12\}$ from all being prime simultaneously. Here again, Hardy and Littlewood have a more general conjecture which encompasses all such patterns, and it has been generalized further by various authors [5, 34, 71]. These all reflect a broad principle: *if there is no simple reason why a given pattern of primes should not occur, then it should occur infinitely often, with an asymptotically predictable frequency.* However, given our inability to prove this for a mere pattern of length 2, there is little hope of establishing such a general claim.

However, there is one other type of pattern for which great success has been achieved, particularly in recent years. Notice that the previous pattern $\{m, m+6, m+12\}$ consists of three equally-spaced values in *arithmetic progression.* If we relax the demand for a particular spacing and allow for any pattern of the form $\{m, m+d, m+2d\}$ with some integer $d \geq 1$, then the question becomes: *do the primes contain infinitely many arithmetic progressions of length 3?*

We will shortly address this question and its analogues to longer arithmetic progressions, but first we take a brief excursion to describe some closely related and beautiful work from the field of additive combinatorics.

1.4 Density and the primes

Given a set $A \subseteq \mathbb{N}$ of natural numbers, one often describes how well-populated A is, relative to $\mathbb{N}$, by the quantity

$$\delta := \lim_{x\to\infty} \frac{\#\{n \leq x : n \in A\}}{x}.$$

We say that A has *asymptotic density* δ, provided this limit exists (if it does not converge, one can still define lower and upper densities by taking the $\liminf$ or $\limsup$). For example, the set of all even numbers has asymptotic density $\frac{1}{2}$, and this could be loosely interpreted as the “probability” that a natural number, chosen at random, is even. However, some care is warranted in this interpretation, as we discuss further in Section 1.10.

Roth proved in 1953 [68] that every set A having a positive density $\delta > 0$ must contain infinitely many arithmetic progressions of length 3. So long as A is “dense”, this pattern — which we abbreviate as 3-AP — is inherently *unavoidable*, no matter how one tries to arrange the elements of A to prevent it from recurring. We shall not describe Roth’s elegant proof in complete detail, but we feel compelled to give a loose account of its clever use of a dichotomy between randomness and structure.

If A is distributed “randomly” enough (in a sense that can be made precise in terms of discrete Fourier transforms), then it can be shown that A contains just as many 3-APs as a typical set whose elements are chosen at random with probability δ. How many is that? In this truly random model, every given subset of size 3 occurs with probability $\delta^3 > 0$, so a random set will typically contain many 3-APs.

On the other hand, if A is not “random” in this specific sense, Roth showed that there must be some structure within A that allows one to pass to a new set A' having two key properties: A' contains fewer 3-APs than A, and the density of A' is appreciably greater than δ.

Suppose now that A doesn't contain any 3-APs. Then it does not fall into the "random" case, and so we may replace it by A' which has higher density. Since A' also lacks 3-APs, it too can be replaced by A'', and so on. Eventually we arrive at a set whose density is so high (namely, $\delta > \frac{2}{3}$) that it cannot avoid any pattern of length 3, a contradiction.

Szemerédi later extended Roth's theorem to arithmetic progressions of any length k, through an extraordinarily intricate argument that introduced entirely new notions of structure and randomness [80]. Since then, much sharper versions of both Roth's and Szemerédi's theorems have been found: the current best results are due to Sanders [70] and Gowers [27].

Let us now return to the question from the end of the previous section. Recall that the Prime Number Theorem predicts the density of primes up to x to be about $1/\log x$, so the asymptotic density of the primes is 0. Unfortunately, Roth's theorem does not apply to sets as sparse as the primes. There aren't enough primes to justify the existence of 3-APs on the sole basis that it is unavoidable for any set with the same density (although the refinement by Sanders comes breathtakingly close to doing so).

Nevertheless, the primes do contain infinitely many 3-APs (with the same frequency as predicted by random models); this was proven by van der Corput [84] in 1939! Essentially, one can show that the primes are distributed sufficiently nicely that they already fall into the "random" case of Roth's proof: the necessary estimates were present two years earlier in Vinogradov's pioneering work on exponential sums [85].

It took many more decades to establish the existence of longer k-APs in the primes. In part this is because the precise sense of "random" that is needed to enumerate k-APs did not begin to emerge until the seminal work of Host and Kra [40], and Gowers [27]. Building on these insights and much more, Green and Tao [29] proved a landmark achievement: *the primes contain arbitrarily long arithmetic progressions.* More recently they have shown, together with Ziegler [30–33], that such progressions occur with the precise frequency predicted by random models: for any $k \geq 2$,

$$\#\{k\text{-APs within primes up to } n\} \sim \frac{D_k}{2(k-1)} \cdot \frac{n^2}{\log^k n}, \tag{1.2}$$

where the constant D_k accounts for all local factors (the other terms arising purely from the naïve model).

1.5 Notation and outline

With the necessary background in place, we now go into some detail about the specific questions considered in this dissertation, which cross through all of the themes of the previous sections. Starting in Section 1.6, we discuss *arithmetic functions*, which constitute the main focus of Chapter 2. This discussion continues until Section 1.11, where we turn to the topic of *random matrices* from Chapter 3. From Section 1.15 onward, we discuss the *combinatorial patterns* central to Chapter 4. Let us take a moment to introduce some common notation that will appear throughout this work.

We use the standard convention that $\mathbb{N}$ refers to the natural numbers beginning with 1, not 0. Most Roman-lettered variables, unless otherwise specified, are assumed to take values in $\mathbb{N}$, especially n, m, k, d; we shall reserve p to indicate a prime value. The variable ϵ merits some attention: it most often refers to a (small) positive real number, but we use ϵ in some parts for a quantity that is ± 1 (the distinction will be obvious in context).

If $g(n)$ is a positive function, the Vinogradov notation $f(n) \ll g(n)$ indicates that there is a constant $C > 0$ such that $|f(n)| \leq Cg(n)$ regardless of the choice of n. This naturally generalizes to functions of multiple variables, and we may use a subscript to denote that the implied constant may depend on some of these variables. For example, it is well-known that $\log n \ll_\epsilon n^\epsilon$ for any $\epsilon > 0$.

When working with error terms it is convenient to use the Landau notation $O(g(n))$ to abbreviate a quantity that is $\ll g(n)$, and likewise $O_\epsilon(g(n))$ if said quantity is $\ll_\epsilon g(n)$. We also use $o(g(n))$ to denote a quantity which is asymptotically much smaller than $g(n)$; that is, $f(n) = o(g(n))$ if and only if $\lim_{n\to\infty} f(n)/g(n) = 0$. In this case, a subscript as in $o_\epsilon(g(n))$ allows the rate of convergence of this ratio to depend non-uniformly on ϵ.

The notation $d \mid n$ means that d divides n (or, n is divisible by d). When p is prime, $p^r \parallel n$ more specifically means that p^r *exactly divides* n, so that $p^r \mid n$ but $p^{r+1} \nmid n$. The notation (m, n) is often used for the GCD or greatest common (positive) divisor of m and n. We say that m and n are *relatively prime* or *coprime* if $(m, n) = 1$. At times we may use $(\cdot, \cdot)$ to instead denote an ordered pair; again, the context should make this clear.

When a set is written in comprehension notation such as $\{n \in \mathbb{N} : n \leq x\}$, we will often prepend # to denote the cardinality, as seen in (1.1) or (1.2).

1.6 Arithmetic functions

The study of *arithmetic functions* is of central interest in analytic number theory. Strictly speaking, any real-valued function $f : \mathbb{N} \to \mathbb{R}$ (or, more generally, a complex-valued function) defined on the natural numbers could be called an "arithmetic function", but in actual usage the term carries a connotation that the value of $f(n)$ has some relation to the divisibility properties of n. The most commonly studied class of arithmetic functions are the *multiplicative functions*, which are defined to satisfy $f(1) = 1$ and

$$f(mn) = f(m)f(n), \qquad \text{whenever } (m, n) = 1. \tag{1.3}$$

(The condition $f(1) = 1$ is nearly redundant, and could be deduced from (1.3) if we ignore the constant function $f \equiv 0$.) It is not hard to see that if the function f is multiplicative then it is uniquely determined by the values $f(p^r)$ taken at all prime powers p^r. It therefore suffices to specify a given multiplicative function only on prime powers, and allow the definition to extend to all natural numbers via (1.3). We give below a brief list of special functions, being already of natural interest, which also enjoy this multiplicative property.

- The *Euler totient function* $\phi(n)$ counts the number of integers between 1 and n that are relatively prime to n (also the number of invertible elements in the ring $\mathbb{Z}/n\mathbb{Z}$). This is multiplicative, with $\phi(p^r) = p^{r-1}(p-1)$ for every $r \geq 1$.

- The *divisor function* $\tau(n)$ counts the number of positive divisors of n (including n itself). This is multiplicative, with $\tau(p^r) = r + 1$.
- The *sum-of-divisors function* $\sigma(n)$ is the sum of all positive divisors of n. This is multiplicative, with $\sigma(p^r) = 1 + p + p^2 + \cdots + p^r = \frac{p^{r+1}-1}{p-1}$.

The above three functions are among the most well-studied arithmetic functions in number theory, yet much about them is still unknown. To give a sobering example, mathematicians in ancient Greece were interested in so-called "perfect" numbers such as $6 = 1 + 2 + 3$ and $28 = 1 + 2 + 4 + 7 + 14$, which are exactly equal to the sum of their respective divisors (not counting the original number itself). In modern notation, the number n is perfect if and only if $\sigma(n)/n = 2$. The simple question "does there exist an *odd* number that is perfect?" remains unsolved despite being posed over two millennia ago (it may thus stake a reasonable claim as the world's oldest unsolved problem).

Note that the function $\sigma(n)/n$ is itself multiplicative (more generally, this is true of any product or quotient of multiplicative functions), and like most such functions, its value depends quite crucially on the prime factors of n. As witnessed by the previous sections, we certainly do not have perfect knowledge of the primes, and multiplicative functions likewise remain an area of intense study. While the previous paragraph highlights the difficulty of understanding when $\sigma(n)/n$ is *exactly* equal to 2, there has been much more success in establishing the *approximate* behaviour of $\sigma(n)/n$: how often does it fall between, say, 1.999 and 2.001?

1.7 Distribution of an arithmetic function

The previous question was settled by Davenport [15], who showed that for any fixed real numbers $1 \leq s < t$, the set of positive integers n for which $s < \sigma(n)/n < t$ has a positive asymptotic density $\delta(s,t) > 0$. (Note that

$\sigma(n)$ is always larger than n, so there is no reason to consider values below 1.) Furthermore, this density varies *continuously* with s and t, so the density associated to any single value of $\sigma(n)/n$ is zero.

An analogous result was established earlier by Schoenberg [73] for the multiplicative function $n/\phi(n)$. The two functions are quite similar in that they both increase as the number of prime factors of n grows, with smaller primes having significantly greater effect. A celebrated theorem of Erdős and Wintner [19] shows that both results fall under a general phenomenon common to many positive-valued multiplicative functions $f(n)$: under mild convergence conditions on the logarithm[2] $g(n) := \log f(n)$, there is a continuous function $F(t)$ such that, for any $t \in \mathbb{R}^+$, the asymptotic density of the set $\{n : f(n) \leq t\}$ exists and is equal to $F(t)$. We reproduce these conditions below, which are in fact both necessary and sufficient (in each case the summation index p is restricted to prime values):

$$\sum_{|g(p)|>1} \frac{1}{p}, \sum_{|g(p)|\leq 1} \frac{g(p)}{p}, \text{ and } \sum_{|g(p)|\leq 1} \frac{g^2(p)}{p} \text{ converge;} \quad \sum_{g(p)\neq 0} \frac{1}{p} \text{ diverges.} \tag{1.4}$$

The function $F(t)$ closely recalls the (cumulative) distribution function of a random variable in probability, and it too is called the *distribution function* of $f(n)$. However, we remind the reader that there is no well-defined random variable that concretely represents all values of $f(n)$; rather, we are taking the limiting distribution of $f(n)$ taken over $n \leq x$ as $x \to \infty$. The reader familiar with measure-theoretic probability will recognize this as a *weak limit* of measures.

While one *can* write down an explicit formula for the distribution of $\sigma(n)/n$ in terms of its *characteristic function* or Fourier transform (first established in a later paper of Schoenberg [74]), such an expression does not lend itself to high-precision computation. It was only very recently that Kobayashi in his doctoral thesis [45] computed $F(2)$ to four decimal places of accuracy, showing that the density $1 - F(2)$ of the *abundant numbers*

[2] The resulting function $g(n)$ is, not surprisingly, called an *additive function*, since it satisfies $g(mn) = g(m) + g(n)$ for $(m, n) = 1$.

$\{n : \sigma(n)/n > 2\}$ is approximately 0.2476.

1.8 Approximation by arithmetic functions

The Erdős–Wintner theorem can be interpreted as saying that many positive multiplicative functions $f(n)$ can be modelled (subject to the above caveat) by a continuous random variable. Furthermore, for the two cases considered by Davenport and Schoenberg, this distribution function is *strictly* increasing, so that $f(n)$ approximates, within any desired accuracy ϵ, every real number $s \geq 1$ for a positive frequency of n ($F(s+\epsilon) - F(s-\epsilon)$, to be exact). In other words, the set of values $\{f(n) : n \in \mathbb{N}\}$ is topologically dense in $[1, \infty)$, and in a rather strong sense (every point has a large number of approximants).

Let us now consider a stronger question that may be reminiscent of the twin primes problem: since n and $n+1$ share no common prime factors, it is reasonable to believe that $f(n)$ and $f(n+1)$ should behave "independently" in some sense. Is it possible for $f(n)$ to approximate one number s_1 while $f(n+1)$ *simultaneously* approximates a different number s_2? Given that the values of $f(n)$ are dense in $[1, \infty)$, is it true that the pairs of values $(f(n), f(n+1))$ are dense in $[1, \infty) \times [1, \infty)$?

Schinzel and Wang [72] proved a result in this general direction for the specific function $\phi(n)$: for any dimension $d \geq 1$, the vector of ratios

$$\left(\frac{\phi(n+1)}{\phi(n)}, \frac{\phi(n+2)}{\phi(n)}, \ldots, \frac{\phi(n+d)}{\phi(n)} \right) \tag{1.5}$$

can approximate any point in $(0, \infty)^d$ arbitrarily well. We remark that while n and $n+k$ may have some prime factors in common, their GCD cannot exceed k, so it is fair to expect $\phi(n)$ and $\phi(n+k)$ to retain some weaker level of independence (when n is large and k is fixed). This is, however, true in a slightly weaker sense than previously: the set of n for which (1.5) approximates a given point to within ϵ (in every coordinate) might have asymptotic density 0.

Shortly afterward, Erdős and Schinzel [20] established a wide generalization for an arbitrary multiplicative $f(n)$ with certain mild hypotheses: for any dimension $d \geq 1$ there is a constant c_d such that the vectors

$$(f(n+1), f(n+2), \ldots, f(n+d)) \tag{1.6}$$

are dense in $[c_d, \infty)^d$. Note that our initial speculation that $c_2 = 1$, in the classical case $f(n) = \sigma(n)/n$, is actually false: it is impossible for $f(n)$ and $f(n+1)$ to both be very close to 1, since one of n or $n+1$ must be even, and for any even number m, $f(m) \geq \frac{3}{2}$. So it is genuinely necessary to prescribe a different c_d for each dimension: the influence of local factors is felt here once more.

Moreover, the Erdős–Schinzel proof shows that the set of n which approximate any given point in $[c_d, \infty)^d$ (within some fixed tolerance ϵ) has a positive asymptotic *lower* density. The hypotheses on $f(n)$ are similar to the Erdős–Wintner conditions, but they do not necessarily imply that f has a distribution function, so it would be unreasonable to expect the aforementioned set to have a precise (nonzero) asymptotic density. However, as an approximation property this is arguably just as strong: we still have that a *positive proportion* of $n \leq x$ are good approximants, and for some values of x this proportion may be higher still.

1.9 Comparing arithmetic progressions

A related problem was more recently considered in a brief note of Newman [62], which is motivated by a simple observation: while $\phi(n)/n$ is well-modelled by a random variable (whose average value works out to exactly $6/\pi^2$), the function $\phi(6n)$ contains an inherent bias: it is easy to verify that if $m = 6n$, then $\phi(m)/m \leq \frac{1}{3}$, which is considerably smaller than average. Therefore, we might expect that the inequality $\phi(6n) > \phi(6n+1)$ should occur very rarely[3], and rarer still for $\phi(30n) > \phi(30n+1)$. Indeed, the former does not occur until $n \approx 4.1 \times 10^{55}$, and the smallest solution to the

[3] In fact, one can show that the average value of $\phi(m)/m$ for $m = 6n+1$ is $9/\pi^2$ (much higher than average), which should lower our expectations even further.

latter, explicitly computed by Martin [53], is 1116 digits long!

Nevertheless, Newman showed that such unusual reversals do occur at infinitely many n, and more generally this is true for any inequality of the form $\phi(an + b) > \phi(cn + d)$ where a, b, c, d are fixed integers satisfying $ad - bc \neq 0$. (This latter constraint is necessary to prevent $cn + d$ from being a scalar multiple of $an + b$, which would force them to have almost exactly the same prime factors: for instance, it is not hard to see that $\phi(n) > \phi(2n)$ can never occur.)

In Chapter 2, we will show that Newman's result may be subsumed into a more general version of the Erdős–Schinzel theorem that applies to such arithmetic progressions. Specifically, if $a_1n+b_1, a_2n+b_2, \ldots, a_dn+b_d$ are any choice of non-constant linear functions (none of which is a scalar multiple of another), then the vectors

$$(f(a_1n + b_1), f(a_2n + b_2), \ldots, f(a_dn + b_d)) \tag{1.7}$$

are dense in $[c, \infty)^d$ for some constant c (which depends on the specific choice of linear functions and not just d), in the same strong sense as before (there are many are good approximants to each point). Our result is more general in one additional sense: each component of the vector may have its own individual multiplicative function in place of f (provided these functions all satisfy similar conditions).

As a consequence of our theorem, not only does the inequality $\phi(6n) > \phi(6n + 1)$ occur with at least some positive (albeit very close to 0) frequency, but this statement remains true even if we specify that the ratio $\phi(6n)/\phi(6n + 1)$ should agree with π to 100 decimal places!

1.10 Probability in number theory

It seems appropriate at this point to elaborate on the loose interpretation of probability which recurred in Sections 1.4 and 1.7. Many results in this area can be paraphrased in probabilistic language: already we have witnessed the terms "probability", "average value" and "distribution function". Certainly, the concept of a "positive integer chosen at random" has a natural intuitive appeal, perhaps even more so than that of a real number chosen randomly from the interval $[0, 1]$. But while the latter concept is readily formalized as a legitimate well-defined random variable, the former is at odds with the measure-theoretic foundations of probability theory. There is no uniform measure that can be placed on a countably infinite set like the naturals (although asymptotic density is uniform, it is only *finitely* additive).

Ultimately, any (non-trivial) statement that refers to "random integers" is actually describing a limiting process (where at each step of the limit, the probability space remains finite and well-defined): *density* as a limit of finite probabilities, *average value* as a limit of finite averages, and *distribution function* as a (weak) limit of discrete distributions. But this interpretation, while appealing, tends to obscure the very real error terms underlying the convergence to such limits.

To give a simple concrete example, the (limiting) probability that a number is even is $\frac{1}{2}$, but the true proportion of even numbers in $\{1, \ldots, n\}$ is $\frac{1}{2} + O(\frac{1}{n})$, the implied constant depending on whether n itself is odd or even. Similarly, the probability that a number is divisible by 3 is $\frac{1}{3} + O(\frac{1}{n})$. The events that a random integer is "divisible by 2" and "divisible by 3" may be independent in the sense of limits (that is, $\frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6}$), but this ignores the error terms that appear along the way.

When working with large numbers of such not-quite-independent events, the accumulation of these error terms can be very significant. For instance, as mentioned in [28], one cannot use this simple approach to accurately count the number of primes between $\sqrt{n}$ and n (by excluding all numbers divisible by some prime $p \leq \sqrt{n}$). Nevertheless, Brun [12] found clever ways to handle these errors to show that the number of twin primes, while still poorly understood, cannot be *vastly* higher than predicted by Hardy and Littlewood's heuristic (1.1). These techniques gave rise to the modern field of *sieve theory*.

Having discussed some of the inherent complications of thinking about integers in probabilistic terms, we feel that it is justified by the fact that many beautiful theorems could not be described so elegantly in any other form. We give two such examples below (the latter is a celebrated result of Erdős and Kac [17]; together with the Erdős–Wintner theorem, these form the cornerstone of *probabilistic number theory*).

- The probability that two random integers are relatively prime is $6/\pi^2$.
- The number of prime factors of a random n is normally distributed with mean and variance both equal to $\log \log n$.

1.11 Random matrices and eigenvalues

Let us now turn to a different type of random structure (still arithmetical in nature), for which it is possible to determine the distribution with a much greater precision than most of our previous examples.

We first recall some terminology from undergraduate linear algebra: given a square matrix A, we say that A has an *eigenvalue* $\lambda \in \mathbb{C}$ if there exists a nonzero vector $\mathbf{x}$ (called an *eigenvector*) such that $A\mathbf{x} = \lambda\mathbf{x}$. Any given matrix possesses only finitely many eigenvalues: to be specific, if A is an $n \times n$ matrix, it has exactly n eigenvalues given by the roots of a certain nth degree polynomial (possibly with some repetitions). These values yield significant information about the structure of A, permitting numerous applications in diverse areas including quantum physics, statistics and epidemiology. Perhaps the most prominent usage of all is the PageRank algorithm pioneered by Google Inc. [11]: if A is suitably built to represent all links between web pages, then the *principal* eigenvector of A (one that is associated with the largest eigenvalue) yields a direct measure of each page's relative importance.

This establishes the set of eigenvalues of a matrix (colourfully known as its *spectrum*) as a central object of study: just as prime factors are intrinsic to integers, so too are eigenvalues intrinsic[4] to matrices. Given a large, or even infinite, family of matrices, we are naturally led to wonder about the statistical distribution of eigenvalues within. In keeping with the arithmetic

[4] The prefix "eigen-" derives from a Germanic word meaning "self" or "innate".

theme of this thesis, we study in Chapter 3 the (initially vague) question: *what can one say about the eigenvalues of a random matrix of integers?*

Before we expand on this question in the next section, we would be remiss not to mention some classical work on random matrices, much of which has been devoted to *continuous* (and often symmetric) random models originating from quantum physics. Even in number theory, one of the most frequently discussed matrix models is based not on integers but on complex numbers: it is a probability distribution defined on $n \times n$ Hermitian matrices[5] known as the *Gaussian unitary ensemble* (GUE). The significance of this model to number theory is that the eigenvalues of random GUE matrices, when n is very large, appear to have spacing that is statistically identical to that of the zeros of the Riemann zeta function; this surprising connection was first made at a (now-famous) chance meeting between Montgomery [58] and physicist Freeman Dyson. (Though it does not occupy a major role in this thesis, the Riemann zeta function may well be the most important object in all of analytic number theory; a greater understanding of its zeros would answer a wide range of outstanding questions about the distribution of primes — see [9] for a nice summary.)

A beautiful result of Edelman [16] is also worth mentioning here: if we populate an $n \times n$ matrix with independent entries randomly drawn from a standard normal distribution, the probability that it is *diagonalizable* over $\mathbb{R}$ (meaning essentially[6] that its eigenvalues are all real) is exactly $2^{-n(n-1)/4}$.

1.12 Random integer matrices

Our own investigation (jointly with Greg Martin) begins with a question similar to Edelman's, which was posed by Hetzel, Liew, and Morrison in [37]: *what is the probability that a random $n \times n$ integer matrix is diagonalizable over the field of rational numbers?*

As we have previously seen, this question does not refer to any single

[5] Hermitian matrices are a complex-valued analogue of real symmetric matrices.

[6] These are not precisely equivalent, should an eigenvalue be repeated; however, this complication occurs with probability 0 so we may safely ignore the distinction.

distribution of integer matrices; it must be understood as a *limiting* process — see the introduction of [37] for a lively discussion justifying the use of the term "probability" for this purpose. Just as before, we will restrict the integers to a finite set: for any $k \geq 1$, we can certainly populate an $n \times n$ matrix $M(k)$ with entries chosen independently at random from the integers between $-k$ and k. The previous question then becomes: *as $k \to \infty$, what happens to the probability that $M(k)$ is diagonalizable over $\mathbb{Q}$?*

We expect that this is equal to the probability that every eigenvalue of $M(k)$ is rational, and indeed Hetzel et al. rigorously justified this equivalence (the two probabilities are not exactly equal but they do have the same limit as $k \to \infty$). Computational experiments led them to conjecture that both probabilities converge to 0 (except in the case $n = 1$, since 1×1 matrices are already diagonal in a vacuous sense).

We confirm in Theorem 3.1 that this is indeed true. In fact we prove a stronger result: the probability that $M(k)$ has *any* rational eigenvalues tends to 0. In other words, a typical matrix does not have any rational eigenvalues, which generalizes earlier results of [47] and [37] for the case $n = 2$. Quantitatively, we show that this probability is $O_{n,\epsilon}(k^{-1+\epsilon})$ for any $\epsilon > 0$, which certainly goes to 0 for k large (we could take, say, $\epsilon = \frac{1}{2}$).

In answering the question of Hetzel et al., we are fixing the matrix size n and allowing the size of the entries k to grow. One might equally ask the converse problem: what happens if we instead fix the size of the entries, but allow the matrix to grow? This has been extensively studied in the more general formulation that the entries are chosen from any fixed probability distribution X. The case where X is a standard normal distribution was studied by Ginibre [24], who proved that as $n \to \infty$ the eigenvalues become *uniformly* distributed — in a limiting sense — throughout the unit disk of radius $\sqrt{n}$ (this random matrix model, now called the *Ginibre ensemble*, was also the setting of Edelman's result). Girko [25] showed that this limit occurs for a wider class of distributions X, and proposed that it should be *universal*, meaning that any random model of this type should converge to this same distribution (provided X has the same mean and variance). Quite recently, a series of papers by Tao and Vu [82, 83] have successfully established Girko's *circular law* in full generality (in fact, their most recent result holds for a slightly wider class of random matrices than originally conjectured).

1.13 Singularity and computation

A closely related, and perhaps simpler, variation of the eigenvalue problem is to find the probability that the random matrix $M(k)$ is *singular* (equivalently, that it has an eigenvalue of 0). In fact, our proof of Theorem 3.1 relies strongly on obtaining a quantitative upper bound of the form $O_{n,\epsilon}(k^{-2+\epsilon})$ for this singularity probability.

Whereas quantum mechanics was a key motivation for the study of Gaussian models, questions about discrete distributions arise more naturally in the field of theoretical computer science. A well-known problem in combinatorial optimization is *integer programming*: given a matrix A and a vector $\mathbf{b}$, one wishes to find an *integer* vector $\mathbf{x}$ such that $A\mathbf{x} \leq \mathbf{b}$ (here the notation $\mathbf{a} \leq \mathbf{b}$ means that $a_i \leq b_i$ for every index i). Were it not for the restriction to integer values, this could be done efficiently by standard techniques for linear programming. With this restriction, the problem is classically known to be *NP-complete*: in a very precise sense, it is just as hard to solve as several thousand other such problems, widely regarded as intractable.

While it is justifiably hard to solve integer programs in generality, the typical case is quite different (say, if A is randomly chosen as before). Our bounds for the singularity probability were adapted by Pataki, Tural and this author [66] to show that *most* instances can be solved very easily, essentially using a naïve algorithm known as *branch-and-bound*. This may seem counterintuitive, but it is not uncommon: many NP-complete problems (not all, though; see [49]) are known to have good typical-case behaviour with very few difficult instances. In contrast, the famous problem of factoring an integer into primes is believed to be quite the opposite, with very few *easy* cases[7] (it is still unknown whether factoring is NP-complete).

[7] Seeing that the security of Internet commerce currently relies on this fact, we sincerely hope this belief is not overturned in the immediate future!

1.14 Rationality versus reality

For the special case of 2 × 2 random matrices, we are able to give a very sharp answer to Hetzel et al.'s original question:

$$\Pr(M(k) \text{ is diagonalizable over } \mathbb{Q}) = \frac{C \log k}{k} + O\big(\frac{1}{k}\big),$$

where $C \approx 0.55873957$ is an explicit constant.

In fact, we devote a significant part of Chapter 3 to calculating the exact limiting distribution of rational eigenvalues of $M(k)$ in the case $n = 2$. Note that we use the term "distribution" in a broader sense than distribution functions: the concept is actually much closer to that of a *probability density function*, except that it has total area 2 rather than 1, since a matrix has two eigenvalues.

To describe this distribution, it is convenient to consider the eigenvalues of the scaled-down matrix $\frac{1}{k}M(k)$ (which are simply the eigenvalues of $M(k)$, divided by k). If one plots a histogram of the eigenvalues of $\frac{1}{k}M(k)$, restricting to only those that are *rational*, these histograms will converge as $k \to \infty$ to the distribution depicted by the solid line in Figure 3.1, which is defined explicitly as a piecewise smooth function.[8]

On the other hand, this limiting process $\frac{1}{k}M(k)$ converges (in the weak sense) to a random matrix with real entries uniformly distributed on the interval $[-1, 1]$. For such a *continuous* random matrices, there is no obvious structure to the set of rational eigenvalues: any small perturbation of a matrix with a rational eigenvalue could easily shift it to some nearby real value.[9] One might therefore ask if the limit distribution of *rational* eigenvalues could be explained simply in terms of the *real* eigenvalue distribution in the continuous case?

To answer this question, we also compute the real eigenvalue distribution for $n = 2$: this is a genuine distribution, though it is also equal to the limiting

[8] By contrast, the distribution of $\sigma(n)/n$ is known to be *singular*, having infinitely many points of non-differentiability [18].

[9] It is not clear whether one can formalize this intuition by placing a natural uniform measure on the restriction to rational eigenvalues.

distribution for the real eigenvalues of $\frac{1}{k}M(k)$. It is depicted by the dashed line in Figure 3.1, which is noticeably different from the previous one. One obvious difference is the fact that the rational distribution is unimodal, while the real distribution is bimodal. In particular, it is more common for $M(k)$ to have a rational eigenvalue near 0 than one near $3k/4$; yet the reverse is true for real eigenvalues!

This curious disparity suggests that there is some *arithmetic* structure of $\frac{1}{k}M(k)$ that is lost upon passing to the continuous limit. It would not be fair to attribute such structure to *local* effects, since the graphs do not display any obvious dependence on the value of k (or on the eigenvalue itself) modulo small primes. However, a powerful result of Katznelson [43] may shed some light on the nature of this structure, as we discuss further in Section 3.10. Indeed, Katznelson's result could be used, with some effort, to recover the rational eigenvalue distribution of $M(k)$ for $n = 2$, or any larger value of n.

1.15 Sums of two squares

Having explored genuinely random phenomena in the previous four sections, we now turn our attention to a third and final arithmetic setting, which bears many similarities to the primes (and poses many of the same challenges). Let us write S for the set of integers which can be written as $x^2 + y^2$, the sum of two square integers. For concreteness we list the first sixteen terms of S below:

$$S = \{0, 1, 2, 4, 5, 8, 9, 10, 13, 16, 17, 18, 20, 25, 26, 29, \ldots\}.$$

It is apparent from this list that not all natural numbers belong to S. Indeed, since x^2 is always 0 or 1 modulo 4, no number of the form $4m + 3$ will appear. This *local* argument succinctly explains why 3 and 7 are not in S, but not why 21 is absent (this could be ruled out by a similar process modulo 9, but we will shortly give a more comprehensive reason).

Despite the seemingly arbitrary construction of S, it actually enjoys a rather rich history, owing to its surprisingly deep structure with roots at the heart of classical number theory. The third-century Greek mathematician Diophantus of Alexandria was the first to record the identity

$$(a^2 + b^2)(c^2 + d^2) = (ac - bd)^2 + (ad + bc)^2, \tag{1.8}$$

which establishes that S is *closed* under multiplication. However, the critical observation indicating the true nature of S was made by Pierre de Fermat[10] in 1640: *every prime of the form* $4n + 1$ *belongs to* S.

A rigorous proof of Fermat's claim did not appear until a century later by Euler (crediting Fermat for the necessary technique of *infinite descent*), who used it to characterize members of S by their prime decomposition: the natural number n belongs to S if and only if, in the prime factorization of n, every prime of the form $4m + 3$ appears to an *even* power. For instance, $7^4 \cdot 17 \cdot 23^2$ belongs to S (since both 7 and 23 have even exponents), while $7^4 \cdot 17 \cdot 23$ does not.

So even though S is constructed through addition, it nonetheless carries a highly multiplicative structure that is not entirely unlike the primes.[11] Indeed, Landau [48] used this very structure to establish a theorem analogous to the Prime Number Theorem, giving the precise density of S:

$$\#\{n \leq x : n \in S\} \sim C \frac{x}{\sqrt{\log x}}, \tag{1.9}$$

where the constant $C \approx 0.76422365$ is largely composed of local factors as in the Hardy–Littlewood conjecture. From this we see that the asymptotic density of S is equal to 0. The quantitative density $O(1/\sqrt{\log x})$ situates S (on a logarithmic scale) exactly halfway between the positive density $O(1)$ and the density of primes $O(1/\log x)$.

[10] Fermat and Diophantus share another connection: Fermat penned his so-called "Last Theorem" (famously proven by Wiles) in his copy of Diophantus's *Arithmetica*.

[11] One could, somewhat facetiously, describe the primes as those numbers n for which every prime less than n appears to a power less than 1.

1.16 Patterns in sums of two squares

We have observed some commonalities between S and the primes: they are roughly comparable in density, and their similar multiplicative structure suggests that a Cramér-like model would be successful in formulating predictions about S (tempered by the local factors that arise from Euler's characterization). It is natural to ask very similar questions about S that were raised about primes in Sections 1.2 and 1.3 (most of which are widely considered hopeless).

However, there is one obvious advantage enjoyed by S, which affords further progress on these combinatorial questions. Not only does S have an exclusive[12] multiplicative structure, it also has an *inclusive*, *additive* structure: by definition, we can show that $n \in S$ simply by finding two integers x and y such that $x^2 + y^2 = n$. There is no comparable way to construct a prime out of arbitrary integers.[13]

As a straightforward example, consider the "twin primes" problem for S: it is obvious that n^2 and $n^2 + 1$ both belong to S for any n, so S contains infinitely many patterns of type $\{m, m+1\}$. Even the longer pattern $\{m, m+1, m+2\}$ is easy to find: given any three consecutive numbers $n-1$, n, $n+1$ belonging to S with $n > 1$ (for instance $n = 9$), it is also true that

$$\{n^2 - 1, n^2, n^2 + 1\} \subset S.$$

The first inclusion follows from the factorization $(n-1)(n+1)$ and Diophantus's formula (1.8); the other two are obvious, as noted before. We will further discuss such *runs* of consecutive elements in Section 1.17, but for S there is no hope of extending this to runs of length 4: every fourth number has the form $4m + 3$, which is excluded.

Let us turn now to the question of *gaps* (spacings between consecutive terms of S). We have just witnessed infinitely many gaps of size 1 in S, so

[12] That is, we can prove certain integers do not belong to S based solely on, say, the power of 3 dividing them. A similar process of exclusion clearly holds for primes.

[13] While "prime-generating" formulae do exist, many are trivial curiosities; even the most interesting one, a polynomial constructed by Jones et al. [41], is not practical here.

we know that gaps can be arbitrarily small, and often. Conversely, it is not hard to see that gaps in S can be arbitrarily large: by Landau's asymptotic (1.9), the average size of the nth gap is roughly $1.3\sqrt{\log n}$. This leaves one existence question: *does every $d \in \mathbb{N}$ occur as a gap somewhere in S?* Spiro and Hensley [79] answered this, showing that every gap size does in fact occur, and infinitely often.

Note that we have constructed only $\sqrt{n}$ instances of the pattern $\{m, m+1\}$ up to n; however, random models predict that there should be about $n/\log n$ (ignoring the multiplicative constant). For the pattern $\{m, m+1, m+2\}$, the above construction gives only $\log\log n$ occurrences; a more efficient construction could improve this to $n^{1/4}$, but this is still a far cry from the expected $n/(\log n)^{3/2}$.

Much stronger results on S were obtained by Hooley [39], who has studied the combinatorics of S extensively. Using analytic techniques (cleverly drawing from sieve theory) he showed that the pattern $\{m, m+1\}$ occurs at least $\gg n/\log n$ times up to n. Since a similar upper bound $\ll n/\log n$ is also known by standard sieve methods, this establishes the correct *order of magnitude* for the frequency of twins $\{m, m+1\}$ in S, if not a precise asymptotic (with a known constant). Hooley also showed in [38] that the pattern $\{m, m+a, m+b\}$ occurs infinitely often in S, for any a and b.

So far, there seems to be greater hope of establishing patterns in S than in the prime case. It is ironic then, that for *arithmetic progressions* the primes have enjoyed far more success than S. For instance, the existence of arbitrarily long k-APs in S was not known until it became an immediate consequence of the Green–Tao theorem.[14] The expected asymptotic for the quantity of 3-APs in S remains (to our best knowledge) unproven, unlike van der Corput's theorem for primes. In general, precise asymptotics for S seem to be quite difficult — very recently, though, Matthiesen [56] has given a complete solution to a modified form of the k-APs problem (we discuss this modification in the conclusion).

[14] Green and Tao actually showed that a set containing any positive proportion of primes contains long progressions, and S contains about half of all primes.

1.17 Patterns in quadratic forms

We can more generally view S within a wider context: for $k \geq 1$, let $S(k)$ be the set of numbers of the form x^2+ky^2, so that S is equivalent to $S(1)$. This extension is not *entirely* arbitrary, in part due to an identity generalizing (1.8), found by the 7th century Indian mathematician Brahmagupta:

$$(a^2 + kb^2)(c^2 + kd^2) = (ac - kbd)^2 + k(ad + bc)^2. \tag{1.10}$$

In Chapter 4 we devote our attention wholly to $S(k)$, but we pause briefly to mention a truly fascinating related area of number theory. The theory of *binary quadratic forms* such as $x^2 + ky^2$ and more generally $ax^2 + bxy + cy^2$ blossomed in the late 18th century under Euler, Lagrange and Legendre, culminating in Gauss's 1801 masterwork *Disquisitiones Arithmeticae* [22]. While not all such forms possess a self-multiplication property[15] such as (1.10), Gauss uncovered a beautiful *group structure* which interconnects them all. Much more recently, Bhargava [7] has shed entirely new light on Gauss's composition laws, and applied these profound insights to solve some outstanding problems in algebraic number theory!

Let us return to the specific topic of $S(k)$. The density of $S(k)$ is very similar to that of S, having an asymptotic formula of the same form as (1.9), albeit with a different constant — here, the constant is more intricate, containing additional arithmetic factors (related to the work mentioned in the previous paragraph). This asymptotic was established by Bernays (then a student of Landau) in his 1912 doctoral thesis [6], and was considerably sharpened recently by Blomer and Granville [8].

The local considerations for $S(k)$ are rather dependent on the value of k. Let us discuss how this affects the two problems of *runs* and *gaps* from the previous section. While $S(1)$ does not contain more than 3 consecutive integers, $S(2)$ contains $\{0, 1, 2, 3, 4\}$, as well as $\{96, 97, 98, 99, 100\}$. On the other hand, one can show that $S(k)$ never admits any run of length 6; in this sense, such quintuples or 5-*runs* are of maximal interest for the $S(k)$

[15] However, a surprising consequence of Gauss's work is that the product of any *three* numbers of the form $ax^2 + bxy + cy^2$ can be expressed in the same form!

setting.[16] This raises the natural question: *for which values of k does $S(k)$ admit (infinitely many) consecutive quintuples?*.

We give a partial answer to this question, showing that $S(2)$ does contain infinitely many 5-runs, which answers a question of Moshe Rosenfeld (personal communication). More generally, we can show that $S(k)$ has the same property for several small values of k, as well as for an infinite sequence of distinct values k. However, this technique does not apply to all values of k: we do not know whether $S(34)$ has infinitely many 5-runs (heuristics suggest that it should).

For the analogous question of which gaps appear in $S(k)$, the only local obstruction occurs when k is divisible by 4 (in which case $S(k)$ only contains numbers of the form 0 or 1 modulo 4). For such k, it is impossible for $S(k)$ to contain gaps of size 2 or any number of the form $4d+2$. In Section 4.7 we fully generalize Spiro and Hensley's result to each $S(k)$: aside from this one class of exceptions, *every $d \in \mathbb{N}$ appears infinitely often as a gap in $S(k)$*.

[16] However, the form xy takes all possible values. While this example is rather trivial, it shows that local factors do not preclude, say, $30!(x^2 + y^2) + xy$ from having long runs.

Chapter 2

Simultaneous Approximation by Arithmetic Functions

2.1 Introduction

In this chapter we study the multidimensional approximation of points using values taken by arithmetic functions on arithmetic progressions, establishing generalizations of some theorems of Erdős and Schinzel [20], as well as Newman's result from Section 1.9. While it is multiplicative functions that comprise our main interest, these theorems are more naturally formulated in terms of the logarithms of such functions, which are additive. We will adopt the (non-standard) notation $\|\cdot\|$ from [20], which is defined by:

$$\|x\| := \begin{cases} x, & \text{if } |x| \le 1; \\ 1, & \text{if } |x| > 1. \end{cases}$$

We first give the formal statement of the Erdős–Schinzel theorem. Define $\mathcal{F}$ to be the class of all additive functions $f : \mathbb{N} \to \mathbb{R}$ satisfying the following:

Condition 2.1. *$f(n)$ is a real-valued additive function such that*

(a) $\sum_p \|f(p)\|^2/p$ is convergent;

(b) There exists a real $c_1 = c_1(f)$ such that, for any integer $M > 0$, the set of values $\{f(N) : (N, M) = 1\}$ is dense in (c_1, ∞);

(c) The partial sums of $\sum_p \|f(p)\|/p$ are bounded (note that the summand is not necessarily positive).

Erdős and Schinzel proved [20] that the following approximation theorem holds for any $f \in \mathcal{F}$:

Theorem 2.2 (Erdős–Schinzel)**.** *For any $h \in \mathbb{N}$ there exists c_h such that, for any $\epsilon > 0$ and every sequence of h numbers $\alpha_1, \alpha_2, \ldots, \alpha_h \geq c_h$, the set of natural numbers n which satisfy the simultaneous inequalities*

$$|f(n+i) - \alpha_i| < \epsilon \quad (i = 1, 2, \ldots, h) \tag{2.1}$$

has positive lower density.

They furthermore showed [20, p. 482] that any additive f for which Theorem 2.2 holds must also satisfy Condition 2.1, so that their result is best possible. It is worth comparing these hypotheses to the Erdős–Wintner conditions which determine the existence of a distribution function for $f(n)$. Those conditions, from (1.4), may be rephrased[17] in the present notation as

- $\sum_p \|f(p)\|^2/p$ is convergent;
- $\sum_p \|f(p)\|/p$ converges.

Relative to the above, we see that Condition 2.1 relaxes the requirement that $\sum_p \|f(p)\|/p$ converge to merely having bounded partial sums. At the same time, it adds a new requirement that the values of f remain dense after excluding certain congruence classes. One fairly common situation, which is sufficient to imply this denseness requirement, is that $f(p) \geq 0$, $f(p) \to 0$ and $\sum_p f(p)$ diverges (or that this is true on some subsequence of primes). In such case we may explicitly take $c_1 = 0$.

We note that if f is a positive multiplicative function for which $\log f \in \mathcal{F}$, then the conclusion of Theorem 2.2 holds equally well for f. The class $\mathcal{F}$ includes a wide range of such logarithmic functions, notably $\log \sigma(n) - \log n$ and $\log n - \log \phi(n)$ (we may take $c_1 = 0$ here, for the reason mentioned above). The theorem can therefore be used to approximate any tuple of sufficiently large reals by consecutive values of $\sigma(n)/n$, or sufficiently *small*

[17] We have suppressed the divergence requirement from (1.4), which is not needed for existence but only for *continuity* of the distribution function.

positive reals by $\phi(n)/n$. Theorem 2.2 also readily implies a similar approximation result for consecutive values of $f(n+1) - f(n)$, and it is easy to see that this yields results for $\sigma(n+1)/\sigma(n)$ and $\phi(n+1)/\phi(n)$; thus, Theorem 2.2 is a powerful generalization of earlier results by Schinzel and Wang [72]. We will return to this "relative" formulation in Section 2.3.

Shao had previously given [77] a generalization of Schinzel and Wang's results for ϕ and σ. However, this result placed additional assumptions on the growth rate of $f(p^r)$ as $p \to \infty$ for each $r \geq 1$. By contrast, Theorem 2.2 needs only bounds on $f(p)$, and it obtains a positive proportion of n rather than just an infinite set. As a cute example, any multiplicative function f with $f(p) = 1$ for all primes p, and each of $f(p^2) = 3$ and $f(p^2) = \frac{1}{2}$ for infinitely many p necessarily satisfies Condition 2.1, and therefore has the approximation property (2.1) for many values of n. (One could freely replace 3 and $\frac{1}{2}$ with any reals α and β such that $\frac{\log \alpha}{\log \beta}$ is a negative irrational.)

Our main theorem is a generalization of Theorem 2.2 to the case of multiple arbitrarily chosen functions from $\mathcal{F}$. The method of proof is essentially the same as that in [20], for which we give a self-contained exposition.

Theorem 2.3. *Let $f_1, f_2, \ldots, f_h \in \mathcal{F}$. Then there exists a constant c such that for any sequence of h numbers $\alpha_1, \alpha_2, \ldots, \alpha_h \geq c$, and $\epsilon > 0$, the set of natural numbers n satisfying the simultaneous inequalities*

$$|f_i(n+i) - \alpha_i| < \epsilon \quad (i = 1, 2, \ldots, h) \tag{2.2}$$

has positive lower density.

In Section 2.3, we generalize this result further: rather than consider the values of f at the consecutive integers $n+1, n+2, \ldots, n+h$, we may take (essentially) arbitrary linear functions in n. The technique is entirely elementary, and for expository purposes we choose to generalize a slightly different theorem from [20]; the application of the same technique to Theorem 2.3 is comparably straightforward.

It is natural to ask how the density of n satisfying (2.2), or the smallest such n, might vary as $\epsilon \to 0$. While estimates for these could be obtained from our construction, they are exponential in $1/\epsilon$, and the parameters depend on specific convergence properties of $f(p)$, which makes it difficult to give a reasonable bound that is uniform over $\mathcal{F}$. However, Alkan, Ford and Zaharescu [2] have recently proved a very sharp result along these lines, after imposing some reasonable regularity constraints on $f(p^r)$. They obtain an analogue of Theorem 2.3 (and its generalization to linear forms, Corollary 2.11) for infinitely many n, where ϵ may be replaced by n^{-c} for some positive constant c, depending only on h and the regularity parameters. This gives a growth rate of just $n \ll (1/\epsilon)^{1/c}$, on some sequence of $\epsilon \to 0$.

2.2 Main theorem

Let us first outline the basic strategy for Theorem 2.3: by restricting n to be divisible by $h!$, we have that $n + i$ is necessarily divisible by i (for $1 \le i \le h$), but we retain the freedom to choose finitely many (small) prime divisors for each $(n + i)/i$ independently. Condition 2.1b then allows us to approximate each target α_i by the contributions from these small prime divisors, by further restricting n to a particular congruence class. We then use Conditions 2.1a and 2.1c to show that the effect of the large primes on $f(n)$ is harmlessly small for a positive proportion of n within this class.

We make use of the following facts, which are well-known or elementary:

Proposition 2.4.

(a) The number of positive integers $n \le x$ which satisfy $n \equiv r \pmod{Q}$ and $d \mid n$ is $x/dQ + O(1)$, uniformly over all naturals d, Q, r where $(d, Q) = 1$.

(b) The number of positive integers $n \le x$ with at most two prime factors is $O(x \log\log x/\log x) = o(x)$.

We will also use a topological observation, essentially a restatement of the simple fact that a bounded subset of $\mathbb{R}^h$ is totally bounded. Note that the symbol $\|\cdot\|_\infty$ below refers to the usual supremum norm; it is unrelated to the $\|\cdot\|$ notation appearing in Condition 2.1.

Proposition 2.5. *Let $h \in \mathbb{N}$ and let $\{\mathbf{s}_n\}_{n>0}$ be a bounded sequence in $\mathbb{R}^h$. Then for any $\eta > 0$, there exists a finite index set $\mathcal{K} \subset \mathbb{N}$, such that for any $n \in \mathbb{N}$ there exists $k \in \mathcal{K}$ such that $k \leq n$ and $\|\mathbf{s}_n - \mathbf{s}_k\|_\infty < \eta$.*

Proof. Let $M > 0$ be an upper bound for $\|\mathbf{s}_n\|_\infty$ (uniform over all n). We can cover the set $\mathcal{S} = \{\mathbf{s}_n : n \in \mathbb{N}\}$ with finitely many open h-dimensional cubes $\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_m$ of side length η, such that $\mathcal{B}_i \cap \mathcal{S} \neq \varnothing$ for each $i = 1, 2, \ldots, m$ (we need at most $(1 + 2M/\eta)^h$ such cubes). We may then take $\mathcal{K} = \{k(i) : 1 \leq i \leq m\}$, where $k(i)$ is defined as the least k for which $\mathbf{s}_k \in \mathcal{B}_i$. For any $n \in \mathbb{N}$, we certainly have $\mathbf{s}_n \in \mathcal{B}_i$ for some i, whence $n \geq k(i)$ and $\|\mathbf{s}_n - \mathbf{s}_{k(i)}\|_\infty < \eta$ since $\mathbf{s}_n$ and $\mathbf{s}_{k(i)}$ lie in the same cube. ∎

Lastly, we require the following lemma, which appears as part of equation (21) in [20]. There is a minor flaw in the argument used there, so we include an independent proof for the sake of completeness.

Lemma 2.6. *There is an absolute constant $C > 0$ such that for any $x > 0$,*

$$\sum_{\sqrt{x} < p \leq x} \sum_{x/p < q < p} \frac{1}{pq} < C,$$

where the sums range over p, q prime.

Proof. We may assume without loss of generality that $x > 4$. Switching the order of summation, the above sum is equal to

$$\begin{aligned} &\sum_{\sqrt{x} < q < x} \frac{1}{q} \Bigg(\sum_{q < p \leq x} \frac{1}{p} \Bigg) + \sum_{q \leq \sqrt{x}} \frac{1}{q} \Bigg(\sum_{x/q < p \leq x} \frac{1}{p} \Bigg) \\ &\quad \leq \Bigg(\sum_{\sqrt{x} < p \leq x} \frac{1}{p} \Bigg)^2 + \sum_{q \leq \sqrt{x}} \frac{1}{q} \Bigg(\sum_{x/q < p \leq x} \frac{1}{p} \Bigg). \end{aligned} \tag{2.3}$$

Recalling Mertens' Theorem ([4, Lemma 4.10]) that for $x \geq 2$,

$$\sum_{p \leq x} \frac{1}{p} = \log\log x + B + O(\log^{-1} x),$$

we see that the first term of (2.3) is equal to $\log^2 2 + O(\log^{-1} x) = O(1)$, so it suffices to bound the second term. The inner sum is

$$\begin{aligned}\sum_{x/q<p\leq x} \frac{1}{p} &= \log\log x - \log\log \frac{x}{q} + O\bigg(\log^{-1} \frac{x}{q}\bigg) \\ &= -\log\left(1 - \frac{\log q}{\log x}\right) + O(\log^{-1} x) \ll \frac{\log q}{\log x}\end{aligned}$$

uniformly for $2 \leq q \leq \sqrt{x}$, so the second term of (2.3) is

$$\sum_{q\leq\sqrt{x}} \frac{1}{q}\bigg(\sum_{x/q<p\leq x} \frac{1}{p}\bigg) \ll \frac{1}{\log x} \sum_{q\leq\sqrt{x}} \frac{\log q}{q},$$

which is $O(1)$ by another Mertens estimate ([4, Lemma 4.8]). ■

Remark. The argument for the above lemma in [20] used an estimate of the form $\sum_{y<p\leq z} 1/p \ll \log\log z - \log\log y$, but this bound does not hold uniformly without an error term such as $O(\log^{-1} y)$ above to account for very short intervals $(y, z]$ containing primes.

Proof of Theorem 2.3. We take $c = \max\{c_1(f_i) + f_i(i)\}_{i=1}^h$, where $c_1(f_i)$ is defined as in Condition 2.1b. Let $\alpha_1, \alpha_2, \ldots, \alpha_h \geq c$ and $\epsilon > 0$ be fixed. By iterative application of Condition 2.1b, we can find pairwise co-prime integers $N_1, N_2, \ldots, N_h > 1$ such that

$$\begin{aligned}(N_i, h!) &= 1 \quad (i = 1, 2, \ldots, h), \\ |f_i(N_i) - \alpha_i + f_i(i)| &< \frac{\epsilon}{2} \quad (i = 1, 2, \ldots, h).\end{aligned} \tag{2.4}$$

Let $N = \prod_{i=1}^h N_i$ and let $k_1 > h$ be the largest prime factor of N. For any $k > k_1$, we can define

$$P_k := \prod_{\substack{h<p\leq k \\ p\nmid N}} p, \qquad Q_k := h!^2 P_k N^2.$$

By the Chinese remainder theorem, the system of congruences

$$n \equiv 0 \pmod{h!^2 P_k}, \qquad n \equiv -i + N_i \pmod{N_i^2} \quad (i = 1, 2, \ldots, h) \tag{2.5}$$

has a unique solution $n \equiv n_0 \pmod{Q_k}$. Let $\mathcal{A}_k$ denote this congruence class. For any $n \in \mathcal{A}_k$, $n + i$ is congruent to i mod i^2, and to N_i mod N_i^2; hence the quantity $(n+i)/iN_i$ is an integer relatively prime to iN_i. Note furthermore that $n + i$ is non-zero mod p for any $p \le h$ with $(p, i) = 1$, for any $p \mid P_k$, and for any $p \mid N_j$ with $j \ne i$. It follows that $(n+i)/iN_i$ is not divisible by any prime $p \le k$, and so

$$f_i(n+i) - f_i(iN_i) = \sum_{\substack{p^r \| n+i \\ p>k}} f_i(p^r). \tag{2.6}$$

We wish to choose k so that this quantity has magnitude at most $\epsilon/2$ for a positive proportion of $n \in \mathcal{A}_k$. Then by (2.4) we will have $|f_i(n+i) - \alpha_i| < \epsilon$ as desired.

We define the following parameters (here C is the constant in Lemma 2.6):

$$\mu := \min(1, \epsilon/\sqrt{96Ch}), \qquad \eta := \epsilon/\sqrt{48h}. \tag{2.7}$$

Let $\mathcal{P}$ denote the set of primes p such that $|f_i(p)| \le \mu$ for $i = 1, 2, \ldots, h$. For convenience we use the notation $\mathcal{P}(k)$ for the subset $\{p \in \mathcal{P} : p > k\}$.

By Condition 2.1a, the sum $\sum_{p \notin \mathcal{P}} 1/p \le \mu^{-2} \sum_{i=1}^{h} \sum_p \|f_i(p)\|^2/p$ must converge. As $\sum_p 1/p^2$ also converges, there is an integer $k_2 \ge k_1$ such that

$$\sum_{p>k_2,\, p\notin\mathcal{P}} \frac{1}{p} + \sum_{p>k_2} \frac{1}{p^2} < \frac{1}{3h}. \tag{2.8}$$

For any $k > k_2 > h$, let $\mathcal{N}_k$ denote the subset of $\mathcal{A}_k$ consisting all $n \in \mathcal{A}_k$ with the following property: $(n+1)(n+2)\cdots(n+h)$ is not divisible by any prime $p > k$ outside of $\mathcal{P}$, nor is it divisible by p^2 for any prime $p > k$ (here and below, P_k, Q_k and $\mathcal{A}_k$ remain as constructed above).

We estimate the counting function $\mathcal{N}_k(x) = \#\{n \le x : n \in \mathcal{N}_k\}$. The counting function of $\mathcal{A}_k$ is clearly $\mathcal{A}_k(x) = x/Q_k + O(1)$. If $n \le x$ and $n \in \mathcal{A}_k \setminus \mathcal{N}_k$, then $n+i$ (for some $1 \le i \le h$) must be divisible by some prime $p \notin \mathcal{P}$ with $k < p \le x+i$, or by p^2 for some $k < p \le \sqrt{x+i}$. For any fixed p and i, Proposition 2.4a gives exactly $x/pQ_k + O(1)$ values of $n \le x$ where $p \mid n+i$, and $x/p^2Q_k + O(1)$ values of $n \le x$ where $p^2 \mid n+i$. Thus

$$\begin{aligned}\mathcal{A}_k(x) - \mathcal{N}_k(x) &\le \sum_{i=1}^{h} \left(\sum_{\substack{k<p\le x+i \\ p\notin\mathcal{P}}} \left(\frac{x}{pQ_k} + O(1) \right) + \sum_{k<p\le\sqrt{x+i}} \left(\frac{x}{p^2Q_k} + O(1) \right) \right) \\ &\le \frac{hx}{Q_k} \left(\sum_{p>k,\, p\notin\mathcal{P}} \frac{1}{p} + \sum_{p>k} \frac{1}{p^2} \right) + O\left(\sum_{p\le x+h} h \right) \le \frac{x}{3Q_k} + o(x).\end{aligned}$$

So $\mathcal{N}_k(x) \ge \frac{2}{3}x/Q_k + o(x)$; in other words, $\mathcal{N}_k$ has lower density at least $\frac{2}{3}$ relative to the progression $\mathcal{A}_k$. For any $n \in \mathcal{N}_k$, the integer $(n+i)/iN_i$ is a squarefree product of primes p with $p > k$ and $p \in \mathcal{P}$. Thus equation (2.6) restricted to $n \in \mathcal{N}_k$ reduces to

$$f_i(n+i) - f_i(iN_i) = \sum_{\substack{p\mid n+i \\ p\in\mathcal{P}(k)}} f_i(p). \tag{2.9}$$

We will show that for x sufficiently large, we can choose $k = k(x)$ so that

$$E(x) = \sum_{n\in\mathcal{N}_k,\, n\le x} \sum_{i=1}^{h} (f_i(n+i) - f_i(iN_i))^2 < \frac{\epsilon^2}{12}\frac{x}{Q_k} + o_k(x), \tag{2.10}$$

and thus there are at most $\frac{1}{3}x/Q_k + o_k(x)$ values of $n \in \mathcal{N}_k$ up to x for which $|f_i(n+i) - f_i(iN_i)| \ge \frac{\epsilon}{2}$ for some $i = 1, 2, \ldots, h$. By the previously estimated density of $\mathcal{N}_k$ there are at least $\frac{1}{3}x/Q_k + o_k(x)$ values of $n \le x$ satisfying inequality (2.2). This will be sufficient to prove Theorem 2.3, provided that we have a uniform upper bound on k (hence also on Q_k) independent of x: we are free to vary k with x provided that we restrict the choice of $k(x)$ to within a *finite* set of possible values, so that the (lower) density of $\mathcal{N}_k$ is bounded below by a positive quantity.

By Condition 2.1a, there exists an integer $k_3 \geq k_2$ such that

$$\sum_{i=1}^{h} \sum_{p \in \mathcal{P}(k_3)} \frac{f_i(p)^2}{p} \leq \sum_{i=1}^{h} \sum_{\substack{p > k_3 \\ |f_i(p)| < \mu}} \frac{f_i(p)^2}{p} < \frac{\epsilon^2}{24}. \tag{2.11}$$

We have by Condition 2.1c that the partial sums of $\sum_p \|f_i(p)\|/p$ are bounded, and since $\sum_{p \notin \mathcal{P}} \|f_i(p)\|/p$ converges absolutely by Condition 2.1a, the partial sums of $\sum_{p \in \mathcal{P}} \|f_i(p)\|/p$ are also bounded. We denote these partial sums by $S_{k,i}$, that is

$$S_{k,i} := \sum_{p \in \mathcal{P},\, p \leq k} \frac{\|f_i(p)\|}{p} = \sum_{p \in \mathcal{P},\, p \leq k} \frac{f_i(p)}{p}.$$

Applying Proposition 2.5 to the sequence of h-tuples $\{(S_{k,i})_{i=1}^{h}\}_{k>k_3}$, we see that there exists a finite index set $\mathcal{K} \subset \mathcal{P}(k_3)$ such that for any $x > k_3$ there exists $k = k(x) \in \mathcal{K}$ such that $k \leq x + h$ and for all $i = 1, 2, \ldots, h$,

$$\left| S_{\lfloor x+h \rfloor, i} - S_{k,i} \right| = \left| \sum_{\substack{k < p \leq x+h \\ p \in \mathcal{P}}} \frac{f_i(p)}{p} \right| < \eta. \tag{2.12}$$

As $\mathcal{K}$ is finite, we have $k = k(x) \leq \max \mathcal{K}$ uniformly for $x > k_3$, as required by the previous discussion. We now estimate the sum in (2.10) for this choice of k. By (2.9) we have

$$\begin{aligned} E(x) = \sum_{\substack{n \in \mathcal{N}_k \\ n \leq x}} \sum_{i=1}^{h} (f_i(n+i) - f_i(iN_i))^2 &= \sum_{i=1}^{h} \sum_{\substack{n \in \mathcal{N}_k \\ n \leq x}} \Bigg(\sum_{\substack{p \mid n+i \\ p \in \mathcal{P}(k)}} f_i(p) \Bigg)^2 \\ \leq \sum_{i=1}^{h} \sum_{\substack{n \in \mathcal{A}_k \\ n \leq x+h-i}} \Bigg(\sum_{\substack{p \mid n+i \\ p \in \mathcal{P}(k)}} f_i(p)^2 &+ 2 \sum_{\substack{pq \mid n+i \\ p,q \in \mathcal{P}(k) \\ p > q}} f_i(p) f_i(q) \Bigg). \end{aligned}$$

Changing the order of summation and applying Proposition 2.4a gives

$$\begin{aligned} E(x) &\le \sum_{i=1}^{h} \left(\sum_{\substack{p \le x+h \\ p \in \mathcal{P}(k)}} f_i(p)^2 \left(\frac{x}{pQ_k} + O(h) \right) + 2 \sum_{\substack{pq \le x+h \\ p,q \in \mathcal{P}(k)}} f_i(p) f_i(q) \left(\frac{x}{pqQ_k} + O(h) \right) \right) \\ &= \frac{x}{Q_k} \Bigg(\sum_{i=1}^{h} \sum_{\substack{p \le x+h \\ p \in \mathcal{P}(k)}} \frac{f_i(p)^2}{p} + 2 \sum_{i=1}^{h} \sum_{\substack{pq \le x+h \\ p,q \in \mathcal{P}(k) \\ p>q}} \frac{f_i(p) f_i(q)}{pq} \Bigg) \\ &\qquad + O\Bigg(\sum_{i=1}^{h} \sum_{p \le x+h} h\mu^2 + \sum_{i=1}^{h} \sum_{pq \le x+h} h\mu^2 \Bigg). \end{aligned}$$

We apply (2.11) to the first double sum and Proposition 2.4b to the error term to obtain

$$E(x) < \frac{x}{Q_k} \Bigg(\frac{\epsilon^2}{24} + 2h \max_{i \le h} \sum_{\substack{pq \le x+h \\ p,q \in \mathcal{P}(k) \\ p>q}} \frac{f_i(p) f_i(q)}{pq} \Bigg) + o(x).$$

To prove (2.10) it suffices to show that for each $i = 1, 2, \ldots, h$,

$$2 \sum_{\substack{pq \le x+h \\ p,q \in \mathcal{P}(k) \\ p>q}} \frac{f_i(p) f_i(q)}{pq} < \frac{\epsilon^2}{24h}. \tag{2.13}$$

By equation (2.12) we have for any $1 \le i \le h$,

$$2 \sum_{\substack{q<p \le x+h \\ p,q \in \mathcal{P}(k)}} \frac{f_i(p) f_i(q)}{pq} \le \Bigg(\sum_{\substack{p \le x+h \\ p \in \mathcal{P}(k)}} \frac{f_i(p)}{p} \Bigg)^2 < \eta^2 = \frac{\epsilon^2}{48h}.$$

Comparing the sum on the left-hand side to the sum in (2.13), we have

$$\begin{aligned}
2 \sum_{\substack{pq \le x+h \\ p,q \in \mathcal{P}(k) \\ p>q}} \frac{f_i(p) f_i(q)}{pq} &< \frac{\epsilon^2}{48h} + 2 \sum_{\substack{pq > x+h \\ p,q \in \mathcal{P}(k) \\ q<p \le x+h}} \frac{\mu^2}{pq} \\
&= \frac{\epsilon^2}{48h} + 2\mu^2 \sum_{\substack{\sqrt{x+h}<p \le x+h \\ p \in \mathcal{P}(k)}} \sum_{\substack{(x+h)/p<q<p \\ q \in \mathcal{P}(k)}} \frac{1}{pq} \\
&< \frac{\epsilon^2}{48h} + 2\mu^2 C = \frac{\epsilon^2}{24h}
\end{aligned}$$

by Lemma 2.6. This completes the proof of Theorem 2.3. ■

2.3 Generalization to linear functions

As we noted in the introduction, one can use Theorem 2.2 to simultaneously approximate arbitrary reals by the differences $f(n+i) - f(n+i-1)$ rather than the absolute values $f(n+i)$. Note that by considering these relative differences we no longer need any lower bound on the targets α_i. Erdős and Schinzel prove in [20] that this formulation actually holds under a significantly weaker hypothesis than Theorem 2.2, specifically:

Theorem 2.7 (Erdős–Schinzel)**.** *Let $f(n)$ be an additive function satisfying Conditions 2.1a and 2.1b. Then, for any given sequence of h real numbers $\alpha_1, \alpha_2, \ldots, \alpha_h$ and $\epsilon > 0$, the set of natural numbers n satisfying the simultaneous inequalities*

$$|f(n+i) - f(n+i-1) - \alpha_i| < \epsilon \quad (i = 1, 2, \ldots, h) \tag{2.14}$$

has positive lower density.

For instance, any additive function f with $f(p) = \log^{-1} p$ satisfies Theorem 2.7 (again taking $c_1(f) = 0$ in Condition 2.1b) but not Condition 2.1c. The elimination of Condition 2.1c relies on a delicate cancellation between values of f, so their proof does not readily adapt to multiple functions f_i

satisfying this weaker condition. However, it can be generalized in a different direction: we may replace the consecutive integers $n, n+1, \ldots, n+h$, at which f is evaluated, by $h+1$ (non-constant) linear functions of n, say

$$a_0n + b_0,\ a_1n + b_1,\ \ldots,\ a_hn + b_h \quad (a_i \in \mathbb{N}, b_i \in \mathbb{Z}). \tag{2.15}$$

We need to impose one condition on the coefficients a_i and b_i, which is that no two forms $a_in + b_i$ and $a_jn + b_j$ may divide one another as polynomials. This requirement is natural as we cannot expect, say, $\phi(3n+6)$ and $\phi(2n+4)$ to approximate independent reals (indeed, the function $\phi(3n+6)/\phi(2n+4)$ takes only four distinct values). Thus we add the constraint that

$$a_ib_j \neq a_jb_i \quad (0 \leq i < j \leq h), \tag{2.16}$$

whereupon the following theorem holds:

Theorem 2.8. *Let $h \in \mathbb{N}$ and let $(a_in + b_i)_{i=0}^{h}$ be linear forms satisfying* (2.15) *and* (2.16)*. Then Theorem 2.7 holds with* (2.14) *replaced by*

$$|f(a_in + b_i) - f(a_{i-1}n + b_{i-1}) - \alpha_i| < \epsilon \quad (i = 1, 2, \ldots, h). \tag{2.17}$$

We will see that Theorem 2.8 actually follows by an elementary argument from Theorem 2.7. We first show that Theorem 2.7 still holds when n is restricted to an arbitrary arithmetic progression. We remark that Erdős and Schinzel's techniques were strong enough to show positive relative density even in the case that n is restricted to the set of primes!

We make use of the following observation which is immediate upon inspection of the proof of Theorem 6 [20, p. 477]. There is a minor flaw in the construction which is immaterial to our proposition, and is easily repaired by replacing $(h+1)!$ with $(h+1)!^2$, similar to our equation (2.5).

Proposition 2.9. *The conclusion of Theorem 2.7 holds when n is restricted to the congruence class $n \equiv 1 \pmod{(h+1)!}$.*

We note that this specific congruence class, just as in the proof of Theorem 2.3, is not essential to the construction and was used only as a means of simplifying the bookkeeping needed in the argument. Indeed, the preceding proposition readily generalizes to the following lemma.

Lemma 2.10. *Let $f(n)$ be an additive function satisfying Conditions 2.1a and 2.1b. Then for any $a \in \mathbb{N}$ and $b \in \mathbb{Z}$, the set of $n \equiv b \pmod{a}$ for which n satisfies* (2.14) *also has positive lower density.*

Proof. Without loss of generality, we may assume that $b \geq a$, so that $a \mid (h+b)!$. Now, let $h' = h + b - 1$ and apply Theorem 2.7 to the h'-tuple $(\alpha'_1, \alpha'_2, \ldots, \alpha'_{h'})$, where $\alpha'_{i+b-1} = \alpha_i$ and we are free to choose α'_i arbitrarily for $i < b$. By Proposition 2.9, any n' satisfying (2.14) for the tuple (α'_i) is congruent to 1 mod $(h+b)!$, hence $n = n' + b - 1$ satisfies (2.14) for the requisite h-tuple $(\alpha_1, \ldots, \alpha_h)$ as well as the modular constraint. ■

Our strategy for proving Theorem 2.8 is simple: we first transform the problem into a specific instance of Lemma 2.10, and then we correct for the discrepancy introduced by the transformation.

Proof of Theorem 2.8. Let A be the lowest common multiple of $a_0, a_1, \ldots, a_h$. By making a substitution of the form $n \to n + k$, we may assume that each $b_i > 0$. We multiply each linear form $a_i n + b_i$ by an appropriate integer to obtain the form $An + B_i$ with $B_i > 0$. Condition (2.16) implies that the natural numbers $B_0, \ldots, B_h$ are distinct. By reordering the a_i and b_i in condition (2.17) and making corresponding adjustments to the values $\alpha_1, \alpha_2, \ldots, \alpha_h$ and ϵ, we may assume that the linear forms $(a_i n + b_i)_{i=0}^h$ are ordered so that $B_0 < B_1 < \cdots < B_h$.

It follows from Lemma 2.10, using $h' = B_h$ and the congruence class $n' \equiv 0 \pmod{A}$, that for any h-tuple $(\alpha'_1, \ldots, \alpha'_h)$, the set of n such that

$$|f(An + B_i) - f(An + B_{i-1}) - \alpha'_i| < \epsilon \quad (i = 1, 2, \ldots, h) \tag{2.18}$$

has positive lower density. This is clear when B_{i-1} and B_i are consecutive integers; if $\Delta B_i = B_i - B_{i-1} > 1$, then we are free to assign arbitrary targets to the values of $f(An + B_{i-1} + j) - f(An + B_{i-1} + j - 1)$ for $1 \leq j \leq \Delta B_i$, provided that the targets sum to α'_i, while replacing ϵ by $\epsilon/\Delta B_i$.

It remains to account for the discrepancy between $f(An+B_i)$ and $f(a_in+b_i)$. Since $An+B_i = M_i(a_in+b_i)$ for some integer $M_i \mid A$, and f is additive, the value of $f(An+B_i) - f(a_in+b_i)$ is precisely determined by the p-adic valuations $v_p(An+B_i)$ and $v_p(a_in+b_i)$ for just the finitely many primes p dividing M_i.

In particular, if we restrict n to a suitable congruence class so that $v_p(An) > v_p(B_i)$ for each prime p dividing A and all $0 \le i \le h$, then $v_p(An+B_i)$ takes the constant value $v_p(B_i)$ and likewise $v_p(a_in+b_i)$ takes the constant value $v_p(b_i)$, so that

$$f(An+B_i) - f(a_in+b_i) = f(B_i) - f(b_i).$$

A second appeal to Lemma 2.10 allows us to do just that, and so Theorem 2.8 holds by taking $\alpha_i' = \alpha_i + f(B_i) - f(b_i) - f(B_{i-1}) + f(b_{i-1})$ in (2.18). ■

The method used to prove Theorem 2.8 is easily adapted to generalize Theorem 2.3, using congruence (2.5) in place of Proposition 2.9, to obtain the corollary below. Note that the constant c now also depends on the coefficients (a_i, b_i).

Corollary 2.11. *Let $f_1, f_2, \ldots, f_h \in \mathcal{F}$, and let $(a_in+b_i)_{i=1}^h$ be linear forms satisfying* (2.15) *and* (2.16) *(ignoring a_0 and b_0). Then Theorem 2.3 holds with* (2.2) *replaced by*

$$|f_i(a_in+b_i) - \alpha_i| < \epsilon \quad (i = 1, 2, \ldots, h). \tag{2.19}$$

Chapter 3

Eigenvalues of Random Integer Matrices

3.1 Introduction

In this chapter we conduct our investigation of the eigenvalues of random integer matrices, in response to the question posed by Hetzel, Liew and Morrison [37]: *what is the probability that a random $n \times n$ integer matrix is diagonalizable over the field of rational numbers?*

Since there is no uniform probability distribution on $\mathbb{Z}$, let us clarify the meaning of this question. For $k \geq 1$, let $\mathcal{I}_k := \{-k, -k+1, \ldots, k-1, k\}$ be the set of integers of height at most k. Since $|\mathcal{I}_k| = 2k+1$ is finite, we can construct a random $n \times n$ matrix M where each entry is chosen independently and uniformly at random from $\mathcal{I}_k$. The probability that M has any given property (such as being diagonalizable over $\mathbb{Q}$) is then a function of the parameter k. Our aim is to determine how this function behaves as $k \to \infty$; in particular, does it approach a limiting value as $k \to \infty$?

For any given integers $n \geq 2$ and $k \geq 1$, the set of random $n \times n$ matrices with entries in $\mathcal{I}_k$ forms a well-defined (finite) probability space; it will be convenient to compute probabilities just by counting matrices, so we introduce some notation for certain sets of matrices. Let $\mathcal{M}_n(k)$ be the set of all $n \times n$ matrices whose entries are in $\mathcal{I}_k$; we are choosing matrices uniformly from $\mathcal{M}_n(k)$, a set with cardinality $(2k+1)^{n^2}$. The probability that a random matrix in $\mathcal{M}_n(k)$ satisfies a particular property is given simply by the number of matrices in $\mathcal{M}_n(k)$ having that property, divided by $(2k+1)^{n^2}$.

For a given integer λ, let $\mathcal{M}_n^\lambda(k)$ denote the set of all matrices in $\mathcal{M}_n(k)$ that have λ as an eigenvalue. In particular, $\mathcal{M}_n^0(k)$ is the subset of singular matrices in $\mathcal{M}_n(k)$. Likewise, we denote the set of matrices in $\mathcal{M}_n(k)$ having at least one integer eigenvalue by $\mathcal{M}_n^{\mathbb{Z}}(k) = \bigcup_{\lambda\in\mathbb{Z}} \mathcal{M}_n^\lambda(k)$. The probability that a random matrix in $\mathcal{M}_n(k)$ has an integer eigenvalue is thus $|\mathcal{M}_n^{\mathbb{Z}}(k)|/(2k+1)^{n^2}$.

Our first result in this chapter affirms and strengthens the conjecture made by Hetzel et al.: *for any $n \geq 2$, the probability that a random $n \times n$ integer matrix has at least one integer eigenvalue tends to 0 as $k \to \infty$.* We give a quantitative upper bound on this probability:

Theorem 3.1. *Given any integer $n \geq 2$ and any real number $\varepsilon > 0$, the probability that a randomly chosen matrix in $\mathcal{M}_n(k)$ has any integer eigenvalues is $\ll_{n,\varepsilon} 1/k^{1-\varepsilon}$. Consequently, the probability that a randomly chosen matrix in $\mathcal{M}_n(k)$ is diagonalizable over $\mathbb{Q}$ is $\ll_{n,\varepsilon} 1/k^{1-\varepsilon}$.*

Given an integer matrix $M \in \mathcal{M}_n(k)$, a necessary (and nearly sufficient) condition for it to be diagonalizable over $\mathbb{Q}$ is that all of its eigenvalues be rational. Moreover, since the characteristic polynomial $\det(\lambda I - M)$ is monic with integer coefficients, the familiar "rational roots theorem" [64, §4.3] implies that every rational eigenvalue of M must be an integer. Hence any integer matrix that is diagonalizable over the rationals must be contained in $\mathcal{M}_n^{\mathbb{Z}}(k)$, and so the second assertion of the theorem follows immediately from the first.

The special case $n = 2$ of Theorem 3.1 was previously obtained in [37] and also earlier by Kowalsky [47]. Unravelling the $\ll$-notation, the theorem states that there exists a constant C, depending on n and ε, such that $|\mathcal{M}_n^{\mathbb{Z}}(k)|/|\mathcal{M}_n(k)| \leq C/k^{1-\varepsilon}$ for all $k \geq 1$. Note that $|\mathcal{M}_n(k)| \ll_n k^{n^2}$ (with the implied constant being highly dependent on n), and so the theorem also gives an upper bound for the number of matrices in $\mathcal{M}_n(k)$ having at least one integer eigenvalue, that is

$$|\mathcal{M}_n^{\mathbb{Z}}(k)| \ll_{n,\varepsilon} k^{n^2-1+\varepsilon}.$$

The key tool used to establish Theorem 3.1 is the following related estimate for $|\mathcal{M}_n^0(k)|$, the number of singular matrices in $\mathcal{M}_n(k)$:

Lemma 3.2. *Given any integer $n \ge 2$ and any real number $\varepsilon > 0$, the probability that a random matrix in $\mathcal{M}_n(k)$ is singular is $\ll_{n,\varepsilon} 1/k^{2-\varepsilon}$. In other words, $|\mathcal{M}_n^0(k)| \ll_{n,\varepsilon} k^{n^2-2+\varepsilon}$.*

This bound is far from optimal in general (as we discuss in Section 3.10), but it is essentially best possible when $n = 2$: in the latter part of this chapter, we conduct a detailed investigation of the distribution of eigenvalues for the 2×2 case, in particular obtaining the precise asymptotic $|\mathcal{M}_2^0(k)| = \frac{96}{\pi^2} k^2 \log k + O(k^2)$. More generally, we determine the asymptotic incidence of all integer eigenvalues of $\mathcal{M}_2(k)$, as given by the following theorem:

Theorem 3.3. *Define the function $V : [-2,2] \to \mathbb{R}$ by $V(-\delta) = V(\delta)$ and*

$$V(\delta) = \begin{cases} 4 - 2\delta - \delta^2 + \delta^2 \log(1+\delta) - 2(1-\delta)\log(1-\delta), & \text{if } 0 \le \delta < 1, \\ 1 + \log 2, & \text{if } \delta = 1, \\ 4 - 2\delta - \delta^2 + \delta^2 \log(\delta+1) + 2(\delta-1)\log(\delta-1), & \text{if } 1 < \delta \le \sqrt{2}, \\ \delta^2 - 2\delta - (\delta^2 - 2\delta + 2)\log(\delta - 1) & \text{if } \sqrt{2} < \delta \le 2. \end{cases} \tag{3.1}$$

Then for any integer λ between $-2k$ and $2k$,

$$|\mathcal{M}_2^\lambda(k)| = \frac{24V(\lambda/k)}{\pi^2} k^2 \log k + O(k^2), \tag{3.2}$$

where the implied constant is absolute. If $|\lambda| > 2k$ then $\mathcal{M}_2^\lambda(k)$ is empty.

We remark that the function $V(\delta)$ is continuous and, with the exception of the points of infinite slope at $\delta = \pm 1$, differentiable everywhere (even at $\delta = \pm 2$, if we further define $V(\delta) \equiv 0$ for $|\delta| > 2$). Technically, equation (3.2) is not an asymptotic formula when λ is extremely close to $\pm 2k$, because then the value of $V(\lambda/k)$ can have order of magnitude $1/\log k$ or smaller, making the "main term" no bigger than the error term. However, it truly is an asymptotic formula for $|\lambda| < 2k - \psi(k)k/(\log k)^{1/3}$, where $\psi(k)$ is any function tending to infinity (the exponent $1/3$ arises because $V(\delta)$ approaches 0 cubically as δ tends to 2 from below).

By summing (3.2) over all possible values of λ, we obtain an asymptotic formula for $|\mathcal{M}_2^\lambda(k)|$, which gives a sharp quantitative answer to the original question, in the special case $n = 2$. We defer the details of this calculation to Section 3.6.

Corollary 3.4. *Let* $C = \big(7\sqrt{2} + 4 + 3\log(\sqrt{2}+1)\big)/3\pi^2 \approx 0.55873957$*. The probability that a random matrix in* $\mathcal{M}_2(k)$ *has integer eigenvalues is asymptotically equal to* $C(\log k)/k$*. More precisely,*

$$|\mathcal{M}_2^{\mathbb{Z}}(k)| = 16Ck^3 \log k + O(k^3). \tag{3.3}$$

If $M \in \mathcal{M}_2(k)$ has eigenvalue λ, then the scaled matrix $k^{-1}M$ has eigenvalue λ/k, which is the argument of V that appears on the right-hand side of (3.2). A natural probabilistic interpretation of Theorem 3.3 is that for large k, the rational eigenvalues of $k^{-1}M$ are distributed according to the density given by V.

Note that the entries of $k^{-1}M$ are sampled from a discrete, evenly-spaced subset of $[-1,1]$. As $k \to \infty$ the random variable describing each entry converges weakly to the uniform distribution on the interval $[-1,1]$. Let $\mathcal{M}_2([-1,1])$ be the probability space of all 2×2 matrices whose entries are independent random variables drawn from this distribution. One might ask whether the distribution given by Theorem 3.3 merely approximates the continuous distribution of (real) eigenvalues in $\mathcal{M}_2([-1,1])$; the answer, perhaps surprisingly, is no, as we establish by also computing this latter distribution explicitly.

Theorem 3.5. *If* M *is chosen uniformly from* $\mathcal{M}_2([-1,1])$*, the expected number of eigenvalues of* M *in the interval* $[s,t]$ *is* $\int_s^t W(\delta)\,d\delta$*, where* $W(\delta)$

is the density function given by $W(-\delta) = W(\delta)$ *and*

$$W(\delta) = \begin{cases} (80 + 20\delta + 90\delta^2 + 52\delta^3 - 107\delta^4)/(144(1+\delta)) \\ \qquad - (5 - 7\delta + 8\delta^2)(1-\delta)\log(1-\delta)/12 \\ \qquad - \delta(1-\delta^2)\log(1+\delta)/4, & \text{if } 0 \le \delta \le 1, \\ \delta(20 + 10\delta - 12\delta^2 - 3\delta^3)/(16(1+\delta)) \\ \qquad + (3\delta - 1)(\delta - 1)\log(\delta - 1)/4 \\ \qquad + \delta(\delta^2 - 1)\log(\delta + 1)/4, & \text{if } 1 \le \delta \le \sqrt{2}, \\ \delta(\delta - 2)(2 - 6\delta + 3\delta^2)/(16(\delta - 1)) \\ \qquad - (\delta - 1)^3\log(\delta - 1)/4, & \text{if } \sqrt{2} \le \delta \le 2, \\ 0, & \text{if } \delta \ge 2. \end{cases} \tag{3.4}$$

Just like $V(\delta)$, the function $W(\delta)$ is continuous and differentiable everywhere, again with the exception of two points of infinite slope at $\delta = \pm 1$. (The value $W(1) = \frac{15}{32}$ makes the function continuous there, although the value of a density function at any one point is irrelevant to the distribution.) It also shares the same cubic decay as δ tends to 2 from below. But there are also striking qualitative differences between the functions V and W. In Figure 3.1 we plot these on the same axes, normalized[18] so that the area under each curve is 2. In the case of $\mathcal{M}_2(k)$, this normalization corresponds to conditioning on having integer eigenvalues: that is, scaling V by the probability $C(\log k)/k$ from Corollary 3.4. For $\mathcal{M}_2([-1,1])$ we are conditioning on having real eigenvalues, an event which occurs with probability $\frac{49}{72}$ (this can be obtained by integrating $W(\delta)$, analogously to the proof of Corollary 3.4, or by a direct integral computation which was done in [37]).

Notice that the distribution $W(\delta)$ is bimodal, having maxima at $\delta \approx \pm 0.75030751$. Thus, a random matrix in $\mathcal{M}_2([-1,1])$ is more likely to have an eigenvalue of near $\frac{3}{4}$ than one near 0. That the opposite is true for $V(\delta)$ shows that the eigenvalue distribution of $\mathcal{M}_2^{\mathbb{Z}}(k)$ is not purely the

[18] We normalize to area 2 since a 2 × 2 matrix has two eigenvalues. One could view this as the sum of respective probability densities for the upper and lower eigenvalues.

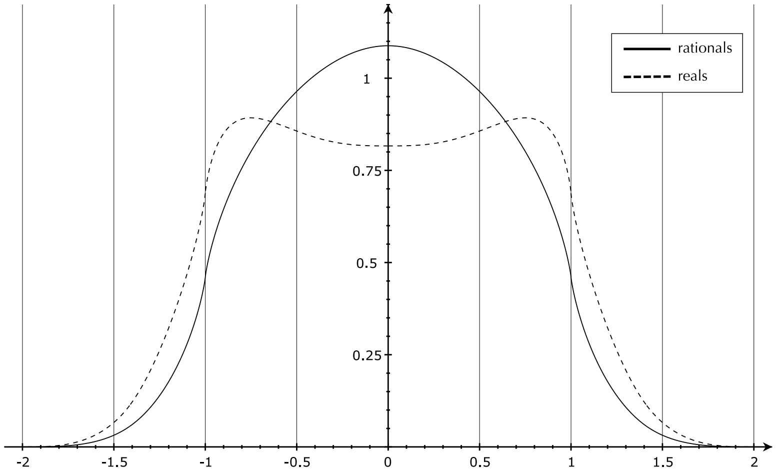

Figure 3.1: Distribution of real and rational eigenvalues for $n = 2$

result of magnitude considerations, but also encodes some of the arithmetic structure of the integers up to k. It is true, however, that if we sample the *real* eigenvalues of matrices in $\mathcal{M}_2(k)$, then these do converge to the exact distribution given by $W(\delta)$: eigenvalues vary continuously with the matrix coefficients, so this sampling limit behaves much like a Riemann sum. We may thus view $V(\delta)$ and $W(\delta)$ as the distributions of integer and real eigenvalues of $\mathcal{M}_2(k)$ for large k.

3.2 Basic bounds on singularity

It is not hard to show a simpler version of Lemma 3.2, with a singularity bound that is qualitatively weaker but more explicit and also uniform in n:

Proposition 3.6. *The probability that a random matrix in $\mathcal{M}_n(k)$ is singular is at most* $1/2k$*.*

To see this, we need a lemma which bounds the size of intersection between a linear subspace and a discrete hypercube in $\mathbb{R}^n$. With minor modifications this previously appeared as problem B4 on the 67th William Lowell Putnam Mathematical Competition [44]. The elegant proof below is due to Catalin Zara.

Lemma 3.7. *Let V be a d-dimensional subspace of $\mathbb{R}^n$, and S be any finite set of m real numbers. Then $|S^n \cap V| \leq m^d$.*

Proof. Form a matrix A whose rows are the points in $S^n \cap V$. This has rank at most d, so we can find d columns of A which span all other columns: each row of A can be uniquely determined by the d entries corresponding to those columns. Since $|S| = m$, there are at most m^d distinct choices for those entries, hence at most m^d distinct rows. ∎

Proof of Proposition 3.6. Let M be randomly chosen from $\mathcal{M}_n(k)$, and let E_i be the event that the ith row of M is contained in the span of the previous $i-1$ rows (E_1 is just the probability that the first row is zero). If M is singular, then at least one of $E_1, \ldots, E_n$ must occur, so the probability that M is singular is at most $\sum_{i=1}^n \Pr(E_i)$. By Lemma 3.7 there are at most $(2k+1)^{i-1}$ possible choices for row i that are spanned by the preceding rows, so $\Pr(E_i) \leq (2k+1)^{i-1-n}$. The result then follows by bounding $\sum_{i=1}^n \Pr(E_i)$ with the geometric series $\sum_{j=1}^\infty (2k+1)^{-j} = 1/2k$. ∎

Remark. Proposition 3.6 is used with minor alterations in [66], where it was beneficial for the application to have explicit constants that do not grow with n. Unfortunately, the decay rate of $O(1/k)$ is not quite strong enough to yield an "almost all" result for the rational eigenvalue problem (we need a bound of $o(1/k)$ to get any qualitative analogue of Theorem 3.1). In order to achieve stronger decay with respect to k, we are willing to sacrifice some uniformity — indeed, our method yields an implicit constant that grows very rapidly with n (like a factorial).

We record without proof a well-known bound (see for instance [59, p. 56]) for the number-of-divisors function $\tau(n)$ that will be useful here and in later sections. Although more explicit bounds are available, the one below is quite sufficient for our purposes.

Proposition 3.8. *For any real $\delta > 0$ and nonzero integer n, $\tau(n) \ll_\delta |n|^\delta$.*

We first prove Lemma 3.2 in the easiest case $n = 2$, but in a slightly more general form that will be useful for induction to larger n. We wish to show that the probability that a random 2×2 matrix is singular is equally small ($\ll_\varepsilon k^{-2+\varepsilon}$, albeit with a different implied constant) even if we choose the entries from different arithmetic progressions (each having the same cardinality as $\mathcal{I}_k$).

Lemma 3.9. *Fix positive real numbers α, B, and ε. Let k be a positive integer, and let $L_1(x)$, $L_2(x)$, $L_3(x)$, and $L_4(x)$ be nonconstant linear polynomials whose coefficients are integers at most Bk^α in absolute value. Then the number of solutions to the equation*

$$L_1(a)L_2(b) = L_3(c)L_4(d) \tag{3.5}$$

with all of a, b, c, and d in $\mathcal{I}_k$ is $\ll_{\alpha,B,\varepsilon} k^{2+\varepsilon}$.

Proof. First we consider the solutions for which both sides of equation (3.5) equal 0. In this case, at least two of the linear factors $L_1(a)$, $L_2(b)$, $L_3(c)$, and $L_4(d)$ equal 0. If, for example, $L_1(a) = 0$ and $L_3(c) = 0$ (the other cases are exactly the same), this completely determines the values of a and c; since there are $2k+1$ choices for each of b and d, the total number of solutions for which both sides of equation (3.5) equal 0 is $\ll k^2$.

Otherwise, fix any values for c and d for which the right-hand side of equation (3.5) is nonzero, a total of at most $(2k+1)^2 \ll k^2$ choices. Then the right-hand side is some nonzero integer that is at most $(k \cdot Bk^\alpha + Bk^\alpha)^2 \le 4B^2k^{2+2\alpha}$ in absolute value, and $L_1(a)$ must be a divisor of that integer.

By Proposition 3.8 with $\delta = \varepsilon/(2+2\alpha)$, the right-hand side of equation (3.5) has only $\ll_{\alpha,\varepsilon} (4B^2k^{2+2\alpha})^{\varepsilon/(2+2\alpha)} \ll_{\alpha,B,\varepsilon} k^\varepsilon$ divisors which may serve as candidates for $L_1(a)$; each of these completely determines a value for a (which might not even be an integer). Then the possible values for $L_2(b)$ and hence b are determined as well. We conclude that there are a total of $\ll_{\alpha,B,\varepsilon} k^{2+\varepsilon}$ solutions to equation (3.5) as claimed. ∎

Remark. It is not important that the L_i be *linear*: the above proof works essentially without change for any four nonconstant polynomials of bounded degree. We will not need such a generalization, however, as the determinant of a matrix depends only linearly on each matrix element.

Remark. For our application, both α and B will depend only on the single parameter n, so the dependence $\ll_{\alpha,B,\varepsilon}$ may be rephrased as $\ll_{n,\varepsilon}$, absorbing the explicit references to α and B.

3.3 Eigenvalues and determinants

We previously noted that if $M \in \mathcal{M}_n(k)$ has a rational eigenvalue λ then λ must be an integer. Furthermore, λ could be as high as nk (if we take M to be the $n \times n$ matrix with all entries equal to k), or as low as $-nk$ (if we take M to be the negation of the previous example). In fact it is not possible for λ to be any larger (even when it is complex), and an analogous bound holds for the continuous case as shown by the lemma below.

Lemma 3.10. *Any eigenvalue of a matrix in $\mathcal{M}_n(k)$ is bounded in absolute value by nk. Any eigenvalue of a matrix in $\mathcal{M}_n([-1,1])$ is bounded in absolute value by n.*

Proof. We invoke Gershgorin's "circle theorem" [23], a standard result in spectral theory: let $M = (m_{ij})$ be an $n \times n$ matrix, and let $D(z, r)$ denote the disk of radius r around the complex number z. Then Gershgorin's theorem says that all of the eigenvalues of M must lie in the union of the disks

$$D\Bigg(m_{11}, \sum_{\substack{1\le j\le n\\ j\ne 1}} |m_{1j}|\Bigg),\ D\Bigg(m_{22}, \sum_{\substack{1\le j\le n\\ j\ne 2}} |m_{2j}|\Bigg),\ \ldots,\ D\Bigg(m_{nn}, \sum_{\substack{1\le j\le n\\ j\ne n}} |m_{nj}|\Bigg).$$

In particular, if all entries of M are bounded in absolute value by B, then all eigenvalues are bounded in absolute value by nB. ■

Remark. The examples preceding the lemma show that the bound is tight; indeed, one could also prove this using Perron–Frobenius theory for the spectral radius of M, which provides some intuition for the fact that the all-k's matrix is maximal.

The next ingredient is a curious determinantal identity, which was classically known but presently seems to have fallen out of common knowledge.

Before we state this identity, it will be very helpful to define some preliminary notation. For the remainder of this section, we will adhere to the convention that capital letters denote matrices, boldface lowercase letters denote column vectors, and regular lowercase letters denote scalars.

Let I_n denote the $n \times n$ identity matrix, whose jth column we denote by $\mathbf{e}_j$ (the jth standard basis vector). Let M be an $n \times n$ matrix, and let $\mathbf{m}_j$ denote its jth column and m_{ij} its ijth entry. Note that $M\mathbf{e}_j = \mathbf{m}_j$ by the definition of matrix multiplication. We caution the reader that, in the definitions below, a_{ij} does *not* denote the entries of A, nor those of $\mathbf{a}_j$.

Let a_{ij} denote the ijth cofactor of M, that is, the determinant of the $(n-1) \times (n-1)$ matrix obtained from M by deleting its ith row and jth column. Let $A = \operatorname{Adj}(M)$ denote the *adjugate* matrix of M, that is, the matrix whose ijth entry is $(-1)^{i+j}a_{ji}$. It is a standard consequence of Laplace's determinant expansion [36, §4.III][19] that $MA = (\det M)I_n$. Finally, let $\mathbf{a}_j$ denote the jth column of A. Note that $M\mathbf{a}_j = (\det M)\mathbf{e}_j$, since both sides are the jth column of $(\det M)I_n$.

Lemma 3.11. *Fix an integer $n \geq 3$. Let M be an $n \times n$ matrix with cofactors a_{ij}. Also let Z denote the $(n-2) \times (n-2)$ matrix obtained from M by deleting the first two rows and first two columns, so that*

$$M = \left(\begin{array}{cc|c} m_{11} & m_{12} & * \\ m_{21} & m_{22} & * \\ \hline * & * & Z \end{array}\right), \tag{3.6}$$

where $*$ represents irrelevant entries. Then $a_{11}a_{22} - a_{12}a_{21} = (\det M)(\det Z)$.

It is important to note that when $\det Z \neq 0$, the cofactor a_{11} is a linear polynomial of m_{22} whose leading coefficient is $\det Z$, and whose constant term depends only on Z and the starred entries; conversely, the cofactor a_{22} is a linear polynomial of m_{11} with leading coefficient $\det Z$ (similar correspondences hold for the pair a_{12} and a_{21}). For example, when $n = 3$ the determinant of the 1×1 matrix Z is simply the lower-right entry m_{33} of M;

[19] Our definition of cofactor differs slightly from that in [36], which includes the $(-1)^{i+j}$ term. We prefer this for simplicity as we work mostly with a_{ij} and not the adjugate A.

the identity in question thus translates to

$$\begin{aligned}(m_{11}m_{33} - m_{13}m_{31})&(m_{22}m_{33} - m_{23}m_{32}) \\ &- (m_{12}m_{33} - m_{13}m_{32})(m_{21}m_{33} - m_{23}m_{31}) = m_{33} \det M. \quad (3.7)\end{aligned}$$

For any fixed dimension n, the assertion of Lemma 3.11 is just a polynomial identity like the above, which is verifiable by direct computation; however, a proof that works for general n requires a bit of cunning.

Proof. Define an $n \times n$ matrix

$$B = \begin{pmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{e}_3 & \cdots & \mathbf{e}_n \end{pmatrix} = \left(\begin{array}{cc|c} a_{11} & -a_{21} & 0 \\ -a_{12} & a_{22} & 0 \\ \hline * & * & I_{n-2} \end{array}\right).$$

Since B is in lower-triangular block form, its determinant is easy to evaluate:

$$\det B = \det \begin{pmatrix} a_{11} & -a_{21} \\ -a_{12} & a_{22} \end{pmatrix} \cdot \det I_{n-2} = a_{11}a_{22} - a_{12}a_{21}.$$

Moreover,

$$\begin{aligned} MB &= \begin{pmatrix} M\mathbf{a}_1 & M\mathbf{a}_2 & M\mathbf{e}_3 & \cdots & M\mathbf{e}_n \end{pmatrix} \\ &= \begin{pmatrix} (\det M)\mathbf{e}_1 & (\det M)\mathbf{e}_2 & \mathbf{m}_3 & \cdots & \mathbf{m}_n \end{pmatrix} = \left(\begin{array}{cc|c} \det M & 0 & * \\ 0 & \det M & * \\ \hline 0 & 0 & \mathrm{Z} \end{array}\right). \end{aligned}$$

Since MB is in upper-triangular block form, its determinant $\det(MB) = (\det M)^2(\det Z)$ is also easy to evaluate. Using the identity $\det M \cdot \det B = \det(MB)$, we conclude that

$$(\det M)(a_{11}a_{22} - a_{12}a_{21}) = (\det M)^2(\det Z).$$

Both sides of this last identity are polynomial functions of the n^2 variables m_{ij} representing the entries of M. The factor $\det M$ on both sides is a nonzero polynomial, and hence it can be canceled to obtain $(\det M)(\det Z) = a_{11}a_{22} - a_{12}a_{21}$ as desired. ■

Remark. This proof generalizes readily to an analogous identity for larger minors of the adjugate matrix A. Indeed, Muir's classic treatise on determinants [60, Ch. VI, §175] includes this general form of Lemma 3.11 in a chapter wholly devoted to *compound* determinants (that is, determinants of matrices whose elements are themselves determinants). The same result can also be found in Scott's reference of equally old vintage [75, p. 62], which was made freely available online by the Cornell University Library Historical Math collection.[20]

3.4 Proof of main results for $n \geq 2$

We are now ready to prove the singularity bound of Lemma 3.2. We proceed by an induction on n, which requires $n = 2$ and $n = 3$ as base cases. In the arguments that follow, let M denote a matrix in $\mathcal{M}_n(k)$ with entries m_{ij}.

Base case $n = 2$: The determinant of $\left(\begin{smallmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{smallmatrix}\right)$ is 0 precisely when $m_{11}m_{22} = m_{12}m_{21}$. By Lemma 3.9, there are $\ll_\varepsilon k^{2+\varepsilon}$ solutions to this equation with the variables m_{11}, m_{12}, m_{21}, and m_{22} all in $\mathcal{I}_k$. This immediately shows that $|\mathcal{M}_2^0(k)| \ll_\varepsilon k^{2+\varepsilon}$ as claimed. Since $1/|\mathcal{M}_2(k)| \ll k^{-4}$, we see that the probability of a randomly chosen matrix from $\mathcal{M}_2(k)$ being singular is $|\mathcal{M}_2^0(k)|/|\mathcal{M}_2(k)| \ll_\varepsilon k^{-2+\varepsilon}$.

Base case $n = 3$: We first estimate the number of matrices in $\mathcal{M}_3^0(k)$ whose lower right-hand entry m_{33} is nonzero. Fix the five entries in the last row and last column of M, with $m_{33} \neq 0$; there are a total of $2k(2k+1)^4 \ll k^5$ possibilities. Using the earlier identity (3.7), we see that if $\det M = 0$ then

[20] `http://ebooks.library.cornell.edu/cgi/t/text/text-idx?c=math;idno=01670002`

we must have

$$\begin{aligned}(m_{11}m_{33} - m_{13}m_{31})(m_{22}m_{33} - m_{23}m_{32})\\ = (m_{12}m_{33} - m_{13}m_{32})(m_{21}m_{33} - m_{23}m_{31}).\end{aligned}$$

This equation is of the form $L_1(m_{11})L_2(m_{22}) = L_3(m_{12})L_4(m_{21})$, where the L_i are nonconstant linear polynomials whose coefficients are at most k^2 in absolute value. (Note that we have used the fact that $m_{33} \neq 0$ in asserting that the L_i are nonconstant.) Applying Lemma 3.9 with $\alpha = 2$, we see that there are $\ll_\varepsilon k^{2+\varepsilon}$ solutions to this equation with m_{11}, m_{12}, m_{21}, and m_{22} all in $\mathcal{I}_k$. This shows that there are $\ll_\varepsilon k^{7+\varepsilon}$ matrices in $\mathcal{M}_3^0(k)$ whose lower right-hand entry m_{33} is nonzero.

If any of the entries in the last row of M is nonzero, then we can swap two columns of M, if necessary, to bring that entry into the lower right-hand position; any matrix so formed corresponds to at most three distinct matrices in $\mathcal{M}_3^0(k)$, and so there are still $\ll_\varepsilon k^{7+\varepsilon}$ matrices in $\mathcal{M}_3^0(k)$ that have any nonzero entry in the last row. Finally, any matrix whose last row consists of all zeros is certainly in $\mathcal{M}_3^0(k)$, but there are only $(2k+1)^6 \ll k^6$ such matrices. We conclude that the total number of matrices in $\mathcal{M}_3^0(k)$ is $\ll_\varepsilon k^{7+\varepsilon}$, so that the probability of a randomly chosen matrix from $\mathcal{M}_3(k)$ being singular is $|\mathcal{M}_3^0(k)|/|\mathcal{M}_3(k)| \ll_\varepsilon k^{-2+\varepsilon}$ as claimed.

Inductive step for $n \geq 4$: Write a matrix $M \in \mathcal{M}_n(k)$ in the form (3.6). Some such matrices will have $\det Z = 0$; however, by the induction hypothesis for $n-2$, the probability that this occurs is $\ll_{n,\varepsilon} k^{-2+\varepsilon}$ (independent of the entries outside Z), which is an allowably small probability.

Otherwise, fix values in $\mathcal{I}_k$ for the $n^2 - 4$ entries other than m_{11}, m_{12}, m_{21}, and m_{22} such that $\det Z \neq 0$. It suffices to show that, conditioned on any such fixed values, the probability that M is singular is $\ll_{n,\varepsilon} k^{-2+\varepsilon}$, as we let the remaining variables m_{11}, m_{12}, m_{21}, and m_{22} range over $\mathcal{I}_k$, .

By Lemma 3.11, we see that $\det M = 0$ is equivalent to $a_{11}a_{22} = a_{12}a_{21}$. Recall that the cofactor a_{11} is a linear polynomial of m_{22} with leading coefficient $\det Z$, while a_{22} is a linear polynomial of m_{11} with leading coefficient $\det Z$ (and similarly for the pair a_{12} and a_{21}). Moreover, the coefficients of these polynomials are at most $(n-1)!\,k^{n-1}$ in magnitude: by Laplace expansion, each coefficient is the sum of at most $(n-1)!$ products, each composed of at most $n-1$ entries of M. We may thus apply Lemma 3.9 with $\alpha = n-1$ and $B = (n-1)!$ to see that the (conditional) probability of $a_{11}a_{22} = a_{12}a_{21}$ is $\ll_{n,\varepsilon} k^{-2+\varepsilon}$, as desired. ∎

We now have everything we need to prove Theorem 3.1: *given any integer $n \ge 2$ and any real number $\varepsilon > 0$, the probability that a randomly chosen matrix in $\mathcal{M}_n(k)$ has an integer eigenvalue is $\ll_{n,\varepsilon} 1/k^{1-\varepsilon}$.* By Lemma 3.10, any such eigenvalue λ is at most nk in absolute value. For each individual λ, we observe that if $M \in \mathcal{M}_n(k)$ has eigenvalue λ, then $M - \lambda I_n$ is a singular matrix with integer entries bounded in size by $k + |\lambda| \le (n+1)k$. Thus, every matrix in $\mathcal{M}_n^{\lambda}(k)$ is also contained in the set

$$\bigl\{M + \lambda I_n \colon M \in \mathcal{M}_n^0\bigl((n+1)k\bigr)\bigr\}.$$

By Lemma 3.2, the cardinality of this set is $\ll_{n,\varepsilon} ((n+1)k)^{n^2-2+\varepsilon} \ll_{n,\varepsilon} k^{n^2-2+\varepsilon}$ for any fixed λ. Summing over all integer values of λ from $-nk$ to nk (admittedly, some matrices are counted multiple times, but the upper bound thus obtained is still valid), we conclude that the total number of matrices in $\mathcal{M}_n^{\mathbb{Z}}(k)$ is $\ll_{n,\varepsilon} k^{n^2-1+\varepsilon}$. In other words, the probability that a matrix in $\mathcal{M}_n(k)$ has an integer eigenvalue is $|\mathcal{M}_n^{\mathbb{Z}}(k)|/|\mathcal{M}_n(k)| \ll_{n,\varepsilon} 1/k^{1-\varepsilon}$, as desired. ∎

3.5 Preliminaries for 2×2 matrices

Having established Theorem 3.1 and Lemma 3.2 for general n, we can give considerably sharper results in the simplest case $n = 2$. The remainder of this chapter is devoted to the proofs of Theorem 3.3 and Corollary 3.4 for integer eigenvalues, and also Theorem 3.5 for real eigenvalues.

The 2×2 case is particularly nice: since the trace of an integer matrix is itself an integer, it follows that if one eigenvalue is an integer then both are. Consequently, there is no significant distinction between belonging to

$\mathcal{M}_2^{\mathbb{Z}}(k)$ and being diagonalizable over $\mathbb{Q}$ (if M has rational eigenvalues but is not rationally diagonalizable, then its eigenvalues must be repeated, and we confirm in Lemma 3.14 that this occurs very rarely).

We begin with some elementary observations for 2×2 matrices (both integer-valued and real-valued) that will simplify our computations of the functions $V(\delta)$ and $W(\delta)$. The key to the precise enumeration of $\mathcal{M}_2^{\lambda}(k)$ is the simple structure of singular integer matrices:

Lemma 3.12. *For any singular matrix $M \in \mathcal{M}_2^0(k)$, either at least two entries of M equal zero, or else there exist nonzero integers a, b, c, d with $(a, b) = 1$ such that*

$$M = \begin{pmatrix} ac & bc \\ ad & bd \end{pmatrix}. \tag{3.8}$$

Moreover, this representation of M is unique up to replacing each of a, b, c, d by its negative.

Proof. If one of the entries of M equals zero, then a second one must equal zero as well for the determinant to vanish. Otherwise, given

$$M = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix}$$

with none of the m_{ij} equal to zero, define $c = (m_{11}, m_{12})$, and set $a = m_{11}/c$ and $b = m_{12}/c$, so that $(a, b) = 1$. Since M is singular, the second row of m must be a multiple of the first row—that is, there exists a real number d such that $m_{21} = ad$ and $m_{22} = bd$. Since a and b are relatively prime, moreover, d must in fact be an integer.

This shows that every such matrix admits a representation of type (3.8). If there is another representation

$$M = \begin{pmatrix} a'c' & b'c' \\ a'd' & b'd' \end{pmatrix},$$

then $(a', b') = 1$ implies $(a'c', b'c') = |c'|$, which shows that $|c'| = c$; the equalities $|a'| = |a|$, $|b'| = |b|$, and $|d'| = |d|$ follow quickly. ∎

For a 2×2 matrix $M = \left(\begin{smallmatrix} a & c \\ b & d \end{smallmatrix}\right)$, we define $\operatorname{disc} M = (\operatorname{tr} M)^2 - 4 \det M = (a-d)^2 + 4bc$, where $\operatorname{tr} M$ is the trace $a + d$ and $\det M$ is the determinant $ad - bc$. It is easily seen that $\operatorname{disc} M$ is the discriminant of the characteristic polynomial of M. We record the following elementary facts, which will be useful in the proofs of Lemma 3.14 and Proposition 3.20.

Lemma 3.13. *Let M be a 2×2 matrix with real entries.*

(a) M has repeated eigenvalues if and only if $\operatorname{disc} M = 0$*.*

(b) M has real eigenvalues if and only if $\operatorname{disc} M \geq 0$*.*

(c) $\det M < 0$ *if and only if M has two real eigenvalues of opposite sign.*

(d) If $\det M > 0$ *and* $\operatorname{disc} M \geq 0$*, then the eigenvalues of M have the same sign as* $\operatorname{tr} M$*.*

Proof. Let λ_1, λ_2 denote the eigenvalues of M, so that $\operatorname{tr} M = \lambda_1 + \lambda_2$, $\det M = \lambda_1 \lambda_2$, and $\operatorname{disc} M = (\lambda_1 - \lambda_2)^2$, each of which is real. Parts (a), (b) and (d) follow immediately from these observations, and part (c) from the fact that $\lambda_2 = \overline{\lambda_1}$ if λ_1, λ_2 are complex. ■

The next lemma gives a bound for the probability of a matrix having repeated eigenvalues. It is natural to expect this probability to approach 0 as k increases, and indeed such a result was obtained in [37] for matrices of arbitrary size, with an upper bound of $O_n(1/k)$. We give a simple proof of a stronger bound for the 2×2 case, as well as an analogous qualitative statement for real matrices which will be helpful in the proof of Theorem 3.5.

Lemma 3.14. *The number of matrices in $\mathcal{M}_2(k)$ with a repeated eigenvalue is $\ll_\varepsilon k^{2+\varepsilon}$ for every $\varepsilon > 0$. The probability that a random matrix in $\mathcal{M}_2([-1,1])$ has a repeated eigenvalue or is singular is* 0*.*

Proof. By Lemma 3.13(a), the 2×2 matrix $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ has a double eigenvalue if and only if $4bc = -(a-d)^2$. For matrices in $\mathcal{M}_2([-1,1])$ this is easily seen to be a zero-probability event, as is the event that $\det M = ad - bc = 0$.

For matrices in $\mathcal{M}_2(k)$, we enumerate how many can satisfy $4bc = -(a-d)^2$. This follows the same argument as Lemma 3.9: if $a = d$ then one of b or c must be 0, yielding at most $\ll k^2$ solutions. Otherwise, this there are $O(k^2)$ choices for a and d that make $(a-d)$ nonzero, each yielding $\ll_\varepsilon k^\varepsilon$ choices for b, c. Combining the two yields $\ll_\varepsilon k^{2+\varepsilon}$ solutions in total. ∎

3.6 Enumeration theorems for integer eigenvalues

Let $\mu(n)$ be the Möbius function, characterized by the identity

$$\sum_{d|n} \mu(d) = \begin{cases} 1, & \text{if } n = 1, \\ 0, & \text{if } n > 1. \end{cases} \tag{3.9}$$

The well-known Dirichlet series identity $1/\zeta(s) = \sum_{n=1}^{\infty} \mu(n) n^{-s}$ is valid for $\Re s > 1$ (see [59, Corollary 1.10], for example). In particular,

$$\sum \frac{\mu(n)}{n^2} = \frac{6}{\pi^2},$$

and we can estimate the tail of this series (using $|\mu(n)| \le 1$) to obtain the quantitative estimate

$$\sum_{d \le k} \frac{\mu(d)}{d^2} = \frac{6}{\pi^2} + O\left(\frac{1}{k}\right). \tag{3.10}$$

Lemma 3.15. *For nonzero integers a, b and parameters k, λ, define the function*

$$N_{k,\lambda}(a,b) = \#\{(c,d) \in \mathbb{Z}^2,\ c \ne 0,\ d \ne 0 : |ac+\lambda|, |bc|, |ad|, |bd+\lambda| \le k\}. \tag{3.11}$$

Then

$$\left|\mathcal{M}_2^\lambda(k)\right| = 4 \sum_{d \le k} \mu(d) \sum_{1 \le \alpha < \beta \le k/d} N_{k,\lambda}(d\alpha, d\beta) + O(k^2),$$

where the implied constant is independent of λ and k.

Proof. Fix an integer $0 \le \lambda \le 2k$, and let $M \in \mathcal{M}_2^\lambda(k)$, so that $M - \lambda I$ is singular. By Lemma 3.12, either at least two entries of $M - \lambda I$ equal zero, or else $M - \lambda I$ has exactly two representations of the form (3.8). In the former case, there are $2k+1$ choices for each of the two potentially nonzero entries, hence $O(k^2)$ such matrices in total (even taking into account the several different choices of which two entries are nonzero). In the latter case, there are exactly two corresponding quadruples a, b, c, d of integers as in Lemma 3.12. Taking into account that each entry of M must be at most k in absolute value, we deduce that

$$\begin{aligned}\left|\mathcal{M}_2^\lambda(k)\right| &= \tfrac{1}{2}\,\#\big\{(a,b,c,d) \in \mathbb{Z}^4 : a,b,c,d \neq 0, \\ &\qquad (a,b)=1,\ |ac+\lambda|, |bc|, |ad|, |bd+\lambda| \le k\big\} + O(k^2) \\ &= \tfrac{1}{2} \sum_{\substack{1 \le |a|,|b| \le k \\ (a,b)=1}} N_{k,\lambda}(a,b) + O(k^2).\end{aligned}$$

Because of the symmetries $N_{k,\lambda}(a,b) = N_{k,\lambda}(-a,b) = N_{k,\lambda}(a,-b) = N_{k,\lambda}(-a,-b)$, we have

$$\left|\mathcal{M}_2^\lambda(k)\right| = 2 \sum_{\substack{1 \le a,b \le k \\ (a,b)=1}} N_{k,\lambda}(a,b) + O(k^2).$$

The only term in the sum where $a = b$ is the term $a = b = 1$, and for all other terms we can invoke the additional symmetry $N_{k,\lambda}(a,b) = N_{k,\lambda}(b,a)$, seen to be valid by switching the roles of c and d in the definition (3.11) of $N_{k,\lambda}(a,b)$. We obtain

$$\begin{aligned}\left|\mathcal{M}_2^\lambda(k)\right| &= 4 \sum_{\substack{1 \le a<b \le k \\ (a,b)=1}} N_{k,\lambda}(a,b) + 2N_{k,\lambda}(1,1) + O(k^2) \\ &= 4 \sum_{\substack{1 \le a<b \le k \\ (a,b)=1}} N_{k,\lambda}(a,b) + O(k^2),\end{aligned}$$

where the last step used the fact that $N_{k,\lambda}(1,1) \le \#\{(c,d) \in \mathbb{Z}^2 : |c|,|d| \le k\} \ll k^2$.

Using the characteristic property of the Möbius function (3.9), we can write the last expression as

$$\begin{aligned}\left|\mathcal{M}_2^\lambda(k)\right| &= 4\sum_{1\le a<b\le k} N_{k,\lambda}(a,b)\sum_{d|(a,b)}\mu(d)+O(k^2)\\ &= 4\sum_{d\le k}\mu(d)\sum_{\substack{1\le a<b\le k\\ d|a,\, d|b}} N_{k,\lambda}(a,b)+O(k^2)\\ &= 4\sum_{d\le k}\mu(d)\sum_{1\le\alpha<\beta\le k/d} N_{k,\lambda}(d\alpha,d\beta)+O(k^2).\end{aligned}$$

■

Lemma 3.16. *Let k and λ be integers with $0\le\lambda\le 2k$, and let x and y be integers with $1\le x\le y\le k$. Then*

$$N_{k,\lambda}(x,y)=k^2C\big(\tfrac{\lambda}{k};x,y\big)D\big(\tfrac{\lambda}{k};x,y\big)+O\big(\tfrac{k}{y}\big),$$

where

$$\begin{aligned}C(\delta;x,y)&=\max\left\{0,\min\left\{\tfrac{1-\delta}{x}+\tfrac{1}{y},\tfrac{2}{y}\right\}\right\},\\ D(\delta;x,y)&=\min\left\{\tfrac{1-\delta}{y}+\tfrac{1}{x},\tfrac{2}{y}\right\}.\end{aligned}\tag{3.12}$$

Proof. We have

$$\begin{aligned}N_{k,\lambda}(x,y)&=\#\{(c,d)\in\mathbb{Z}^2,\ c\neq 0,\ d\neq 0: |xc+\lambda|,|yc|,|xd|,|yd+\lambda|\le k\}\\ &=\#\{c\in\mathbb{Z},\ c\neq 0: -k\le xc+\lambda\le k,\ -k\le yc\le k\}\\ &\qquad\times\#\{d\in\mathbb{Z},\ d\neq 0: -k\le xd\le k,\ -k\le yd+\lambda\le k\}.\end{aligned}$$

Since x and y are positive, we can rewrite this product as

$$\begin{aligned} N_{k,\lambda}(x,y) &= \#\{c \in \mathbb{Z},\, c \neq 0 : (-k-\lambda)/x \le c \le (k-\lambda)/x,\ -k/y \le c \le k/y\} \\ &\qquad \times \#\{d \in \mathbb{Z},\, d \neq 0 : -k/x \le d \le k/x,\ (-k-\lambda)/y \le d \le (k-\lambda)/y\} \\ &= \#\big\{c \in \mathbb{Z},\, c \neq 0 : -k/y \le c \le \min\{(k-\lambda)/x, k/y\}\big\} \qquad (3.13) \\ &\qquad \times \#\big\{d \in \mathbb{Z},\, d \neq 0 : \max\{-k/x, (-k-\lambda)/y\} \le d \le (k-\lambda)/y\big\}, \end{aligned}$$

where we have used $\lambda \ge 0$ and $x \le y$ to simplify the inequalities slightly. The first factor on the right-hand side of equation (3.13) is

$$\min\{(k-\lambda)/x, k/y\} - (-k/y) + O(1) = k \min\big\{(1-\tfrac{\lambda}{k})/x + 1/y, 2/y\big\} + O(1)$$

if this expression is positive, and 0 otherwise; it is thus precisely $kC(\frac{\lambda}{k}; x, y) + O(1)$. Similarly, the second factor on the right-hand side of (3.13) is

$$(k-\lambda)/y - \max\{-k/x, (-k-\lambda)/y\} + O(1) = k \min\{(1-\tfrac{\lambda}{k})/y + 1/x, 2/y\} + O(1)$$

(note that this expression is always positive under the hypotheses of the lemma), which is simply $kD(\frac{\lambda}{k}; x, y) + O(1)$. Multiplying these two factors yields

$$N_{k,\lambda}(x,y) = k^2 C\big(\tfrac{\lambda}{k}; x, y\big) D\big(\tfrac{\lambda}{k}; x, y\big) + k \cdot O(C\big(\tfrac{\lambda}{k}; x, y\big) + D\big(\tfrac{\lambda}{k}; x, y\big)) + O(1).$$

The lemma follows upon noting that both $C(\frac{\lambda}{k}; x, y)$ and $D(\frac{\lambda}{k}; x, y)$ are $\ll 1/y$ by definition, so the second summand becomes simply $O(\frac{k}{y})$, and the $O(1)$ term may be subsumed into $O(\frac{k}{y})$ since $y \le k$. ■

We have already used the trivial estimate

$$\sum_{L \le \alpha < U} 1 = (U - L) + O(1),$$

provided $0 < L < U$. We will also use, without further comment, the estimates

$$\sum_{L\le\alpha<U} \frac{1}{\alpha} = \log\frac{U}{L} + O\Big(\frac{1}{L}\Big),$$

with its particular case

$$\sum_{1\le\alpha<U} \frac{1}{\alpha} = \log U + O(1),$$

and

$$\sum_{L\le\alpha<U} \frac{1}{\alpha^2} = \frac{1}{L} - \frac{1}{U} + O\Big(\frac{1}{L^2}\Big).$$

These estimates (also valid for $0 < L < U$) follow readily from comparison to the integrals $\int_L^U 1/x\,dx$ and $\int_L^U 1/x^2\,dx$.

Most of the technical work in proving Theorem 3.3 lies in establishing an estimate for a sum of the form $\sum_{1\le\alpha<\beta} C(\delta;\alpha,\beta)D(\delta;\alpha,\beta)$ for a fixed β. The following proposition provides an asymptotic formula for this sum; we defer the proof until the next section. Assuming this proposition, though, we can complete the proof of Theorem 3.3, as well as Corollary 3.4.

Proposition 3.17. *Let $\beta \ge 1$ and $0 \le \delta \le 2$ be real numbers, and let C and D be the functions defined in equation* (3.12)*. Then*

$$\sum_{1\le\alpha<\beta} C(\delta;\alpha,\beta)D(\delta;\alpha,\beta) = \frac{V(\delta)}{\beta} + O\Big(\frac{1+\log\beta}{\beta^2}\Big),$$

where $V(\delta)$ was defined in equation (3.1)*.*

Proof of Theorem 3.3 assuming Proposition 3.17. The functions C and D defined in equation (3.12) are homogeneous of degree -1 in the variables x and y, so that Lemma 3.16 implies

$$N_{k,\lambda}(d\alpha, d\beta) = \frac{k^2}{d^2} C\big(\tfrac{\lambda}{k};\alpha,\beta\big) D\big(\tfrac{\lambda}{k};\alpha,\beta\big) + O\big(\tfrac{k}{d\beta}\big).$$

Inserting this formula into the conclusion of Lemma 3.15 yields

$$\begin{aligned}\bigl|\mathcal{M}_2^\lambda(k)\bigr| = 4k^2 \sum_{d\le k} \frac{\mu(d)}{d^2} \sum_{1\le\alpha<\beta\le k/d} C\bigl(\tfrac{\lambda}{k};\alpha,\beta\bigr) D\bigl(\tfrac{\lambda}{k};\alpha,\beta\bigr)& \\ + O\biggl(\sum_{d\le k}\sum_{1\le\alpha<\beta\le k/d} \frac{k}{d\beta}\biggr) + O(k^2).&\end{aligned}$$

We bound the first error term by summing over $1 \le \alpha < \beta$ to obtain

$$\sum_{d\le k}\sum_{1\le\alpha<\beta\le k/d} \frac{k}{d\beta} \le \sum_{d\le k}\sum_{1<\beta\le k/d} \frac{k}{d} < \sum_{d\le k} \frac{k^2}{d^2} \ll k^2,$$

so that we have the estimate

$$\bigl|\mathcal{M}_2^\lambda(k)\bigr| = 4k^2 \sum_{d\le k} \frac{\mu(d)}{d^2} \sum_{1\le\alpha<\beta\le k/d} C\bigl(\tfrac{\lambda}{k};\alpha,\beta\bigr) D\bigl(\tfrac{\lambda}{k};\alpha,\beta\bigr) + O(k^2). \tag{3.14}$$

We now apply Proposition 3.17 to obtain

$$\begin{aligned}\bigl|\mathcal{M}_2^\lambda(k)\bigr| &= 4k^2 \sum_{d\le k} \frac{\mu(d)}{d^2} \sum_{1\le\beta\le k/d} \biggl(\frac{V(\lambda/k)}{\beta} + O\biggl(\frac{1+\log\beta}{\beta^2}\biggr)\biggr) + O(k^2) \\ &= 4k^2 \sum_{d\le k} \frac{\mu(d)}{d^2} \biggl(V(\lambda/k)\biggl(\log\frac{k}{d} + O(1)\biggr) + O(1)\biggr) + O(k^2) \\ &= 4k^2 \biggl(V(\lambda/k)\log k \sum_{d\le k} \frac{\mu(d)}{d^2} + O\biggl(\sum_{d\le k} \frac{\log d}{d^2}\biggr)\biggr) + O(k^2) \\ &= 4k^2 \biggl(V(\lambda/k)\log k\biggl(\frac{6}{\pi^2} + O\biggl(\frac{1}{k}\biggr)\biggr) + O(1)\biggr) + O(k^2) \\ &= \frac{24V(\lambda/k)}{\pi^2} k^2 \log k + O(k^2),\end{aligned}$$

where we have used equation (3.10) and the convergence of $\sum 1/n^2$ and $\sum(\log n)/n^2$ (so the partial sums are uniformly bounded). ∎

Proof of Corollary 3.4 from Theorem 3.3. For any $M \in \mathcal{M}_2(k)$, if one eigenvalue is an integer then they both are (since the trace of M is an integer).

Thus if we add up the cardinalities of all of the $\mathcal{M}_2^\lambda(k)$, we get twice the cardinality of $\mathcal{M}_2^{\mathbb{Z}}(k)$, except that matrices with repeated eigenvalues only get counted once. However, the number of such matrices is $\ll_\varepsilon k^{2+\varepsilon}$ by Lemma 3.14. Therefore

$$\begin{aligned} 2|\mathcal{M}_2^{\mathbb{Z}}(k)| &= \sum_{\lambda\in\mathbb{Z}} |\mathcal{M}_2^\lambda(k)| + O_\varepsilon(k^{2+\varepsilon}) \\ &= \sum_{-2k\le\lambda\le 2k} \left(\frac{24V(\lambda/k)}{\pi^2} k^2 \log k + O(k^2) \right) + O_\varepsilon(k^{2+\varepsilon}) \\ &= \frac{24k^3\log k}{\pi^2} \sum_{-2k\le\lambda\le 2k} \frac{V(\lambda/k)}{k} + O(k^3). \end{aligned}$$

The sum is a Riemann sum of a function of bounded variation, so this becomes

$$2|\mathcal{M}_2^{\mathbb{Z}}(k)| = \frac{24k^3\log k}{\pi^2}\left(\int_{-2}^{2} V(\delta)\, d\delta + O\big(\tfrac{1}{k}\big) \right) + O(k^3).$$

The corollary then follows from the straightforward computation of the integral $\int_{-2}^{2} V(\delta)\, d\delta = \frac{4}{9}(7\sqrt{2}+4+3\log(\sqrt{2}+1))$, noting that $\log(\sqrt{2}-1) = -\log(\sqrt{2}+1)$. ■

3.7 Proof of key proposition for enumeration

It remains to prove Proposition 3.17. Recalling that the functions C and D defined in (3.12) are formed by combinations of minima and maxima, we need to separate our arguments into several cases depending on the range of δ. The following lemma addresses a sum that occurs in two such cases ($0 < \delta < 1$ and $1 < \delta < \sqrt{2}$). Note that the formula for $V(\delta)$ contains terms like $\log(\delta - 1)$, so we need to exercise some caution near $\delta = 1$.

Lemma 3.18. *Let $\beta \ge 1$ and $0 \le \delta \le \sqrt{2}$ be real numbers, with $\delta \ne 1$.*

Then

$$\sum_{\max\{1,|1-\delta|\beta\}\le\alpha<(1+\delta)^{-1}\beta}\left(\frac{1-\delta}{\alpha}+\frac{1}{\beta}\right)\frac{2}{\beta}$$
$$=\frac{2}{\beta}\left(\frac{1}{1+\delta}-|1-\delta|-(1-\delta)\log|1-\delta^2|\right)+O\left(\frac{1+\log\beta}{\beta^2}\right).$$

Proof. Suppose first that $|1-\delta|\beta\ge 1$. Then the sum in question is

$$\frac{2(1-\delta)}{\beta}\sum_{|1-\delta|\beta\le\alpha<(1+\delta)^{-1}\beta}\frac{1}{\alpha}+\frac{2}{\beta^2}\sum_{|1-\delta|\beta\le\alpha<(1+\delta)^{-1}\beta}1$$
$$=\frac{2(1-\delta)}{\beta}\left(\log\frac{\beta/(1+\delta)}{|1-\delta|\beta}+O\left(\frac{1}{|1-\delta|\beta}\right)\right)$$
$$+\frac{2}{\beta^2}\left(\frac{\beta}{1+\delta}-|1-\delta|\beta+O(1)\right)$$
$$=\frac{2}{\beta}\left(\frac{1}{1+\delta}-|1-\delta|-(1-\delta)\log|1-\delta^2|\right)+O\left(\frac{1}{\beta^2}\right),$$

which establishes the lemma in this case. On the other hand, if $|1-\delta|\beta<1$ then the sum in question is

$$\frac{2(1-\delta)}{\beta}\sum_{1\le\alpha<(1+\delta)^{-1}\beta}\frac{1}{\alpha}+\frac{2}{\beta^2}\sum_{1\le\alpha<(1+\delta)^{-1}\beta}1$$
$$=\frac{2(1-\delta)}{\beta}\left(\log\frac{\beta}{1+\delta}+O(1)\right)+\frac{2}{\beta^2}\left(\frac{\beta}{1+\delta}+O(1)\right)$$
$$=\frac{2}{\beta}\left(\frac{1}{1+\delta}-(1-\delta)\log(1+\delta)\right)+O\left(\frac{1}{\beta^2}+\frac{|1-\delta|\log\beta}{\beta}\right).$$

We subtract $2(|1-\delta|+(1-\delta)\log|1-\delta|)/\beta$ from the main term and compensate in the error term to obtain

$$\begin{aligned}\frac{2(1-\delta)}{\beta} & \sum_{1\le\alpha<(1+\delta)^{-1}\beta}\frac{1}{\alpha}+\frac{2}{\beta^2}\sum_{1\le\alpha<(1+\delta)^{-1}\beta}1\\ &=\frac{2}{\beta}\biggl(\frac{1}{1+\delta}-|1-\delta|-(1-\delta)\log|1-\delta^2|\biggr)\\ &\qquad+O\biggl(\frac{1}{\beta^2}+\frac{|1-\delta|\log\beta}{\beta}+\frac{|1-\delta|+|1-\delta|\log|1-\delta|^{-1})}{\beta}\biggr)\\ &=\frac{2}{\beta}\biggl(\frac{1}{1+\delta}-|1-\delta|-(1-\delta)\log|1-\delta^2|\biggr)\\ &\qquad+O\biggl(\frac{1+\log\beta}{\beta^2}+\frac{|1-\delta|\log|1-\delta|^{-1})}{\beta}\biggr),\end{aligned}$$

since we are working with the assumption that $|1-\delta|<1/\beta$. Because the function $t\log t^{-1}$ is increasing on the interval $(0,1/e)$ and bounded on the interval $(0,1]$, we have $|1-\delta|\log|1-\delta|^{-1}<(1/\beta)\log\beta$ if $\beta>e$ and $|1-\delta|\log|1-\delta|^{-1}\ll 1\ll 1/\beta$ if $1\le\beta\le e$. In either case, the last error term can be simplified to $O((1+\log\beta)/\beta^2)$, which establishes the lemma in second case. ∎

Proof of Proposition 3.17. We consider separately the four cases corresponding to the different parts of the definition (3.1) of $V(\delta)$. For brevity we will suppress the dependence on δ, α and β from the notation for C and D.

- *Case 1:* $0\le\delta<1$. In this case we have $0<1-\delta<(1+\delta)^{-1}\le 1$ and

$$C=\begin{cases}\frac{2}{\beta}, & \text{if } \alpha\le(1-\delta)\beta,\\ \frac{1-\delta}{\alpha}+\frac{1}{\beta}, & \text{if } (1-\delta)\beta\le\alpha\end{cases}\quad\text{and}\quad D=\begin{cases}\frac{2}{\beta}, & \text{if } \alpha\le(1+\delta)^{-1}\beta,\\ \frac{1-\delta}{\beta}+\frac{1}{\alpha}, & \text{if } (1+\delta)^{-1}\beta\le\alpha.\end{cases}$$

Therefore

$$\begin{aligned}\sum_{1\le\alpha<\beta}CD=\sum_{1\le\alpha<(1-\delta)\beta}\frac{2}{\beta}\cdot\frac{2}{\beta}&+\sum_{\max\{1,(1-\delta)\beta\}\le\alpha<(1+\delta)^{-1}\beta}\biggl(\frac{1-\delta}{\alpha}+\frac{1}{\beta}\biggr)\frac{2}{\beta}\\ &+\sum_{(1+\delta)^{-1}\beta\le\alpha<\beta}\biggl(\frac{1-\delta}{\alpha}+\frac{1}{\beta}\biggr)\biggl(\frac{1-\delta}{\beta}+\frac{1}{\alpha}\biggr).\end{aligned}\qquad(3.15)$$

(The first sum might be empty, but this does not invalidate the argument that follows.) The first sum is simply

$$\frac{4}{\beta^2}\sum_{1\le\alpha<(1-\delta)\beta} 1 = \frac{4}{\beta^2}\big((1-\delta)\beta + O(1)\big) = \frac{4(1-\delta)}{\beta} + O\bigg(\frac{1}{\beta^2}\bigg).$$

By Lemma 3.18, the second sum is

$$\begin{aligned}\sum_{\max\{1,(1-\delta)\beta\}\le\alpha<(1+\delta)^{-1}\beta} &\bigg(\frac{1-\delta}{\alpha}+\frac{1}{\beta}\bigg)\frac{2}{\beta}\\ &= \frac{2}{\beta}\bigg(\frac{1}{1+\delta}+\delta-1-(1-\delta)\log(1-\delta^2)\bigg)+O\bigg(\frac{1+\log\beta}{\beta^2}\bigg),\end{aligned}$$

while the third sum is

$$\begin{aligned}\sum_{(1+\delta)^{-1}\beta\le\alpha<\beta} &\bigg(\frac{1-\delta}{\alpha}+\frac{1}{\beta}\bigg)\bigg(\frac{1-\delta}{\beta}+\frac{1}{\alpha}\bigg)\\ &= (1-\delta)\sum_{(1+\delta)^{-1}\beta\le\alpha<\beta}\frac{1}{\alpha^2} + \frac{\delta^2-2\delta+2}{\beta}\sum_{(1+\delta)^{-1}\beta\le\alpha<\beta}\frac{1}{\alpha}\\ &\quad+\frac{1-\delta}{\beta^2}\sum_{(1+\delta)^{-1}\beta\le\alpha<\beta}1\\ &= (1-\delta)\bigg(\frac{1+\delta}{\beta}-\frac{1}{\beta}+O\bigg(\frac{1}{\beta^2}\bigg)\bigg)\\ &\quad+\frac{\delta^2-2\delta+2}{\beta}\bigg(\log\frac{\beta}{(1+\delta)^{-1}\beta}+O\bigg(\frac{1}{\beta}\bigg)\bigg)\\ &\quad+\frac{1-\delta}{\beta^2}\bigg(\beta-\frac{\beta}{1+\delta}+O(1)\bigg)\\ &= \frac{1}{\beta}\bigg(2-\delta^2-\frac{2}{1+\delta}+(\delta^2-2\delta+2)\log(1+\delta)\bigg)+O\bigg(\frac{1}{\beta^2}\bigg).\end{aligned} \tag{3.16}$$

This case of the proposition then follows from (3.15) upon noting that

$$4-4\delta+\frac{2}{1+\delta}+2\delta-2-2(1-\delta)\log(1-\delta^2)+2-\delta^2-\frac{2}{1+\delta}+(\delta^2-2\delta+2)\log(1+\delta)$$
$$= 4-2\delta-\delta^2+\delta^2\log(1+\delta)-2(1-\delta)\log(1-\delta).$$

- *Case 2:* $\delta = 1$. In this case we have

$$C = \frac{1}{\beta} \quad \text{and} \quad D = \begin{cases} 2/\beta, & \text{if } \alpha \le \beta/2, \\ 1/\alpha, & \text{if } \beta/2 \le \alpha. \end{cases}$$

Therefore

$$\begin{aligned}\sum_{1\le\alpha<\beta} CD &= \sum_{1\le\alpha<\beta/2} \frac{1}{\beta}\cdot\frac{2}{\beta} + \sum_{\beta/2\le\alpha<\beta} \frac{1}{\beta}\cdot\frac{1}{\alpha} \\ &= \frac{2}{\beta^2}\bigg(\frac{\beta}{2}+O(1)\bigg) + \frac{1}{\beta}\bigg(\log\frac{\beta}{\beta/2}+O\bigg(\frac{1}{\beta}\bigg)\bigg) \\ &= \frac{1+\log 2}{\beta}+O\bigg(\frac{1}{\beta^2}\bigg),\end{aligned}$$

as desired.

- *Case 3:* $1 < \delta \le \sqrt{2}$. In this case we have

$$C = \begin{cases} 0, & \text{if } \alpha \le (\delta-1)\beta, \\ \frac{1-\delta}{\alpha}+\frac{1}{\beta}, & \text{if } (\delta-1)\beta \le \alpha \end{cases} \quad \text{and} \quad D = \begin{cases} \frac{2}{\beta}, & \text{if } \alpha \le (\delta+1)^{-1}\beta, \\ \frac{1-\delta}{\beta}+\frac{1}{\alpha}, & \text{if } (\delta+1)^{-1}\beta \le \alpha. \end{cases}$$

Therefore

$$\begin{aligned}\sum_{1\le\alpha<\beta} CD = &\sum_{\max\{1,(\delta-1)\beta\}\le\alpha<(\delta+1)^{-1}\beta} \bigg(\frac{1-\delta}{\alpha}+\frac{1}{\beta}\bigg)\frac{2}{\beta} \\ &+ \sum_{(\delta+1)^{-1}\beta\le\alpha<\beta} \bigg(\frac{1-\delta}{\alpha}+\frac{1}{\beta}\bigg)\bigg(\frac{1-\delta}{\beta}+\frac{1}{\alpha}\bigg). \qquad (3.17)\end{aligned}$$

(We note that $(\delta-1)\beta \le (\delta+1)^{-1}\beta$ for δ between 1 and $\sqrt{2}$. For very small β we might have $1 > (\delta+1)^{-1}\beta$, in which case the first sum is empty, but that does not invalidate the argument that follows.) By Lemma 3.18, the first sum is

$$\sum_{\max\{1,(\delta-1)\beta\}\le\alpha<(1+\delta)^{-1}\beta}\bigg(\frac{1-\delta}{\alpha}+\frac{1}{\beta}\bigg)\frac{2}{\beta}$$
$$=\frac{2}{\beta}\bigg(\frac{1}{1+\delta}+1-\delta-(1-\delta)\log(\delta^2-1)\bigg)+O\bigg(\frac{1+\log\beta}{\beta^2}\bigg),$$

while the second sum has already been evaluated in equation (3.16) above. This case of the proposition then follows from (3.17) by noting that

$$\frac{2}{1+\delta}+2-2\delta-2(1-\delta)\log(\delta^2-1)+2-\delta^2-\frac{2}{\delta+1}+(\delta^2-2\delta+2)\log(\delta+1)$$
$$=4-2\delta-\delta^2+\delta^2\log(\delta+1)+2(\delta-1)\log(\delta-1).$$

- *Case 4:* $\sqrt{2}<\delta\le 2$. Just as in Case 3, we have

$$C=\begin{cases}0, & \text{if } \alpha\le(\delta-1)\beta,\\ \frac{1-\delta}{\alpha}+\frac{1}{\beta}, & \text{if } (\delta-1)\beta\le\alpha\end{cases}\quad\text{and}\quad D=\begin{cases}\frac{2}{\beta}, & \text{if } \alpha\le(\delta+1)^{-1}\beta,\\ \frac{1-\delta}{\beta}+\frac{1}{\alpha}, & \text{if } (\delta+1)^{-1}\beta\le\alpha.\end{cases}$$

However, the inequality $(\delta-1)\beta\le\alpha$ automatically implies that $(\delta+1)^{-1}\beta\le\alpha$ when $\delta\ge\sqrt{2}$. Therefore

$$\sum_{1\le\alpha<\beta}CD=\sum_{(\delta-1)\beta\le\alpha<\beta}\bigg(\frac{1-\delta}{\alpha}+\frac{1}{\beta}\bigg)\bigg(\frac{1-\delta}{\beta}+\frac{1}{\alpha}\bigg).$$

(In this case we will not need to use the precise lower bound $\max\{1,(\delta-$

$1)\beta\} \le \alpha$ for the summation over α.) This yields

$$\begin{aligned}\sum_{1\le\alpha<\beta} CD &= (1-\delta)\sum_{(\delta-1)\beta\le\alpha<\beta}\frac{1}{\alpha^2} \quad + \quad \frac{\delta^2-2\delta+2}{\beta}\sum_{(\delta-1)\beta\le\alpha<\beta}\frac{1}{\alpha} \\ &\quad + \frac{1-\delta}{\beta^2}\sum_{(\delta-1)\beta\le\alpha<\beta} 1 \\ &= (1-\delta)\biggl(\frac{1}{(\delta-1)\beta} - \frac{1}{\beta} + O\Bigl(\frac{1}{(\delta-1)^2\beta^2}\Bigr)\biggr) \\ &\quad + \frac{\delta^2-2\delta+2}{\beta}\biggl(\log\frac{\beta}{(\delta-1)\beta} + O\Bigl(\frac{1}{(\delta-1)\beta}\Bigr)\biggr) \\ &\quad + \frac{1-\delta}{\beta^2}\Bigl(\beta-(\delta-1)\beta+O(1)\Bigr) \\ &= \frac{1}{\beta}\Bigl(\delta^2-2\delta-(\delta^2-2\delta+2)\log(\delta-1)\Bigr) + O\Bigl(\frac{1}{\beta^2}\Bigr),\end{aligned}$$

where the error terms have been simplified since $\delta-1$ is bounded away from 0. This concludes the proof of the key proposition which yields Theorem 3.3. ■

3.8 Distribution of real eigenvalues

In proving Theorem 3.5, it will be convenient to define the odd function

$$G(z) = \int_0^z -\log|t|\,dt = z(1-\log|z|), \tag{3.18}$$

whose relevance is demonstrated by the following lemma.

Lemma 3.19. *If B and C are independent random variables uniformly distributed on $[-1,1]$, then the product BC has the distribution function $F_{BC}(z) = \Pr(BC < z) = \frac{1}{2}(1+G(z))$ for $z \in [-1,1]$.*

(Of course $F_{BC}(z) = 0$ for any $z < -1$, and $F_{BC}(z) = 1$ for any $z > 1$.)

Proof. Note that $|B|$ and $|C|$ are uniformly distributed on $[0,1]$. For $0 \le z \le 1$, we easily check that $\Pr(|BC| < z) = \int_0^1 \min\{1, z/s\}\,ds = G(z)$. Thus $|BC|$ is distributed on $[0,1]$ with density $f_{|BC|}(z) = -\log z$, and by symmetry BC has density $f_{BC}(z) = -\frac{1}{2}\log|z|$ on $[-1,1]$. The lemma follows upon computing $F_{BC}(z) = \int_{-1}^{z} f_{BC}(s)\,ds$. ■

It will also be helpful to define the following functions, which are symmetric in x and y:

$$\nu_1(x,y) = 1/2 + G(xy)/2 + G((x-y)^2/4), \tag{3.19}$$

$$\nu_2(x,y) = 1/2 - G(xy)/2, \tag{3.20}$$

$$\nu(x,y) = \begin{cases} \nu_1(x,y), & \text{if } xy < 1 \text{ and } x+y<0, \\ \nu_2(x,y), & \text{if } xy < 1 \text{ and } x+y>0, \\ 1 + G((x-y)^2/4), & \text{if } xy > 1 \text{ and } x+y<0, \\ 0, & \text{otherwise.} \end{cases} \tag{3.21}$$

To prove Theorem 3.5, we first consider the distribution function

$$F_W(\delta) = \int_{t<\delta} W(t)\,dt$$

associated to the density $W(\delta)$. For a random matrix M in $\mathcal{M}_2([-1,1])$ and a real number δ, we will derive an expression for the expected number of real eigenvalues of M falling below δ, then differentiate it to obtain $W(\delta)$.

It is clear that $W(-\delta) = W(\delta)$ since the set $\mathcal{M}_2([-1,1])$ is closed under negation, so it suffices to compute $W(\delta)$ for $\delta \in [0,2]$. It turns out that our calculations for F_W will be somewhat simplified by considering $F_W(-\delta)$ rather than $F_W(\delta)$.

Proposition 3.20. *We have*

$$F_W(-\delta) = \frac{1}{4}\int_{-1+\delta}^{1+\delta}\int_{-1+\delta}^{1+\delta} \nu(x,y)\,dx\,dy$$

for all $0 \le \delta \le 2$, *where* ν *is defined in equation* (3.21).

Proof. We denote the entries of M by A, B, C, D; by assumption these are independent random variables uniformly distributed in $[-1,1]$. Let δ be fixed in the range $[0,2]$, and consider the shifted matrix $M' = M + \delta I$, which we write as

$$M' = \begin{pmatrix} X & B \\ C & Y \end{pmatrix},$$

where X, Y range independently and uniformly in $[-1+\delta, 1+\delta]$ and B, C are as before. Clearly the eigenvalues of M lying below $-\delta$ correspond to the negative (real) eigenvalues of M'. By Lemma 3.14, we are free to exclude the null set where M' is singular or has repeated eigenvalues. Outside of this null set, M' has exactly one negative eigenvalue if and only if

$$\det M' = XY - BC < 0,$$

by Lemma 3.13(c). Likewise by Lemma 3.13(d), M' has exactly two negative eigenvalues if and only if

$$XY - BC > 0, \quad X + Y < 0 \quad \text{and} \quad \operatorname{disc} M' = (X-Y)^2 + 4BC > 0.$$

We thus have:

$$F_W(-\delta) = \Pr(BC > XY) + 2\Pr\left(X+Y < 0 \text{ and } -\frac{(X-Y)^2}{4} < BC < XY\right).$$

By conditioning on the values of X and Y, we may express this probability as the average value

$$F_W(-\delta) = \frac{1}{4}\int_{-1+\delta}^{1+\delta}\int_{-1+\delta}^{1+\delta} \rho(x,y)\,dx\,dy,$$

where for fixed x and y,

$$\begin{aligned} \rho(x,y) &= \Pr(BC > xy) + 2\Pr(x+y<0 \text{ and } -(x-y)^2/4 < BC < xy) \\ &= \Pr(BC > xy) + 2\Pr(-(x-y)^2/4 < BC < xy)\mathbf{1}\{x+y<0\} \end{aligned} \tag{3.22}$$

(here $\mathbf{1}\{\cdot\}$ denotes the indicator function of the indicated relation). To complete the proof it suffices to show that ρ equals the function ν defined in equation (3.21).

The probabilities appearing in equation (3.22) are effectively given by Lemma 3.19. However, there is some case-checking involved in applying this lemma, since the value of, say, $\Pr(BC > xy) = 1 - F_{BC}(xy)$ depends on whether $xy < -1$, $-1 \leq xy \leq 1$, or $xy < -1$. We make some observations to reduce the number of cases we need to examine.

Note that $(x-y)^2/4$ is bounded between 0 and 1 for any $x, y \in [-1+\delta, 1+\delta]$, so that $-(x-y)^2/4$ always lies in the interval $[-1, 1]$ prescribed by Lemma 3.19. From the identity $(x+y)^2 - (x-y)^2 = 4xy$ we see also that $xy \geq -(x-y)^2/4$. Thus xy is never lower than -1, and we need only consider whether $xy > 1$ (in which case $F_{BC}(xy) = 1$). We therefore have

$$\Pr(BC > xy) = 1 - F_{BC}(xy) = \tfrac{1}{2}(1 - G(xy))\mathbf{1}\{xy < 1\}$$

and

$$\begin{aligned} 2\Pr(-(x-y)^2/4 < BC < xy) &= 2F_{BC}(xy) - 2F_{BC}(-(x-y)^2/4) \\ &= \mathbf{1}\{xy > 1\} + G(xy)\mathbf{1}\{xy < 1\} + G((x-y)^2/4). \end{aligned}$$

Inserting these two evaluations into the formula (3.22), we obtain

$$\begin{aligned} \rho(x,y) = {} & \tfrac{1}{2}(1 - G(xy))\mathbf{1}\{xy < 1\} \\ & + \Big(\mathbf{1}\{xy > 1\} + G(xy)\mathbf{1}\{xy < 1\} + G\Big(\frac{(x-y)^2}{4}\Big)\Big)\mathbf{1}\{x + y < 0\}. \end{aligned}$$

It can be verified that this last expression is indeed equal to the right-hand side of the definition (3.21) of ν, which establishes the proposition. ∎

Since $W(\delta) = W(-\delta) = -\frac{d}{d\delta}F_W(-\delta)$, to finish the proof of Theorem 3.5 it suffices to show that $-\frac{d}{d\delta}F_W(-\delta)$ equals the expression in formula (3.4).

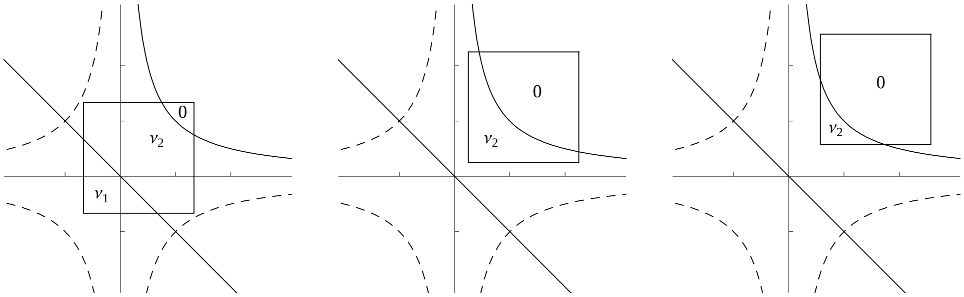

Figure 3.2: The three cases $0 \leq \delta \leq 1$, and $1 \leq \delta \leq \sqrt{2}$, and $\sqrt{2} \leq \delta \leq 2$

3.9 The derivative of the distribution

Proposition 3.20 expresses $F_W(\delta)$ as an integral of a function ν (which is independent of δ) over the square $S_\delta := [-1+\delta, 1+\delta]^2 \subset \mathbb{R}^2$. Since the region S_δ varies continuously with δ, the derivative $-\frac{d}{d\delta}F_W(-\delta)$ can be computed by an appropriate line integral around the *boundary* of S_δ. Indeed, by the fundamental theorem of calculus, we have

$$\begin{aligned}
-\frac{d}{d\delta}F_W(-\delta) &= -\frac{1}{4}\frac{d}{d\delta}\left(\int_{-1+\delta}^{1+\delta}\int_{-1+\delta}^{1+\delta}\nu(x,y)\,dx\,dy\right)\\
&= -\frac{1}{4}\left(\int_{-1+\delta}^{1+\delta}\nu(1+\delta,y)\,dy - \int_{-1+\delta}^{1+\delta}\nu(-1+\delta,y)\,dy\right.\\
&\qquad\left.+\int_{-1+\delta}^{1+\delta}\nu(x,1+\delta)\,dx - \int_{-1+\delta}^{1+\delta}\nu(x,-1+\delta)\,dx\right)\\
&= \frac{1}{2}\int_{-1+\delta}^{1+\delta}\nu(x,-1+\delta)\,dx - \frac{1}{2}\int_{-1+\delta}^{1+\delta}\nu(x,1+\delta)\,dx, \qquad (3.23)
\end{aligned}$$

where we have used the symmetry $\nu(x,y) = \nu(y,x)$ to reduce the integral to just the top and bottom edges of S_δ (where $y = 1+\delta$ and $y = -1+\delta$, respectively).

The evaluation of (3.23) divides into three cases depending on the behaviour of the indicator functions $\mathbf{1}\{x+y<0\}$ and $\mathbf{1}\{xy<1\}$ on the boundary of S_δ. Figure 3.2 depicts the possible intersections of S_δ with the line $x+y=0$ and the hyperbola $xy=1$.

- *Case 1:* $0 \leq \delta \leq 1$. For this range of δ, the line $x+y=0$ intersects the bottom edge of S_δ at $x = 1 - \delta$, while the hyperbola $xy = 1$ intersects the top edge at $x = (1+\delta)^{-1}$. Thus by the definition of ν, equation (3.23) becomes

$$-\frac{d}{d\delta}F_W(-\delta) = \frac{1}{2}\bigg(\int_{-1+\delta}^{1-\delta} \nu_1(x,-1+\delta)\,dx \\ + \int_{1-\delta}^{1+\delta} \nu_2(x,-1+\delta)\,dx - \int_{-1+\delta}^{(1+\delta)^{-1}} \nu_2(x,1+\delta)\,dx \bigg).$$

The following elementary antiderivative (easily obtained by substitution and integration by parts) hold for any fixed nonzero real number y using the definitions (3.18), (3.19), and (3.20) of G, ν_1, and ν_2:

$$\int \nu_1(x,y)\,dx = \tfrac{1}{2}x + \tfrac{1}{8}x^2y(3 - 2\log|xy|) + \tfrac{1}{36}(x-y)^3(5 - 6\log|(x-y)/2|), \\ \int \nu_2(x,y)\,dx = \tfrac{1}{2}x - \tfrac{1}{8}x^2y(3 - 2\log|xy|). \tag{3.24}$$

Therefore in this case

$$-\frac{d}{d\delta}F_W(-\delta) = \frac{1}{2}\Bigg(\Big(\tfrac{1}{2}x + \tfrac{1}{8}x^2(-1+\delta)(3 - 2\log|x(-1+\delta)|) \\ + \tfrac{1}{36}(x+1-\delta)^3(5 - 6\log|(x+1-\delta)/2|)\Big)\bigg|_{x=-1+\delta}^{1-\delta} \\ + \big(\tfrac{1}{2}x - \tfrac{1}{8}x^2(-1+\delta)(3 - 2\log|x(-1+\delta)|)\big)\Big|_{x=1-\delta}^{1+\delta} \\ - \big(\tfrac{1}{2}x - \tfrac{1}{8}x^2(1+\delta)(3 - 2\log|x(1+\delta)|)\big)\Big|_{x=-1+\delta}^{(1+\delta)^{-1}}\Bigg) \\ = \frac{80 + 20\delta + 90\delta^2 + 52\delta^3 - 107\delta^4}{144(1+\delta)} - \frac{(5 - 7\delta + 8\delta^2)(1-\delta)}{12}\log(1-\delta) \\ - \frac{\delta(1-\delta^2)}{4}\log(1+\delta)$$

(after some algebraic simplification), which verifies the first case of Theorem 3.5. (The integrands really are continuous, despite the presence of terms that could be $\log 0$, because the function G is continuous at 0; hence evaluating the integrals by antiderivatives is valid.)

- *Case 2:* $1 \le \delta \le \sqrt{2}$. Now, the line $x + y = 0$ does not intersect S_δ, while the hyperbola $xy = 1$ intersects the top edge at $x = (1+\delta)^{-1}$. Thus by the definition of ν and the antiderivative (3.24) of ν_2, equation (3.23) becomes

$$\begin{aligned}
-\frac{d}{d\delta}F_W(-\delta) &= \frac{1}{2}\left(\int_{-1+\delta}^{1+\delta} \nu_2(x, -1+\delta)\,dx - \int_{-1+\delta}^{(1+\delta)^{-1}} \nu_2(x, 1+\delta)\,dx\right) \\
&= \frac{1}{2}\Bigg(\big(\tfrac{1}{2}x - \tfrac{1}{8}x^2(-1+\delta)(3 - 2\log|x(-1+\delta)|)\big)\Bigg|_{x=-1+\delta}^{1+\delta} \\
&\qquad - \big(\tfrac{1}{2}x - \tfrac{1}{8}x^2(1+\delta)(3 - 2\log|x(1+\delta)|)\big)\Bigg|_{x=-1+\delta}^{(1+\delta)^{-1}}\Bigg) \\
&= \frac{\delta(20 + 10\delta - 12\delta^2 - 3\delta^3)}{16(1+\delta)} + \frac{(3\delta-1)(\delta-1)}{4}\log(\delta-1) \\
&\qquad + \frac{\delta(\delta^2-1)}{4}\log(\delta+1),
\end{aligned}$$

which verifies the second case of Theorem 3.5.

- *Case 3:* $\sqrt{2} < \delta \le 2$. As before, the line $x + y = 0$ does not intersect S_δ, while the hyperbola $xy = 1$ intersects the bottom edge at $x = (\delta-1)^{-1}$. Thus by the definition of ν and the antiderivative (3.24) of ν_2, equation (3.23) becomes

$$\begin{aligned}
-\frac{d}{d\delta}F_W(-\delta) &= \frac{1}{2}\int_{-1+\delta}^{(-1+\delta)^{-1}} \nu_2(x, -1+\delta)\,dx \\
&= \frac{1}{2}\big(\tfrac{1}{2}x - \tfrac{1}{8}x^2(1+\delta)(3 - 2\log|x(1+\delta)|)\big)\Bigg|_{x=-1+\delta}^{(-1+\delta)^{-1}} \\
&= \frac{\delta(\delta-2)(2 - 6\delta + 3\delta^2)}{16(\delta-1)} - \frac{(\delta-1)^3}{4}\log(\delta-1),
\end{aligned}$$

which verifies the third case of Theorem 3.5.

Since the last case of Theorem 3.5 is a consequence of Lemma 3.10, the proof of the theorem is complete. ■

Remark. One could also use the same method to extract the individual distributions of the upper and lower eigenvalues of M: for instance, eliminating the factor of 2 from equation (3.22) would yield an expression for the distribution of the lower eigenvalue of M.

3.10 Further remarks

Theorem 3.3 can also be obtained from a powerful result of Y. R. Katznelson [43], which gives an asymptotic formula for the number of singular integer matrices contained in the dilate $k \cdot \mathcal{B}$ of any convex body $\mathcal{B} \subseteq \mathbb{R}^4$ (here we are embedding the 2×2 matrices into $\mathbb{R}^4$ in the obvious way).[21] In this more general setting, the count remains proportional to $k^2 \log k$, with a constant determined by integrating $\mathcal{B}$ with respect to a certain measure.

This measure is fairly unusual: it is singular, being supported only on the hypersurface of $\mathbb{R}^4$ consisting of singular matrices. The explicit evaluation of this measure is roughly analogous to our case-by-case considerations in Section 3.7, modulo the significant complications of carrying error terms in our case. In some sense, this measure must encode the structural difference between rational eigenvalues and real eigenvalues. We expect that real eigenvalues are likewise governed by a measure that is absolutely continuous.

Katznelson's result for larger n shows that the true proportion of singular matrices in $\mathcal{M}_n(k)$ decays as $(\log k)/k^n$, with a constant depending on n that is precisely computable (at least in principle). This is a considerable sharpening of our Lemma 3.2 for $n \geq 3$, and it is very close to the obvious *lower* bound of order $1/k^n$ obtained by counting matrices that have a row of all zeros. From this one can deduce a sharper form of Theorem 3.1, that $(\log k)/k^{n-1}$ is the correct order of magnitude for the probability of having at least one integer eigenvalue (see [78] for further results along this line).

This probability is also an upper bound for the probability of diagonalizability over $\mathbb{Q}$, but when $n > 2$ we do not know the precise decay rate of the latter. We note that an obvious lower bound of order $1/k^{(n^2-n)/2}$ is given by counting upper-triangular matrices having distinct diagonal entries, and it seems reasonable to conjecture that the true order of magnitude only differs from this bound by some power (possibly dependent on n) of $\log k$.

[21] We thank Martin Widmer for directing our attention to this paper.

Chapter 4

Patterns in $x^2 + ky^2$

4.1 Introduction

In this chapter we investigate the existence of certain combinatorial patterns in the set $S(k) := \{x^2 + ky^2 : x, y \in \mathbb{Z}\}$ (we will confine our attention to the positive definite case $k \geq 1$). Let us define the two primary patterns of interest, each of which addresses a basic natural question that one might ask of any integer sequence: "how many consecutive integers can it have?" and "how far apart can two successive elements be?"

For $d \in \mathbb{N}$, we define a *run of length* d (or d*-run*) to be any string of d consecutive numbers $\{m, m+1, \ldots, m+d-1\}$. When further specificity is required, we call this the d-run starting at m. Should a given run be entirely contained in another set $S \subseteq \mathbb{N}$, we say that S *contains* this run (or some intuitively similar turn of phrase). By a *gap of size* d (or d-gap) we mean an ordered pair $(m, m+d)$, which we also think of as starting at m. We will say that S contains the gap $(m, m+d)$ when S contains m and $m+d$ but no other integer between m and $m+d$. (When $d > 1$, this is equivalent to the complement $\mathbb{N} \setminus S$ having the $(d-1)$-run starting at $m+1$.)

Our primary concern in this chapter shall be the qualitative question of whether $S(k)$ contains *infinitely* many d-runs or d-gaps for a given d. We simply wish to know whether a locally feasible pattern must occur infinitely often in $S(k)$, with comparatively less regard for the much harder problem of quantifying the precise frequency of occurrence (we discuss some quantitative aspects in the conclusion).

In the case of $S(1)$, which is the classical sums of two squares, both runs and gaps are fairly well understood. With regard to runs, one sees easily (by local considerations modulo 4) that $S(1)$ contains no run of length 4.

Conversely, it is well-known (see, for example, [14]) that $S(1)$ has infinitely many runs of length 3, which is made apparent by the relations:

$$\begin{aligned} 4n^4 + 4n^2 &= (2n^2)^2 + (2n)^2, \\ 4n^4 + 4n^2 + 1 &= (2n^2+1)^2 + 0^2, \\ 4n^4 + 4n^2 + 2 &= (2n^2+1)^2 + 1^2. \end{aligned}$$

Sharper quantitative results are known, but only for short runs: Hooley [39] proved that the number of 2-runs in $S(1)$ up to x satisfies the lower bound

$$\#\{n \le x : \{n, n+1\} \subset S\} \gg \frac{x}{\log x}. \tag{4.1}$$

By standard sieve methods (for instance, Selberg's sieve as used in [14]), there is an upper bound of exactly the same order of magnitude, meaning that Hooley's bound is best possible, aside from a constant factor.

In general, the length of runs appearing in any $S(k)$ is *uniformly bounded*: we will see in the next section that $S(k)$ never contains any run of length 6 or higher, which follows from a simple local argument. However, it is possible to find runs of length 5 in certain sets $S(k)$, and our main interest here lies in this extremal case: *which sets $S(k)$ admit infinitely many 5-runs?*

M. Rosenfeld (personal communication) previously observed that $S(2)$ has many such runs, the first of which begin at 0, 96, 800, 2400, 3200 and 3648. This led him to conjecture that there are infinitely many 5-runs in $S(2)$, and we affirm this with the first main result of this chapter:

Theorem 4.1. *There are infinitely many numbers m for which*

$$\{18m^2,\, 18m^2+1,\, 18m^2+2,\, 18m^2+3,\, 18m^2+4\} \subset S(2).$$

The proof of Theorem 4.1 uses a direct construction, and the same technique can be used (with slight modifications) with certain other values of k, such as $k = 46$. However, we will see in Section 4.2 that there is a positive density of k for which $S(k)$ "should" contain 5-runs (in that there are no local obstructions prohibiting so); our method, unfortunately, only works for k in a set of asymptotic density 0. (The precise conditions are somewhat tedious, and we defer their discussion to Section 4.3.) However, it does allow us to establish the following general theorem about larger values of k:

Theorem 4.2. *There are infinitely many $k > 2$ for which $S(k)$ contains infinitely many 5-runs.*

The question of gaps in $S(1)$ was answered by Spiro and Hensley [79], who proved the universal existence of gaps: *for any $d \in \mathbb{N}$, $S(1)$ contains infinitely many gaps of size d.* Just as in our theorems for runs, their proof is explicitly constructive; we greatly generalize this construction to all sets $S(k)$, proving universal existence with a single type of exception:

Theorem 4.3. *Fix any $k, d \in \mathbb{N}$ subject to the restriction that either $4 \nmid k$ or $d \not\equiv 2 \pmod 4$. Then there exists an explicitly computable quadratic polynomial $P(n)$ such that $S(k)$ has the gap $(P(n), P(n) + d)$ for all n.*

Note that the restriction in the hypotheses is entirely necessary: if $4 \mid k$, then $S(k)$ consists solely of integers congruent to 0 or 1 modulo 4, so the difference between two elements of $S(k)$ is never of the form $4m + 2$.

4.2 Local feasibility of runs

We first examine all local obstructions to the existence of d-runs in $S(k)$, in particular obtaining necessary conditions for $S(k)$ to admit 5-runs. The underlying principle here is to consider the set of values attained by the quadratic form $x^2 + ky^2$ for x, y in the p-adic integers $\mathbb{Z}_p$ for some p; more directly, we are considering the range of $x^2 + ky^2$ modulo p^r for every prime p and each $r \geq 1$. If there is any prime p for which this set does not contain d consecutive values, then clearly $S(k)$ does not have any d-runs.

Let us first dispense with the 2-adic case, which already reveals a uniform upper bound on d, as shown by the proposition below.

Proposition 4.4. *The length of any run in $S(k)$ is at most 5. Moreover, if $S(k)$ achieves this length, then $k \equiv 2 \pmod 4$.*

$k \pmod 8$	$x^2 + ky^2 \pmod 8$	Run length
$0, 4$	$0, 1, 4$	2
1	$0, 1, 2, 4, 5$	3
2	$0, 1, 2, 3, 4, 6$	5
$3, 7$	$7, 0, 1, 3, 4$	3
5	$0, 1, 4, 5, 6$	3
6	$6, 7, 0, 1, 2, 4$	5

Table 4.1: Distribution of $x^2 + ky^2$ modulo 8

Proof. This follows by straightforward (but tedious) calculation of all possible residues of $x^2 + ky^2$ modulo 8. Table 4.1 summarizes the possible values of $x^2 + ky^2$ depending on the congruence class of k modulo 8, and the longest resulting string of residues. We have ordered the residues so as to make the maximum number of consecutive values more apparent. ∎

We know from the introduction that the upper bound of the proposition is tight; for instance $S(2)$, does contain 5-runs. This motivates us to try to characterize all k for which $S(k)$ has any 5-runs. Let us complete our analysis of local obstructions with an eye toward this central question. We first recall the well-known structure of p-adic squares [76, Ch. II, §3.3]:

Proposition 4.5. *If p is odd and $p \nmid x$, then x is a square modulo p^r if and only if x is a square mod p. If $p = 2$, $r \geq 3$ and x is odd, then x is a square modulo 2^r if and only if $x \equiv 1 \pmod 8$.*

In other words (after factoring out powers of p from x), the "squareness" of x modulo some high prime power p^r depends only on its quadratic character modulo a fixed prime power (either p, or 2^3). The next fact shows that for most p, $S(k)$ covers all congruence classes mod p. It follows from a well-known counting argument [65, Theorem 5.14]: the sets $\{x^2 : x \in \mathbb{Z}/p\mathbb{Z}\}$ and $\{b - ky^2 : y \in \mathbb{Z}/p\mathbb{Z}\}$ are simply too large not to intersect.

Proposition 4.6. *If $p \nmid k$, then the equation $x^2 + ky^2 \equiv b \pmod p$ has a solution for any fixed $b \in \mathbb{Z}/p\mathbb{Z}$.*

We will also need a result of Buell and Hudson [13] on runs of consecutive quadratic residues.

Proposition 4.7. *For any d, every sufficiently large prime p has d consecutive quadratic residues (also d consecutive nonresidues). In particular, for $d = 5$ this holds for all $p \geq 197$.*

Theorem 4.8. *The existence of 5-runs in $S(k)$ is locally feasible if and only if $k \equiv 2 \pmod 4$ and none of the following primes divide k:*

$$3, 5, 7, 11, 13, 19, 29, 31, 37, 53, 149.$$

Proof. For 2-adic feasibility, we look to Proposition 4.4, which clearly shows $k \equiv 2 \pmod 4$ is necessary; since this bounds the power of 2 dividing k, Table 4.1 essentially also shows 2-adic sufficiency.[22]

Let us now consider p-adic feasibility for some odd $p \nmid k$. By Proposition 4.6, $x^2 + ky^2$ attains all possible residues modulo p, giving an "infinite" run in $\mathbb{Z}/p\mathbb{Z}$. Not all of these necessarily lift to arbitrary p-adic values, but certainly any nonzero residue will (Proposition 4.5), and 0 itself is certainly attained. This yields the run $\{-p+1, \ldots, p-1\}$ of length $2p - 1 \geq 5$; such primes pose no obstruction to local feasibility.

It remains to consider odd primes p which do divide k. Since $x^2 + ky^2$ reduces to x^2 modulo p, a necessary condition for p-adic feasibility is that there are 5 consecutive squares in $\mathbb{Z}/p\mathbb{Z}$. We claim it is also sufficient (note that $p > 5$ in such case): this is obvious by Proposition 4.5 if the 5 consecutive squares are nonzero, but even when one of the squares is 0, the same proposition allows us to lift the other 4 squares so that they remain consecutive modulo p^r for all r.

By Proposition 4.7, certainly all $p \geq 197$ have 5 consecutive squares (since we count 0 as a square but it is not a quadratic residue, our requirement is weaker). The theorem follows by explicitly computing which primes $p < 197$ fail to have 5 consecutive squares mod p. These are precisely the primes listed in the statement of the theorem. ∎

[22] To be sure, one could extend the table to modulo 16, but nothing new is revealed.

As an immediate consequence, the set of k for which $S(k)$ *locally* admits 5-runs has a positive density of about 8%.

Corollary 4.9. *The set of $k \in \mathbb{N}$ satisfying the conditions of Theorem 4.8 has density equal to*

$$\frac{2^{15} \cdot 3^6 \cdot 5}{11 \cdot 19 \cdot 29 \cdot 31 \cdot 53 \cdot 149} \approx 0.0804969.$$

Proof. This is just $\frac{1}{4}\phi(n)/n$, where n is the product of the listed primes. ∎

4.3 Construction for $k = 2$

Clearly $k = 2$ is the first integer to satisfy the conditions of Theorem 4.8, and we observed in the first section that $S(2)$ contains several 5-runs. Let us now give the simple construction of infinitely many quintuples in $S(2)$.

Proof of Theorem 4.1. Three of the stated inclusions are immediate for any m, from the easy identities

$$\begin{aligned} 18m^2 &= (4m)^2 + 2(m)^2 \\ 18m^2 + 1 &= (1)^2 + 2(3m)^2 \\ 18m^2 + 4 &= (2)^2 + 2(3m)^2. \end{aligned}$$

For the other two terms, we will show that m may be chosen simultaneously equal to $9a^2 + 4a$ and $3b^2 + b$, whence m satisfies the additional identities

$$\begin{aligned} 18m^2 + 2 &= (4m + 2a)^2 + 2(m - 4a - 1)^2, \\ 18m^2 + 3 &= (4m + 2b - 1)^2 + 2(m - 4b - 1)^2. \end{aligned}$$

Let us verify that the Diophantine equation $9a^2 + 4a = 3b^2 + b$ has infinitely many solutions (it has at least the basic solution $m = a = b = 0$). Making the substitution $c = 18a + 4, d = 6b + 1$ yields a Pell-type equation with modular restrictions:

$$c^2 - 3d^2 = 13, \quad c \equiv 4 \pmod{18}, \quad d \equiv 1 \pmod{6}. \tag{4.2}$$

This has an infinite family of solutions with the initial solution $(c_0, d_0) = (4, 1)$ and recurrence $(c_{n+1}, d_{n+1}) = (2c_n + 3d_n, c_n + 2d_n)$. One easily verifies that the subsequence (c_{6n}, d_{6n}) satisfies the necessary congruences (4.2). ■

Remark. Since the ring $\mathbb{Z}[\sqrt{-2}]$ is a unique factorization domain, the existence of *rational* values x, y with $n = x^2 + 2y^2$ is enough to conclude that $n \in S(2)$. Thus, the above construction actually produces runs in $S(2)$ whenever $18m^2$ is an integer, regardless of the integrality of c and d. It turns out that this relaxation is satisified for the denser subsequence (c_{2n}, d_{2n}), which yields the run at 800; in contrast, the first run given by (c_{6n}, d_{6n}) occurs at the much larger 1224111751200.

4.4 Construction for $k > 2$

We now generalize the construction of Theorem 4.1 to certain other values of k. We will see that under certain conditions it is possible to take a single 5-run in $S(k)$, and generate from it an infinite sequence of 5-runs. Note that $S(2)$ is unique in that it is the only $S(k)$ containing the run $\{0, 1, 2, 3, 4\}$, which one might call the "generator" of Theorem 4.1 since it corresponds to (c_0, d_0). While the strategy in this section also works for $S(2)$, it cannot be used with the run $\{0, 1, 2, 3, 4\}$ and thereby yields a less efficient construction. We first require an elementary proposition regarding generalized Pell equations. It is certainly well-known in various equivalent forms; the version below will be convenient to us.

Proposition 4.10. *Let A, B, C, D be fixed integers such that $BD > 0$ is non-square. If the system*

$$m = \frac{x^2 + A}{B} = \frac{y^2 + C}{D} \tag{4.3}$$

has at least one integer solution (m, x, y) with $(x, y) \neq (0, 0)$, then it has infinitely many.

Proof. Let (m_0, x_0, y_0) be an initial solution to (4.3). Rearranging the right-hand equality and substituting $z = By$ yields the Pell equation

$$z^2 - BDx^2 = B(AD - BC). \tag{4.4}$$

Define $N := BD$ and let $a + b\sqrt{N}$ be a fundamental unit in $\mathbb{Z}[\sqrt{N}]$. Then (4.4) has an infinite family of integer solutions (z_n, x_n) given by

$$\begin{pmatrix} z_n \\ x_n \end{pmatrix} = \begin{pmatrix} a & Nb \\ b & a \end{pmatrix}^n \begin{pmatrix} z_0 \\ x_0 \end{pmatrix},$$

where $z_0 = By_0$. Let M denote the 2×2 matrix above. The terms of this sequence are necessarily distinct because $(z_0, x_0) \neq (0, 0)$ and the eigenvalues $a \pm b\sqrt{N}$ of M are not roots of unity. If r is the order of M in the finite group $\mathrm{SL}_2(\mathbb{Z}/B\mathbb{Z})$, then the subsequence (z_{rn}, x_{rn}) is congruent to (z_0, x_0) modulo B. Thus, $x_{rn} \equiv x_0 \pmod{B}$ and $z_{rn} \equiv z_0 \equiv 0 \pmod{B}$, which ensures that both $y_{rn} = z_{rn}/B$ and $m_{rn} = (x_{rn}^2 + A)/B$ are integers. ■

We can now give a sufficient condition for the existence of infinitely many 5-runs within a particular $S(k)$.

Theorem 4.11. *Let $k > 2$ be fixed, and suppose there exists $m > 0$ such that $km^2 + 2$ and $km^2 + 3$ belong to $S(k)$, hence $km^2 + 2 = a^2 + kb^2$ and $km^2 + 3 = c^2 + kd^2$ for some $a, b, c, d \in \mathbb{Z}$. If $(m - b)(m - d)$ is not a square, then $S(k)$ contains infinitely many 5-runs.*

Proof. Just as for Theorem 4.1, the inclusion $\{km^2, km^2+1, km^2+4\} \subset S(k)$ is obvious, so it suffices to produce infinitely many m for which $\{km^2 + 2, km^2 + 3\} \subset S(k)$.

We make the substitutions $B = m - b$ and $D = m - d$, whereupon the conditions $km^2 + 2 = a^2 + kb^2$ and $km^2 + 3 = c^2 + kd^2$ are equivalent to

$$m = \frac{a^2 + (kB^2 - 2)}{2Bk} = \frac{c^2 + (kD^2 - 3)}{2Dk}. \tag{4.5}$$

In light of Proposition 4.10, we may hold B and D fixed and obtain infinitely many solutions to (4.5) in m, a, c, provided that the initial solution satisfies $(a, c) \neq (0, 0)$ and $BD > 0$ (we have assumed by hypothesis that BD is not a square, and so neither is $4BDk^2$). But any solution to (4.5) for $k > 2$ satisfies $a^2 \equiv 2 \pmod{k}$, which means $a \neq 0$. Furthermore, we must have $b \neq m$ or else $a^2 = 2$; we must have $b \leq m$ or else $a^2 = 2 + k(m^2 - b^2) < 2 - k < 0$. Thus $B > 0$ and likewise $D > 0$ by an identical argument. ∎

The condition on $(m-b)(m-d)$ in Theorem 4.11 can easily be weakened by observing that we may freely substitute $b \leftarrow -b$ or $d \leftarrow -d$ to modify the quantity $(m-b)(m-d)$. The conclusion can only fail if *all* such values that result are simultaneously squares. We conjecture that such a coincidence cannot occur (in which case Theorem 4.11 would hold without this additional restriction), but we are unable to prove this. In any case, the previous observation immediately gives the following.

Corollary 4.12. *The conclusion of Theorem 4.11 also holds if any of the following quantities is not a square:* $(m-b)(m-d), m^2 - b^2$, *or* $m^2 - d^2$.

4.5 How many values of k?

The preceding corollary allows us to rapidly establish the existence of infinitely many 5-runs in $S(k)$ for a specific value of k, by finding a single one of the appropriate form. We have used this to verify, for 91 different values of $k \leq 5000$, the existence of infinitely many 5-runs — we have yet to find any generator that fails to satisfy the non-squareness conditions, which are presumably unnecessary.

However, there is a rather strict condition on k implicit in this method: in order for $km^2 + 2$ and $km^2 + 3$ to belong to $S(k)$, both 2 and 3 *must* be squares modulo k, which limits which primes p can divide k. Combining this observation with Theorem 4.8, we see by quadratic reciprocity that the following condition is necessary:

Condition 4.13. *k is equal to* 2 *times a product of (not necessarily distinct) primes* $p \equiv \pm 1 \pmod{24}$.

Our computations suggest that this is the only restriction: the 91 values of k mentioned above are exactly those values of $k \leq 5000$ which satisfy Condition 4.13. Even so, the set of such k still has asymptotic density 0, in contrast to the positive density given by Corollary 4.9. The method of Wirsing and Odoni (for example, [21, Proposition 4]) can be used to make this quantitatively precise: the number of $k \leq x$ satisfying this condition is asymptotic to $x/(\log x)^{3/4}$ times an explicit constant.

While $S(34)$ does contain 5-runs (and computations suggest that they occur with approximately the expected frequency), it does not satisfy Condition 4.13, so we have no proof that it has infinitely many. It is natural to ask if there are infinitely many k for which the construction necessarily works: this is answered by Theorem 4.2, which we now prove. Conveniently, the proof uses Proposition 4.10 again; however, the Pellian construction inevitably gives a sequence of k that is far sparser than $x/(\log x)^{3/4}$.

Proof of Theorem 4.2. We apply Theorem 4.11, which requires that we find families of solutions to $km^2 + 2 = a^2 + kb^2$ and $km^2 + 3 = c^2 + kd^2$. One such solution is given for $k = 2$ by

$$\begin{aligned} 2(20)^2 + 2 &= 802 = 28^2 + 2(3)^2 \\ 2(20)^2 + 3 &= 803 = 9^2 + 2(19)^2. \end{aligned}$$

Now, observe that holding $(m, b, d) = (20, 3, 19)$ fixed and solving for k gives

$$k = \frac{a^2 - 2}{m^2 - b^2} = \frac{c^2 - 3}{m^2 - d^2}. \tag{4.6}$$

We now apply Proposition 4.10 to obtain an infinite sequence of tuples (k, a, c) satisfying (4.6), starting with $(2, 28, 9)$ and noting that $m^2 - b^2 = 17 \cdot 23$ and $m^2 - d^2 = 3 \cdot 13$ satisfy the conditions on B and D. Since m, b and d are held constant and $(m-b)(m-d) = 17$, the conditions of Theorem 4.11 are satisfied across this entire sequence. The resulting values of k are clearly distinct, since a and c are determined up to sign by k. ∎

4.6 Preliminaries for gaps

We now turn to the question of gaps in $S(k)$, with the eventual goal of proving Theorem 4.3. The general strategy for the proof is simple: we first construct a polynomial $P_0(y)$ such that $\{P_0(y), P_0(y)+d\} \subset S(k)$ for any y. We then choose an arithmetic progression for y so that $P_0(y)+j \notin S(k)$ for $0 < j < d$ for any y in this progression; passing to this progression immediately yields the desired polynomial $P_0(an+b)$ in the conclusion of the theorem. While this basic outline is no different from that used by Spiro and Hensley [79], for general k there are some subtleties in choosing P_0 that did not appear there.

We will require several propositions which should be considered well-known (the first was essentially due to Euler in his characterization of sums of two squares). We provide the simple proofs here.

Proposition 4.14. *If a prime p satisfies $\left(\frac{-k}{p}\right) = -1$, and $p \parallel n$, then $n \notin S(k)$.*

Proof. This follows immediately from the fact that $x^2 + ky^2 \equiv 0 \pmod{p}$ has only the trivial solution $x \equiv y \equiv 0 \pmod{p}$ (else $-k$ would have the square root xy^{-1} modulo p). ■

Proposition 4.15. *If $q_1 < q_2 < \cdots < q_r$ are primes and $q_0 = -1$, then for any choice of $\epsilon_0, \epsilon_1, \epsilon_2, \ldots, \epsilon_r \in \{\pm 1\}$, there are infinitely many odd primes p which satisfy the constraints $\left(\frac{q_i}{p}\right) = \epsilon_i$ for each $0 \le i \le r$.*

Proof. Recall that the constraint $\left(\frac{-1}{p}\right) = \epsilon_0$ uniquely determines p as either 1 or 3 mod 4. For each odd q_i, quadratic reciprocity then gives $\left(\frac{q_i}{p}\right) = \epsilon_i$ for p in any of $\frac{1}{2}(q_i - 1)$ congruence classes mod $4q_i$ (depending on the choice of ϵ_0). Similarly, if $q_1 = 2$, we easily check that the equation $\left(\frac{q_1}{p}\right) = \epsilon_i$ admits a (unique) solution modulo 8, for any choice of ϵ_0 and ϵ_1. By Chinese remaindering, we obtain a congruence class for p modulo $4\prod_i q_i$, in which all constraints are simultaneously satisfied. The proposition now follows from Dirichlet's theorem. ■

Lemma 4.16. *Let A, B be positive integers such that AB is not a square. There are infinitely many primes p such that $\left(\frac{-A}{p}\right) = -1$ and $\left(\frac{-B}{p}\right) = 1$.*

Proof. We assume without loss of generality that A and B are squarefree, as A and An^2 have identical Legendre symbols, apart from the finitely many primes dividing n. Since AB is non-square (and positive), there exists some prime q dividing exactly one of A and B. If $q \mid A$, we may use Proposition 4.15 to choose p such that $\left(\frac{-1}{p}\right) = 1$, $\left(\frac{q}{p}\right) = -1$, and $\left(\frac{q'}{p}\right) = 1$ for all other primes q' dividing AB. Likewise, if $q \mid B$, we choose $\left(\frac{-1}{p}\right) = \left(\frac{q}{p}\right) = -1$, and $\left(\frac{q'}{p}\right) = 1$ for all other q' dividing AB. ∎

Remark. This lemma is a special case of the more general principle that, if $A_1, A_2, \ldots, A_r$ are multiplicatively independent integers (in the sense that no non-empty subset has a product equal to a square), then for any assignment of ± 1 values to the Legendre symbols $\left(\frac{A_i}{p}\right)$, there are infinitely many primes p which satisfy that assignment (in fact the set of such p has density exactly 2^{-r} relative to the set of all primes).

Proposition 4.17. *Let $f(x) = ax^2 + bx + c$ be a quadratic with integer coefficients and discriminant $\Delta = b^2 - 4ac$. Suppose p is an odd prime such that $p \nmid a$ and $\left(\frac{\Delta}{p}\right) = 1$. Then there exists a number m such that $p \parallel f(m)$.*

Proof. Under the given assumptions, the quadratic formula readily yields a solution to $f(m) \equiv p \pmod{p}$, namely $m \equiv (-b + \sqrt{\Delta})/2a \pmod{p}$. Furthermore, since Δ is nonzero mod p, the derivative $f'(m) = 2am + b$ is non-vanishing, and so Hensel's lemma ensures that we can lift m to a solution of $f(m) \equiv p \pmod{p^2}$, whence p exactly divides $f(m)$. ∎

Proposition 4.18. *Let be $P(n)$ be a nonconstant polynomial of degree r, and $B > 0$ a fixed integer. Then the asymptotic density of n for which $P(n)$ is B-friable (has no prime factors exceeding B) is 0.*

Proof. By an elementary counting argument, the number of distinct B-friable values up to y is $\ll (\log y)^B$. For any $n \leq x$, $P(n)$ is bounded by Cx^r (for some constant C) and takes each value at most r times, so there are at most $\ll_{r,P} r(r\log x + C)^B \ll_{r,P} (\log x)^B$ values of $n \le x$ where $P(n)$ is B-friable. Since B is fixed (one could even allow B to grow slowly with x), this set has asymptotic density 0. ∎

4.7 Proof of existence of gaps

Proof of Theorem 4.3. Let $P_0(y)$ be the quadratic polynomial $x^2 + ky^2$, where x is a linear function of y with integer coefficients to be chosen shortly. Clearly, we have $P_0(y) \in S(k)$ for any y. To ensure that $P_0(y) + d \in S(k)$, we write $P_0(y) + d = (x+\delta)^2 + k(y-\eta)^2$, where δ and η are constants to be chosen below. This is equivalent to $d = 2\delta x + \delta^2 - 2k\eta y + k\eta^2$, or

$$x = \frac{d - \delta^2 - k\eta^2}{2\delta} + \frac{k\eta}{\delta} y. \tag{4.7}$$

For a given k and d, we will fix δ and restrict η so that the above expression, viewed as a linear function of y, has integer coefficients. This can be done in several cases as follows:

- If d is odd, we may simply choose $\delta = 1$ and $\eta = 0$ (so x is constant).
- If d is even and k is odd, we choose $\delta = 1$ and restrict η to be odd.
- If $d \equiv 0 \pmod 4$, we choose $\delta = 2$ and η even.
- If $d \equiv k \equiv 2 \pmod 4$, we choose $\delta = 2$ and η odd.

Note that these cases cover all possible choices of k and d satisfying the hypotheses (there is a small amount of overlap between the cases, which is not important). For any η subject to the above restrictions, we now choose x as determined by (4.7) (which is linear in y), so that $x^2 + ky^2 + d \in S(k)$ for any y.

It remains to choose a specific η subject to the above restrictions, as well as an arithmetic progression for y that induces a genuine gap of size d at $x^2 + ky^2$: we need $x^2 + ky^2 + j \notin S(k)$ for every $0 < j < d$.

In the first case that d is odd, we have already chosen $\eta = 0$. For all other cases, we fix η to be any integer (subject to the parity constraints above) such that $\delta^2 + k\eta^2$ has a prime factor q which exceeds $2d$. The existence of such an η is guaranteed by Proposition 4.18, as very few choices of $\delta^2 + k\eta^2$ are $2d$-friable.

To exclude $x^2 + ky^2 + j$ from $S(k)$ for a given j, it suffices by Proposition 4.14 to exhibit a prime p_j such that $\left(\frac{-k}{p_j}\right) = -1$ and $p_j \parallel x^2 + ky^2 + j$ for some appropriate choice of y. Recall that x and y are not independently chosen but related by (4.7), which we abbreviate as $x = Cy + D$ by taking

$$C := \frac{k\eta}{\delta}, \quad D := \frac{d - \xi}{2\delta}, \qquad \text{where } \xi := \delta^2 + k\eta^2.$$

Then $x^2 + ky^2 + j$, as a function of y, is the quadratic $P_0(y) = (C^2 + k)y^2 + 2CDy + (D^2 + j)$. In order to apply Proposition 4.17, we must verify that the discriminant of P_0 is a nonzero square mod p_j (we also need p_j to be coprime to $C^2 + k$, but since C and k are fixed, there are only finitely many values of p_j to exclude). Specifically, we need

$$\left(\frac{C^2D^2 - (C^2 + k)(D^2 + j)}{p_j}\right) = \left(\frac{-kD^2 - j(C^2 + k)}{p_j}\right) = 1. \tag{4.8}$$

By Lemma 4.16, there are infinitely many odd primes p_j satisfying both $\left(\frac{-k}{p_j}\right) = -1$ and (4.8), provided that the product $k^2D^2 + jk(C^2 + k)$ is not a square. Noting that $C^2 + k = k\xi/\delta^2$, this is equivalent, up to a square factor of $k^2/4\delta^2$, to

$$\Delta_j := (d - \xi)^2 + 4\xi j. \tag{4.9}$$

We claim that Δ_j is not square for any value of $0 < j < d$. Note that this quantity is unavoidably square at the endpoints $\Delta_0 = (d - \xi)^2$ and $\Delta_d = (d + \xi)^2$, by virtue of the construction.

When d is odd, the fact that Δ_j is not square is immediate: we chose $\delta = 1$ and $\eta = 0$, so $\xi = 1$ and there are no even squares between $(d - 1)^2$ and $(d + 1)^2$.

When d is even, recall that we chose η so that $\xi = \delta^2 + k\eta^2$ has a large prime factor $q > 2d$. In this case ξ is now larger than d, so it is more natural to write Δ_j as $(\xi - d)^2 + 4\xi j$.

We now show that $(\xi - d)^2 + 4\xi j = A^2$ admits no integer solutions for $0 < j < d$, or equivalently with $\xi - d < A < \xi + d$. Since $q \mid \xi$, any such solution must have $A^2 \equiv d^2 \pmod{q}$, whence $A - \xi \equiv A \equiv \pm d \pmod{q}$. But d and $-d$ are necessarily the smallest representatives of their respective congruence classes mod q (because $q > 2d$), so we must have $|A - \xi| \geq d$.

Putting this all together, for each $0 < j < d$ we may now choose a distinct prime p_j satisfying $\left(\frac{-k}{p_j}\right) = -1$ so that $p_j \parallel P_0(m_j) + j$ for some m_j. By choosing $y \equiv m_j \pmod{p_j^2}$ simultaneously for all $0 < j < d$, we obtain a congruence class for y modulo $\prod_j p_j^2$ such that $P_0(y) + j \notin S(k)$, so that $(P_0(y), P_0(y) + d)$ is indeed a gap of size d. ■

Remark. The above proof yields a quantitative lower bound of $\gg_{k,d} \sqrt{x}$ gaps of size d in $S(k)$ up to x, where the implied constant decays rather rapidly with d. In the case of $S(1)$ we can do much better: Hooley's quantitative lower bounds for the number of pairs $\{m, m+1\}$ in $S(1)$ can be generalized to yield a lower bound of $\gg_d x/\log x$ for the number of $\{m, m+d\}$ pairs in $S(1)$. While not every such pair is necessarily a gap, Selberg's sieve shows for any $0 < c < d$, the triple $\{m, m+c, m+d\}$ appears only $\ll_{c,d} x/(\log x)^{3/2}$ times. Since there are only $O_d(1)$ possible choices for c, this shows that the majority of pairs $\{m, m+d\}$ have no other elements in between, and are genuine gaps.

Chapter 5

Conclusion

5.1 Arithmetic functions

We studied a broad class of arithmetic functions $f(n)$ in Chapter 2, showing that (with some restrictions) for any $\varepsilon > 0$ and any choices for $\alpha_1, \alpha_2, \ldots, \alpha_h$, the approximations

$$|f_i(a_i n + b_i) - \alpha_i| < \epsilon \quad (i = 1, 2, \ldots, h) \tag{5.1}$$

occur simultaneously for a *positive proportion* of values n. This generalizes the Erdős–Schinzel theorem in two directions: one may freely compose the functions with different arithmetic progressions, and also freely choose a function at each coordinate $1 \leq i \leq h$.

Let us offer an interpretation of this latter aspect that seems intuitively attractive. To paraphrase the earlier results, a single value of n may aim at several targets simultaneously, using the values of a fixed arithmetic function; moreover, each individual function (such as $\sigma(n)/n$ or $\phi(n)/n$) gives rise to a large set of such approximants n. What our Corollary 2.11 adds is that the sets arising from different functions also have large *intersection*: they cannot conspire to avoid one another — unless condition (2.16) is violated, making them too interdependent. We might say they are distributed "randomly" enough that their intersections behave (up to an extremely large local factor) like those of random sets having similar size.

We suspect that there is not much room for further generalization if our goal is to retain a positive proportion of approximants n. However, we have already described the contemporary work of Alkan, Ford and Zaharescu which passes up this requirement in exchange for a much more effective approximation (one whose accuracy improves nicely with n).

More recently, Luca, Mejía Huguet and Nicolae [51] have taken such results in a novel direction: they prove a result analogous to Schinzel and Wang's (1.5), except where $\phi(n)$ is composed with the Fibonacci sequence F_n. In other words, the following sequence of points is also dense in the positive orthant of $\mathbb{R}^d$:

$$\left(\frac{\phi(F_{n+1})}{\phi(F_n)}, \frac{\phi(F_{n+2})}{\phi(F_n)}, \dots, \frac{\phi(F_{n+d})}{\phi(F_n)}\right).$$

Their proof makes use of the fact that Fibonacci numbers, although exponentially sparse, have a rich divisor structure that enables the sort of constructions we saw in the proof of Theorem 2.3; the same techniques work for similar sequences like $2^n - 1$ (but not 2^n, which is very poor in divisors). Many difficult problems on arithmetic functions can be solidly answered when restricted to these sequences. For instance, while no one has entirely ruled out odd perfect numbers, Luca [50] has shown that no Fibonacci number is perfect, and Pollack [67] proved that there are no perfect numbers composed of the same digit repeatedly (excluding the trivial example 6).

Seeing that approximation properties are robust against composing with such exotic sequences rather than linear functions, we think it would be interesting to consider compositions with more natural (but less charming) sequences, such as higher-degree polynomials.

5.2 Random matrices

In Chapter 3, we studied questions regarding the number of $n \times n$ integer matrices having a particular eigenvalue. In particular, we established upper bounds for the singularity probability in general, as well as the exact distribution of real and rational eigenvalues for $n = 2$. The latter investigation revealed some rather unexpected differences between the two, and the location of the mode(s) strikes the author as the most surprising aspect worthy of investigation: for instance, might further modes appear as n increases, and is there a natural explanation for their appearance?

Our singularity bounds have been considerably improved by Shparlinski [78], who obtained stronger bounds, more generally for matrices shifted by adding a constant matrix (also more generally, having a prescribed determinant rather than just 0). As mentioned in the last section of said chapter, the precise number of integer matrices which are diagonalizable over $\mathbb{Q}$, or have all eigenvalues simultaneously rational, remains open for any $n \geq 3$. We know of only one result specific to such matrices, in the special case of symmetric $(0, 1)$-matrices, also corresponding to *integral graphs.* These were considered by Ahmadi et al. [1] and shown to occur with exponentially small probability as n increases. Also, Maze [57] has given very precise distributional results on the fine structure[23] of random integer matrices.

Just as in the eigenvalue problem, there is a wealth of results on the singularity probability in the converse case where we fix the size of entries k (thus fixing the distribution of each entry) and let n increase. Komlós [46] first showed that large random matrices are rarely singular, regardless of the distribution; that is, as $n \to \infty$, the singularity probability shrinks to 0 for any fixed k. In fact, it decreases exponentially quickly, as first shown by Kahn, Komlós and Szemerédi [42] for $(0, 1)$-matrices, and in much greater generality by Rudelson and Vershynin [69].

Recently, Bourgain, Vu and Wood [10] have proved an impressive general bound for the singularity probability, one that incorporates both the polynomial decay observed in k (for fixed n), and the exponential decay seen in n (when k fixed). For n sufficiently large their bound, in the case of the matrices $\mathcal{M}_n(k)$, is essentially $(2k)^{-n/2}$. It would be interesting to explore how explicit such a bound could be made, which might yield a far more effective version of Proposition 3.6. The author's recent efforts in this area are concerned with obtaining explicit lower bounds for the magnitude of the determinant of a typical random matrix (the determinant may be viewed as a crude measure of how far away a matrix is from being singular; once again, there is a rich history of bounding determinants away from zero in various other contexts, for instance [69, 81]).

[23] However, this concerns the Hermite normal form (a division-free version of row echelon form) rather than the spectrum.

5.3 Patterns in quadratic forms

We have studied the patterns of runs and gaps in the set $S(k) = \{x^2 + ky^2 : x, y \in \mathbb{Z}\}$. We gave a qualitatively complete answer to the existence of gaps for all $S(k)$, but significant quantitative questions remain. For instance, the *largest* gap in $S(k)$ up to x is certainly at least $\sqrt{\log x}$, and heuristically should be not much larger (a power of $\log x$). However, the best known upper bound for this quantity is $O(x^{1/4})$, and it comes from an obvious argument: [59, p. 43] describes this state of the art as "somewhat embarrassing".[24]

Even qualitatively, much remains unknown about runs, with $S(34)$ being the first of many examples for which the existence of infinitely many 5-runs is open. One avenue of pursuit considered by the author has been to find new polynomial identities which could take the place of the trivial inclusions in the proofs of Theorem 4.1 and Theorem 4.11, thus extending the reach of this method beyond current limitations.

For more general binary quadratic forms, we noted previously that they may (locally) admit extremely long runs, for instance the (positive definite) form $xy+k!(x^2+y^2)$ for large k. Such forms are certainly worth investigating and are likely amenable to the methods of Theorem 4.3, but clearly new ideas are needed to tackle runs of any significant length. It seems unlikely that analytic methods can prevail for runs of such length, either: even for 3-runs, we know of no result remotely comparable in strength to (4.1).

Considering the spectacular success of Green, Tao and Ziegler in enumerating arithmetic progressions of primes, one naturally expects to obtain better quantitative results for APs. Matthiesen [56] has very recently used those methods to obtain an asymptotic formula for the number of ℓ-APs in $S(k)$, vastly generalizing earlier work of Heath-Brown [35]. The catch is that in both results, the elements of $S(k)$ are weighted *with multiplicity*: that is, since there are 4 solutions to $x^2 + y^2 = 1$, and 4 solutions to $x^2 + y^2 = 2$, the progression $\{0, 1, 2\}$ counts as appearing 16 times in $S(1)$.

Such an interpretation is fairly natural, and it does not detract from the

[24] Upper bounds for prime gaps are also very far from predictions, but they may take pride in actually *progressing*: the current best is due to Baker, Harman and Pintz [3].

beauty of Matthiesen's sharp result. However, the multiplicity $r_k(n)$ of a given $n \in S(k)$ could be as small as 2, or as large as $\exp(c_k \log n / \log\log n)$, depending on n. This makes for a rather wide variation in the number of times each arithmetic progression gets counted, so it is difficult to make any deductions about the number of *distinct* progressions this way.

The primary focus of the author's current research is to obtain sharper results for this *unweighted* problem in $S(k)$. Such precision is typically more accessible in the weighted case considered by Heath-Brown and Matthiesen.[25] Even though $r_k(n)$ fluctuates highly with n, its sum $\sum_{n=1}^{x} r_k(n)$ is understood with reasonable accuracy (at least for k fixed), certainly better than the error term in Landau's unweighted count (1.9).

Still, one might hope to obtain estimates akin to Hooley's (4.1), establishing at least the correct order of magnitude which is not currently known from existing results. A result which dates back to de la Vallée Poussin's original 1897 proof of the Prime Number Theorem [61, p. 72] is that $S(k)$ contains a specific positive fraction of all primes. The Green–Tao theorem therefore shows that the number of r-APs in $S(k)$ up to x is at least $\gg_{k,d} x^2/(\log x)^d$. It is plausible that a careful unweighting of Matthiesen's estimate might do better (by some positive power of $\log x$).

However, we expect that the powerful Green–Tao machinery can be used to improve this all the way to $\gg_{k,d} x^2/(\log x)^{d/2}$, as predicted by random heuristics (and matching the upper bounds given by sieves). In fact, it should even be possible to get lower bounds for some *higher-order* arithmetic structures, such as *pairs* of arithmetic progressions $\{m, m+d, m+2d\} \cup \{m+1, m+d+1, m+2d+1\}$: Zhou [87] has established this for primes, on assumption of the twin primes conjecture (and unconditionally for *almost-primes* having only 1 or 2 prime factors).

[25] If the reader will forgive the anthropomorphism, it is as though $n \in S(k)$ *wants* to be counted with weight $r_k(n)$. The primes also *want* to be counted with the *von Mangoldt* weight $\Lambda(n)$, but this varies very smoothly, making it a simple matter to remove.

Bibliography

[1] O. Ahmadi, N. Alon, I. F. Blake, and I. E. Shparlinski. Graphs with integral spectrum. *Linear Algebra Appl.*, 430(1):547–552, 2009.

[2] Emre Alkan, Kevin Ford, and Alexandru Zaharescu. Diophantine approximation with arithmetic functions. I. *Trans. Amer. Math. Soc.*, 361(5):2263–2275, 2009.

[3] R. C. Baker, G. Harman, and J. Pintz. The difference between consecutive primes. II. *Proc. London Math. Soc. (3)*, 83(3):532–562, 2001.

[4] Paul T. Bateman and Harold G. Diamond. *Analytic number theory.* World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2004.

[5] Paul T. Bateman and Roger A. Horn. A heuristic asymptotic formula concerning the distribution of prime numbers. *Math. Comp.*, 16:363–367, 1962.

[6] P. Bernays. *Über die Darstellung von positiven, ganzen Zahlen durch die primitiven, binären quadratischen Formen einer nicht-quadratischen Diskriminante.* PhD thesis, Georg-August-Universität, Göttingen, 1912.

[7] Manjul Bhargava. Higher composition laws. I. A new view on Gauss composition, and quadratic generalizations. *Ann. of Math. (2)*, 159(1):217–250, 2004.

[8] Valentin Blomer and Andrew Granville. Estimates for representation numbers of quadratic forms. *Duke Math. J.*, 135(2):261–302, 2006.

[9] Peter Borwein, Stephen Choi, Brendan Rooney, and Andrea Weirathmueller, editors. *The Riemann Hypothesis: A Resource for the Afficionado and Virtuoso Alike.* CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC. Springer, New York, 2008.

[10] Jean Bourgain, Van H. Vu, and Philip Matchett Wood. On the singularity probability of discrete random matrices. *J. Funct. Anal.*, 258(2):559–603, 2010.

[11] S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. In *Seventh International World-Wide Web Conference (WWW 1998)*, 1998.

[12] V. Brun. 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. *Darboux Bull. (2) 43*, 100-104:124–128, 1919.

[13] D. A. Buell and R. H. Hudson. On runs of consecutive quadratic residues and quadratic nonresidues. *BIT*, 24(2):243–247, 1984.

[14] Todd Cochrane and Robert E. Dressler. Consecutive triples of sums of two squares. *Arch. Math. (Basel)*, 49(4):301–304, 1987.

[15] Harold Davenport. Über numeri abundantes. *Sitzungsber. Preuß. Akad. Wiss., Phys.-Math. Kl.*, 1933(26-29):830–837, 1933.

[16] Alan Edelman. The probability that a random real Gaussian matrix has k real eigenvalues, related distributions, and the circular law. *J. Multivariate Anal.*, 60(2):203–232, 1997.

[17] P. Erdős and M. Kac. The Gaussian law of errors in the theory of additive number theoretic functions. *Amer. J. Math.*, 62:738–742, 1940.

[18] Paul Erdős. On the smoothness of the asymptotic distribution of additive arithmetical functions. *Amer. J. Math.*, 61:722–725, 1939.

[19] Paul Erdős and Aurel Wintner. Additive arithmetical functions and statistical independence. *Amer. J. Math.*, 61:713–721, 1939.

[20] P. Erdős and A. Schinzel. Distributions of the values of some arithmetical functions. *Acta Arith.*, 6:473–485, 1960/1961.

[21] Steven Finch, Greg Martin, and Pascal Sebah. Roots of unity and nullity modulo *n*. *Proc. Amer. Math. Soc.*, 138(8):2729–2743, 2010.

[22] Carl Friedrich Gauss. *Disquisitiones arithmeticae.* Springer-Verlag, New York, 1986. Translated and with a preface by Arthur A. Clarke, Revised by William C. Waterhouse, Cornelius Greither and A. W. Grootendorst and with a preface by Waterhouse.

[23] S. Gershgorin. Über die Abgrenzung der Eigenwerte einer Matrix. *Izv. Akad. Nauk SSSR, Otd. Mat. Estest. Nauk, VII. Ser. No.*, 6:749–754, 1931.

[24] Jean Ginibre. Statistical ensembles of complex, quaternion, and real matrices. *J. Mathematical Phys.*, 6:440–449, 1965.

[25] V. L. Girko. The circular law. *Teor. Veroyatnost. i Primenen.*, 29(4):669–679, 1984.

[26] Daniel A. Goldston, János Pintz, and Cem Y. Yıldırım. Primes in tuples. I. *Ann. of Math. (2)*, 170(2):819–862, 2009.

[27] W. T. Gowers. A new proof of Szemerédi's theorem. *Geom. Funct. Anal.*, 11(3):465–588, 2001.

[28] Andrew Granville. Harald Cramér and the distribution of prime numbers. *Scand. Actuar. J.*, 1:12–28, 1995. Harald Cramér Symposium (Stockholm, 1993).

[29] Ben Green and Terence Tao. The primes contain arbitrarily long arithmetic progressions. *Ann. of Math. (2)*, 167(2):481–547, 2008.

[30] Ben Green and Terence Tao. Linear equations in primes. *Ann. of Math. (2)*, 171(3):1753–1850, 2010.

[31] Ben Green and Terence Tao. The Möbius function is strongly orthogonal to nilsequences. *Ann. of Math. (2)*, 175(2):541–566, 2012.

[32] Ben Green and Terence Tao. The quantitative behaviour of polynomial orbits on nilmanifolds. *Ann. of Math. (2)*, 175(2):465–540, 2012.

[33] Ben Green, Terence Tao, and Tamar Ziegler. An inverse theorem for the Gowers $U^{s+1}[N]$-norm. *Ann. of Math. (2).* To appear. Preprint available at arXiv:1009.3998v2.

[34] G. H. Hardy and J. E. Littlewood. Some problems of 'Partitio numerorum'; III: On the expression of a number as a sum of primes. *Acta Math.*, 44(1):1–70, 1923.

[35] D. R. Heath-Brown. Linear relations amongst sums of two squares. In *Number theory and algebraic geometry*, volume 303 of *London Math. Soc. Lecture Note Ser.*, pages 133–176. Cambridge Univ. Press, Cambridge, 2003.

[36] Jim Hefferon. *Linear Algebra (online book).* `http://joshua.smcvt.edu/linearalgebra/`, accessed May 17, 2012.

[37] Andrew J. Hetzel, Jay S. Liew, and Kent E. Morrison. The probability that a matrix of integers is diagonalizable. *Amer. Math. Monthly*, 114(6):491–499, 2007.

[38] C. Hooley. On the intervals between numbers that are sums of two squares. II. *J. Number Theory*, 5:215–217, 1973.

[39] Christopher Hooley. On the intervals between numbers that are sums of two squares. III. *J. Reine Angew. Math.*, 267:207–218, 1974.

[40] Bernard Host and Bryna Kra. Nonconventional ergodic averages and nilmanifolds. *Ann. of Math. (2)*, 161(1):397–488, 2005.

[41] James P. Jones, Daihachiro Sato, Hideo Wada, and Douglas Wiens. Diophantine representation of the set of prime numbers. *Amer. Math. Monthly*, 83(6):449–464, 1976.

[42] J. Kahn, J. Komlós, and E. Szemerédi. On the probability that a random ±1-matrix is singular. *J. Amer. Math. Soc.*, 8(1):223–240, 1995.

[43] Y. R. Katznelson. Singular matrices and a uniform bound for congruence groups of $\mathrm{SL}_n(\mathbf{Z})$. *Duke Math. J.*, 69(1):121–136, 1993.

[44] K. Kedlaya and L. Ng. Solutions to the 67th William Lowell Putnam Mathematical Competition. `http://amc.maa.org/a-activities/a7-problems/putnam/-pdf/2006s.pdf` (online resource), accessed May 23, 2012.

[45] Mitsuo Kobayashi. *On the Density of Abundant Numbers.* PhD thesis, Dartmouth College, 2010.

[46] J. Komlós. On the determinant of random matrices. *Studia Sci. Math. Hungar.*, 3:387–399, 1968.

[47] Hans-Joachim Kowalsky. Ganzzahlige Matrizen mit ganzzahligen Eigenwerten. *Abh. Braunschweig. Wiss. Ges.*, 34:15–32, 1982.

[48] E. Landau. Über die Einteilung der positiven ganzen Zahlen in vier Klassen nach der Mindestzahl der zu ihrer additiven Zusammensetzung erforderlichen Quadrate. *Arch. der Math. u. Phys. (3)*, 13:305–312, 1908.

[49] Leonid A. Levin. Average case complete problems. *SIAM J. Comput.*, 15(1):285–286, 1986.

[50] Florian Luca. Perfect Fibonacci and Lucas numbers. *Rend. Circ. Mat. Palermo (2)*, 49(2):313–318, 2000.

[51] Florian Luca, V. Janitzio Mejía Huguet, and Florin Nicolae. On the Euler function of Fibonacci numbers. *J. Integer Seq.*, 12(6):Article 09.6.6, 15pp., 2009.

[52] Helmut Maier. Primes in short intervals. *Michigan Math. J.*, 32(2):221–225, 1985.

[53] Greg Martin. The smallest solution of $\phi(30n+1) < \phi(30n)$ is .... *Amer. Math. Monthly*, 106(5):449–451, 1999.

[54] Greg Martin and Erick Wong. The number of 2×2 integer matrices having a prescribed integer eigenvalue. *Algebra Number Theory*, 2(8):979–1000, 2008.

[55] Greg Martin and Erick B. Wong. Almost all integer matrices have no integer eigenvalues. *Amer. Math. Monthly*, 116(7):588–597, 2009.

[56] L. Matthiesen. Linear correlations amongst numbers represented by positive definite binary quadratic forms. *Acta Arith.* To appear. Preprint available at arXiv:1106.4690.

[57] Gérard Maze. Natural density distribution of Hermite normal forms of integer matrices. *J. Number Theory*, 131(12):2398–2408, 2011.

[58] H. L. Montgomery. The pair correlation of zeros of the zeta function. In *Analytic number theory (Proc. Sympos. Pure Math., Vol. XXIV, St. Louis Univ., St. Louis, Mo., 1972)*, pages 181–193. Amer. Math. Soc., Providence, R.I., 1973.

[59] Hugh L. Montgomery and Robert C. Vaughan. *Multiplicative number theory. I. Classical theory*, volume 97 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2007.

[60] T. Muir. *A treatise on the theory of determinants.* Revised and enlarged by William H. Metzler. Dover Publications Inc., New York, 1960.

[61] Władysław Narkiewicz. *The development of prime number theory.* Springer Monographs in Mathematics. Springer-Verlag, Berlin, 2000. From Euclid to Hardy and Littlewood.

[62] D. J. Newman. Euler's ϕ function on arithmetic progressions. *Amer. Math. Monthly*, 104(3):256–257, 1997.

[63] Thomas R. Nicely. Enumeration to 10^{14} of the twin primes and Brun's constant. *Virginia J. Sci.*, 46(3):195–204, 1995.

[64] Ivan Niven. *Numbers: rational and irrational*, volume 1 of *New Mathematical Library*. Random House, New York, 1961.

[65] Ivan Niven, Herbert S. Zuckerman, and Hugh L. Montgomery. *An introduction to the theory of numbers*. John Wiley & Sons Inc., New York, fifth edition, 1991.

[66] Gábor Pataki, Mustafa Tural, and Erick B. Wong. Basis reduction and the complexity of branch-and-bound. In *Proceedings of the Twenty-First Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1254–1261, Philadelphia, PA, 2010. SIAM.

[67] Paul Pollack. Perfect numbers with identical digits. *Integers*, 11A:A18, 11pp., 2011.

[68] K. F. Roth. On certain sets of integers. *J. London Math. Soc.*, 28:104–109, 1953.

[69] Mark Rudelson and Roman Vershynin. The Littlewood-Offord problem and invertibility of random matrices. *Adv. Math.*, 218(2):600–633, 2008.

[70] Tom Sanders. On Roth's theorem on progressions. *Ann. of Math. (2)*, 174(1):619–636, 2011.

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

[72] A. Schinzel and Y. Wang. A note on some properties of the functions $\varphi(n)$, $\sigma(n)$ and $\theta(n)$. *Ann. Polon. Math.*, 4:201–213, 1958.

[73] I. Schoenberg. Über die asymptotische Verteilung reeller Zahlen mod 1. *M. Z.*, 28:171–199, 1928.

[74] I. J. Schoenberg. On asymptotic distributions of arithmetical functions. *Trans. Amer. Math. Soc.*, 39(2):315–330, 1936.

[75] Robert Forsyth Scott. *The theory of determinants and their applications*. Revised by G. B. Mathews. Cambridge University Press, Cambridge, 1904.

[76] J.-P. Serre. *A course in arithmetic*. Springer-Verlag, New York, 1973. Translated from the French, Graduate Texts in Mathematics, No. 7.

[77] Pin-Tsung Shao. On the distribution of the values of a class of arithmetical functions. *Bull. Acad. Pol. Sci., Cl. III*, 4:569–572, 1956.

[78] I. E. Shparlinski. Some counting questions for matrices with restricted entries. *Linear Algebra Appl.*, 432(1):155–160, 2010.

[79] Claudia A. Spiro and Douglas A. Hensley. Problems and Solutions: Solutions of Advanced Problems: 6539. *Amer. Math. Monthly*, 97(3):253–254, 1990.

[80] E. Szemerédi. On sets of integers containing no k elements in arithmetic progression. *Acta Arith.*, 27:199–245, 1975. Collection of articles in memory of Juriĭ Vladimirovič Linnik.

[81] Terence Tao and Van Vu. On random ±1 matrices: singularity and determinant. *Random Structures Algorithms*, 28(1):1–23, 2006.

[82] Terence Tao and Van Vu. Random matrices: the circular law. *Commun. Contemp. Math.*, 10(2):261–307, 2008.

[83] Terence Tao and Van Vu. Random matrices: universality of ESDs and the circular law. *Ann. Probab.*, 38(5):2023–2065, 2010. With an appendix by Manjunath Krishnapur.

[84] J. G. van der Corput. Über Summen von Primzahlen und Primzahlquadraten. *Math. Ann.*, 116(1):1–50, 1939.

[85] I.M. Vinogradov. Some theorems concerning the theory of primes. *Rec. Math. Moscou, n. Ser.*, 2:179–195, 1937.

[86] Erick B. Wong. Simultaneous approximation of reals by values of arithmetic functions. In *Anatomy of Integers*, volume 46 of *CRM Proc. Lecture Notes*, pages 289–297. Amer. Math. Soc., Providence, RI, 2008.

[87] Binbin Zhou. The Chen primes contain arbitrarily long arithmetic progressions. *Acta Arith.*, 138(4):301–315, 2009.