ANNALES
DE LA FACULTÉ
DES SCIENCES
DE
TOULOUSE
Mathématiques

ALEX KONTOROVICH

Levels of Distribution and the Affine Sieve

Tome XXIII, n° 5 (2014), p. 933-966.

<http://afst.cedram.org/item?id=AFST_2014_6_23_5_933_0>

Annales de la Faculté des Sciences de Toulouse Vol. XXIII, n° 5, 2014
pp. 933-966

Levels of Distribution and the Affine Sieve

Alex Kontorovich[1]

Abstract. — We discuss the notion of a "Level of Distribution" in two settings. The first deals with primes in progressions, and the role this plays in Yitang Zhang's theorem on bounded gaps between primes. The second concerns the Affine Sieve and its applications.

Résumé. — Nous discutons de la notion de "Niveau de Distribution" dans deux contextes. Le premier concerne les nombres premiers en progression, et le rôle qu'elle joue dans le théorème de Yitang Zhang sur les écarts bornés entre nombres premiers. Le second concerne le Crible Affine et ses applications.

Contents

[1] Math Department, Yale University, New Haven, CT 06511 USA, and School of Mathematics, Institute for Advanced Study, Princeton, NJ 08540 USA
alex.kontorovich@yale.edu

The author gratefully acknowledges support from an NSF CAREER grant DMS-1254788, an Alfred P. Sloan Research Fellowship, a Yale Junior Faculty Fellowship, and support at IAS from The Fund for Math and The Simonyi Fund.

1. Introduction

This article is an expanded version of the author's lecture in the Basic Notions Seminar at Harvard, September 2013. Our goal is a brief and introductory exposition of aspects of two topics in sieve theory which have received attention recently: (1) the spectacular work of Yitang Zhang, under the title "Level of Distribution," and (2) the so-called "Affine Sieve," introduced by Bourgain-Gamburd-Sarnak.

2. Level of Distribution for the Primes

Let p_n be the nth prime number. We begin with the infamous

Twin Prime Conjecture:

$$\liminf_{n\to\infty}(p_{n+1}-p_n) \;=\; 2.$$

A slight weakening of this problem is called the

Bounded Gaps Conjecture:

$$\liminf_{n\to\infty}(p_{n+1}-p_n) \;<\; \infty.$$

A tremendous shock ran through the mathematical community in April 2013 when Yitang Zhang [104] proved

Zhang's Theorem (2013). — *The Bounded Gaps Conjecture is true. In particular,*

$$\liminf_{n\to\infty}(p_{n+1}-p_n) \;<\; 7\times 10^7.$$

Our goal in this section is to explain what is meant by a "level of distribution" for the primes, and give some hints of the role it plays in the proof of Zhang's theorem.

2.1. The Distribution of Primes

First we recall the Prime Number Theorem (PNT), proved independently and simultaneously by Hadamard [49] and de la Vallée Poussin [24] in 1896, following the strategy introduced in Riemann's 1859 epoch-making memoir [85]. It is often stated as:

$$\pi(x) \ := \sum_{p<x} 1 \quad \sim \quad \frac{x}{\log x}, \qquad\qquad x \to \infty,$$

where, as throughout, log is to base e, and p denotes a prime. The first Basic Notion is that this is the "wrong" formula, not in the sense of being untrue, but in the sense that

$$\pi(x) \ = \ \frac{x}{\log x} + \Omega\left(\frac{x}{\log^2 x}\right), \tag{2.1}$$

the error term being unnecessarily large. (Here Ω is the negation of little-oh.) A more precise statement of PNT, not far from the best currently known, is the following.

Prime Number Theorem. — *For any $A > 1$,*

$$\pi(x) \ = \ \mathrm{Li}(x) + O_A\left(\frac{x}{\log^A x}\right), \qquad\qquad \text{as } x \to \infty. \tag{2.2}$$

Here the subscript A in the big-Oh means that the implied constant depends on A, and Li is the "logarithmic integral" function

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

By an exercise in partial integration, we have that

$$\mathrm{Li}(x) \ = \ \frac{x}{\log x} + \frac{x}{\log^2 x} + O\left(\frac{x}{\log^3 x}\right),$$

which together with (2.2) implies (2.1). On the other hand, the Riemann Hypothesis (RH) predicts that

$$\pi(x) \ = \ \mathrm{Li}(x) + O(\sqrt{x}\log x),$$

epitomizing the "square-root cancellation" phenomenon. If true, this estimate would be best possible (up to log factors), as Littlewood proved in 1914 that

$$\pi(x) \ = \ \mathrm{Li}(x) + \Omega\left(\sqrt{x}\,\frac{\log\log\log x}{\log x}\right).$$

In fact, he showed that the difference $\pi(x) - \mathrm{Li}(x)$ infinitely-often attains both positive and negative values of this order of magnitude.

2.2. Primes in Progressions

The next most basic question is: How are the primes distributed in arithmetic progressions? Given an integer $q \geqslant 1$, often called the "level" in this context, and a coprime number $(a,q)=1$, let

$$\pi(x;a,q) \ := \sum_{\substack{p<x \\ p\equiv a(\mathrm{mod}\, q)}} 1$$

denote the number of primes up to x in the progression $a(\mathrm{mod}q)$. A relatively minor modification to the proof of (2.2) gives

PNT in Progressions. — *For any $A>1$,*

$$\pi(x;a,q) \ = \ \frac{\mathrm{Li}(x)}{\phi(q)} + O_{A,q}\left(\frac{x}{\log^A x}\right), \qquad x\to\infty. \tag{2.3}$$

Meanwhile, the Generalized Riemann Hypothesis (GRH) predicts

$$\pi(x;a,q) \ = \ \frac{\mathrm{Li}(x)}{\phi(q)} + O_{\varepsilon}\left(x^{1/2+\varepsilon}\right), \tag{2.4}$$

for any $\varepsilon>0$. These estimates confirm our intuition that primes should not favor one primitive (meaning a and q are coprime) arithmetic progression mod$q$ over others, there being $\phi(q)=|(\mathbb{Z}/q\mathbb{Z})^\times|$ of them total.

In applications, it is often important to be able to use formulae like (2.3) while allowing q to vary with x. For example, it does not seem unreasonable that we should be able to use (2.3) to estimate, say, the number of primes up to e^{100} which are 1 mod 100^3, or primes up to e^{1000} which are 1 mod 1000^3, or more generally, primes up to $x=e^{\ell}$ which are $1\mathrm{mod}\ell^3$. In these examples, we have taken a level of size $q=\ell^3=\log^3 x$ which, it turns out, is growing too rapidly relative to x to obtain a meaningful asymptotic from present methods; the error terms in all of these questions might swamp the main terms, giving no estimate at all.

To address this issue of uniformity in the level q, there is a famous estimate proved by Walfisz [103] in 1936 by adapting work of Siegel [99].

Siegel-Walfisz Theorem. — *Given any positive constants A and B, any $q<\log^B x$, and any $(a,q)=1$, we have*

$$\pi(x;a,q) \ = \ \frac{\mathrm{Li}(x)}{\phi(q)} + O_{A,B}\left(\frac{x}{\log^A x}\right). \tag{2.5}$$

It may appear that the uniformity issue in the range $q < \log^B x$ has been completely resolved, but there's a catch: the implied constant in (2.5) coming from the proof is "ineffective." This means that, once the parameters A and B are supplied, there is no known procedure to determine the constant. Thus we still have no way of verifying that $\mathrm{Li}(e^\ell)/\phi(\ell^3)$ is an accurate estimate for $\pi(e^\ell; 1, \ell^3)$. This so-called "Siegel zero" phenomenon is the sense in which we do not know the PNT in progressions.

The danger of an ineffective constant is beautifully illustrated by Iwaniec's (facetious)

THEOREM. — *There exists a constant $C > 0$ such that, if RH holds up to height C (meaning $\zeta(\sigma + it) \neq 0$ for all $\frac{1}{2} < \sigma < 1$, $|t| < C$), then RH holds everywhere.*

This fantastic result seems to reduce RH to a finite computation; before we get too excited, let's have a look at the

Proof. — There are two cases.

Case 1: Assume RH is true. Set $C = 1$, and RH holds.

Case 2: Assume RH is false, that is, $\zeta(\sigma + it) = 0$ for some $\frac{1}{2} < \sigma < 1$ and some $t > 0$. Set $C = t + 1$. The statement is vacuously true. □

As an aside, we briefly recall that a similar phenomenon occurs in the study of Gauss's Class Number Problem. Let $-d < 0$ be the discriminant of an imaginary quadratic field and let $h(-d)$ be the corresponding class number (see wikipedia for definitions, which will not be needed for our discussion). In 1935, Siegel and Landau [66] (based on work of Hecke [65], Deuring [23], Mordell [78], and Heilbronn [51]) independently proved that

$$h(-d) \gg_\varepsilon d^{1/2-\varepsilon}, \tag{2.6}$$

for any $\varepsilon > 0$. Again this implied constant is ineffective, and thus does not allow one to, e.g., tabulate all d with $h(-d) = 1$ (the Class Number One Problem). Much later, Goldfeld [39, 40] (1976) together with Gross-Zagier [48] (1985) managed to circumvent the ineffectivity, proving

$$h(-d) > \frac{1}{55} \log d, \tag{2.7}$$

whenever d is prime (we make this restriction only to give the simplest formula). Thanks to (2.7) (and much other work), we now have complete tables of all d with class number up to 100. The point of this aside is that, just because one proof gives an ineffective constant, there might be a

completely different proof for which the constants are absolute. Resolving this "Siegel zero" issue is one of the main outstanding problems in analytic number theory.

2.3. Primes in Progressions on Average

In many applications, what is needed is not uniformity for a single level q, but over a range of q. This is the heart of what is meant by a "level of distribution," as explained below.

Assuming GRH, we see from (2.4) that

$$\sum_{q<Q} \max_{(a,q)=1} \left| \pi(x;a,q) - \frac{\mathrm{Li}(x)}{\phi(q)} \right| \ll_\varepsilon \sum_{q<Q} x^{1/2+\varepsilon} < Qx^{1/2+\varepsilon}. \tag{2.8}$$

So if we take $Q = x^{1/2-2\varepsilon}$, say, then the error terms add up to at most $x^{1-\varepsilon}$, while there are about $x/\log x$ primes up to x. That is, all of these errors summed together still do not exceed the total number of primes. This immediately leads us to the

Definition: Level of Distribution (for primes in progressions). — We will say that the primes have a *level of distribution* Q if, for all $A < \infty$,

$$\sum_{q<Q} \max_{(a,q)=1} \left| \pi(x;a,q) - \frac{\mathrm{Li}(x)}{\phi(q)} \right| = O_A\left(\frac{x}{\log^A x} \right). \tag{2.9}$$

When Q can be taken as large as $x^{\vartheta-\varepsilon}$ for some $\vartheta > 0$, we call ϑ an *exponent of distribution* for the primes.

Note that level of distribution is not a quantity inherent to the sequence of primes, but is instead a function of what one can *prove* about the primes. While GRH implies the level $Q = x^{1/2-\varepsilon}$ (or exponent $\vartheta = 1/2$), the unconditional Siegel-Walfisz estimate (2.5) gives only a level of size $Q = \log^A x$, which is not even a positive exponent ϑ.

It was a dramatic breakthrough when Bombieri [16] and A. I. Vinogradov [102] (based on earlier work of Linnik [67], Renyi [83], Roth [86], and Barban [1]) independently and simultaneously proved the

Bombieri-Vinogradov Theorem (1965). — *The primes have exponent of distribution $\vartheta = 1/2$. More precisely, for any constant $A > 1$, there exists a constant $B > 1$ so that*

$$\sum_{q<\frac{x^{1/2}}{\log^B x}} \max_{(a,q)=1} \left| \pi(x;a,q) - \frac{\mathrm{Li}(x)}{\phi(q)} \right| \ll_A \frac{x}{\log^A x}.$$

The Bombieri-Vinogradov theorem (B-V) is thus an unconditional substitute for GRH on average, since both produce the same exponent of distribution $\vartheta = 1/2$! (The implied constant is still ineffective, as the proof uses Siegel-Walfisz; we have not escaped the "Siegel zero" problem.)

Being even more ambitious, one may ask for variation in the size of the error term; after all, we crudely imported the worst possible error from (2.4) into (2.8). Applying a "square-root cancellation" philosophy yet again, one might boldly posit that the term $Qx^{1/2}$ on the right side of (2.8) can be replaced by $Q^{1/2}x^{1/2}$, in which case Q can be taken as large as $x^{1-\varepsilon}$. This is the

Elliott-Halberstam Conjecture [25] (1968). — *The primes have exponent of distribution $\vartheta = 1$. That is, for any $\varepsilon > 0$ and $A < \infty$,*

$$\sum_{q<x^{1-\varepsilon}} \max_{(a,q)=1} \left| \pi(x;a,q) - \frac{\mathrm{Li}(x)}{\phi(q)} \right| \;=\; O_{A,\varepsilon}\left(\frac{x}{\log^A x} \right). \tag{2.10}$$

The Elliott-Halberstam Conjecture (E-H), if true, goes far beyond any RH-type statement, as far as we are aware. As long as we are already dreaming, we may as well suppose that this further square-root cancellation happens not only on average, as E-H claims, but individually; by this we mean the following. Returning to (2.4), the "main" term is very roughly of size x/q, so might not the error be of square-root the main term, not just square-root of x? This is

Montgomery's Conjecture [77] (1971). — For all $\varepsilon > 0$,

$$\pi(x;a,q) \;=\; \frac{\mathrm{Li}(x)}{\phi(q)} + O_\varepsilon\left(\frac{x^{1/2+\varepsilon}}{q^{1/2}} \right).$$

Montgomery's Conjecture immediately implies E-H, but we emphasize again that both of these assertions are not, as far as we know, consequences of any RH-type statement.

Nothing beyond B-V has ever been proved towards the pure level of distribution defined in (2.9). But if one drops the absolute values, fixes one non-zero integer a, and weights the errors at level q by a function $\lambda(q)$ which is "well-factorable" (the precise meaning of which we shall not give here), then one can go a bit into the E-H range. Building on work by Fouvry-Iwaniec [28, 35], we have the

Bombieri-Friedlander-Iwaniec Theorem [5] (1986). — *Fix any $a \neq 0$ and let $\lambda(q)$ be a "well-factorable" function. Then for any $A > 1$ and $\varepsilon > 0$,*

$$\sum_{q<x^{4/7-\varepsilon}} \lambda(q) \left(\pi(x;a,q) - \frac{\mathrm{Li}(x)}{\phi(q)} \right) \ll_{a,A,\varepsilon} \frac{x}{\log^A x}. \tag{2.11}$$

Thus in the weighted sense above, the Bombieri-Friedlander-Iwaniec Theorem (BFI) gives a weighted exponent of distribution

$$\vartheta \ = \ 4/7 \quad > \quad 1/2,$$

giving some partial evidence towards E-H. But before we get too optimistic about the full E-H, let us point out just how delicate the conjecture is. Building on work of Maier [72], Friedlander and Granville [27] showed that the level $x^{1-\varepsilon}$ in (2.10) cannot be replaced by $x(\log x)^{-A}$. More precisely, we have the following

Friedlander-Granville Theorem (1989). — *For any $A > 0$, there exist arbitrarily large values of a and x for which*

$$\sum_{\substack{q<x(\log x)^{-A} \\ (q,a)=1}} \left| \pi(x;a,q) - \frac{\mathrm{Li}(x)}{\phi(q)} \right| \gg_A \frac{x}{\log x}.$$

In particular, the asymptotic formula $\pi(x;a,q) \sim \mathrm{Li}(x)/\phi(q)$ can be false for q as large as $x/\log^A x$.

2.4. Small Gaps Between Primes

Let us return to the Bounded Gaps Conjecture (now Zhang's Theorem). Recall that p_n is the nth prime. Before studying absolute gaps, what can we say about gaps relative to the average? PNT tells us that $p_n \sim n \log n$, so the average gap $p_{n+1} - p_n$ is of size about $\log p_n$. Hence

$$\Delta \ := \ \liminf_{n\to\infty} \frac{p_{n+1} - p_n}{\log p_n} \ \leqslant \ 1. \tag{2.12}$$

Here is a very abbreviated history on reducing the number on the right side of (2.12).

Hardy-Littlewood [84] (1926):	$\Delta \leqslant 2/3$,	assuming GRH.
Erdös [26] (1940):	$\Delta < 1$,	unconditional; by sieving.
Bombieri-Davenport [4] (1966):	$\Delta \leqslant 1/2$,	unconditional; by refining Hardy-Littlewood and replacing GRH by B-V.
Maier [56, 73] (1988):	$\Delta < 1/4$,	unconditional; using a radically different method
Goldston [41] (1992):	$\Delta = 0$,	assuming E-H and another E-H type conjecture.
Goldston-Pintz-Yıldırım [43] (2005):	$\Delta = 0$,	unconditional.

In fact, Goldston-Pintz-Yıldırım (GPY) were able to push their method even further to show unconditionally [44] that

$$\liminf_{n\to\infty} \frac{p_{n+1} - p_n}{(\log p_n)^{1/2}(\log\log p_n)^2} \ < \ \infty.$$

So consecutive primes infinitely often differ by about square-root of the average gap. Moreover, assuming the primes have *any* level of distribution $\vartheta > 1/2$, that is, any level in the E-H range (going beyond B-V), the GPY method gives a conditional proof of the Bounded Gaps Conjecture.

The GPY method has been explained in great detail in a number of beautiful expositions (e.g. [100, 42]) so we will not repeat the discussion here, contenting ourselves with just a few words on Zhang's advances. Once GPY was understood by the community, the big open question, in light of BFI, was whether the "weights" $\lambda(q)$ from (2.11) could somehow be incorporated into the GPY method, so that in the resulting error analysis, B-V could be replaced by BFI. There was even a meeting at the American Institute of Mathematics in November 2005, at which one working group was devoted to exactly this problem. At the time, at least to some, it did not seem promising.

Yitang Zhang's accomplishment, then, was threefold. He first changed the GPY weighting functions in a clever way (in fact a similar change had been observed independently by Motohashi-Pintz [79] and others), then he proved an analogue of the GPY sieving theorem with his new weights (as Motohashi-Pintz had also done), and finally (and most spectacularly!), he proved a more flexible[1] analogue of BFI which incorporates his new weights. In this technical tour-de-force, he was able to break the $\vartheta = 1/2$ barrier in the weighted level of distribution of the primes, and complete the program initiated by Goldston in [41].

(1) The most important aspect of Zhang's version is that the shift variable a is allowed to vary, as opposed to (2.11) where it must be fixed.

Here is one final Basic Notion on this topic: Zhang's Theorem, at least as it currently stands, is ineffective! What he actually proves is a twin-prime analogue of Bertrand's Postulate (that for any $x > 1$, there is a prime between x and $2x$).

Zhang's Theorem, Again. — *For every x sufficiently large, there is a pair of primes with difference at most 7×10^7 in the range $[x, 2x]$.*

How large is sufficiently large? It depends on whether or not GRH is true! Like most others, Zhang too relies at some early stage on the ineffective Siegel-Walfisz Theorem, and for this reason cannot escape Siegel zeros. (On the other hand, Heath-Brown [50] has famously shown that if GRH fails and there *is* a particularly "bad" sequence of Siegel zeros, then the Twin Prime Conjecture would follow!)

For further reading, we recommend any number of excellent texts, e.g. [22, 57, 31], and of course the original papers.

Added in proof. — It is very fortunate for the author that he chose to focus this survey on the "level of distribution" aspect of Zhang's work. In November 2013, James Maynard [74] (and independently, Terry Tao), developing an earlier attempted version of a method by Goldston-Yıldırım, succeeded in proving the even more shocking result:

Maynard's Theorem. — *For any $\ell \geqslant 1$,*

$$\liminf_{n\to\infty}(p_{n+\ell} - p_n) < \infty.$$

That is, one can find not only prime pairs which differ by a bounded amount, but also prime triples, quadruples, etc. Most remarkably, Maynard only needs the primes to have *any* exponent of distribution $\vartheta > 0$ for his method to work (so now not even B-V is needed)! Nevertheless, Zhang's spectacular achievement in going beyond the Riemann hypothesis in giving a flexible (weighted) exponent of distribution beyond $\theta = 1/2$ will stand the test of time, and will surely find other applications. For a beautiful exposition of this aspect of Zhang's work, see the recent arXiv posting by Friedlander-Iwaniec [32].

3. The Affine Sieve

The goal of the Affine Sieve, initiated by Bourgain-Gamburd-Sarnak and completed by Salehi Golsefidy-Sarnak, is to extend to the greatest generality possible the mechanism of the Brun sieve. We will first discuss the latter

in §3.1 before turning our attention to the former. In §3.2 we motivate the general theory with an elementary problem, before presenting (some aspects of) the general theory in §3.3. This will again not be a rigorous or comprehensive survey (for which we refer the reader to any number of expositions, e.g. [91, 92, 46, 64, 96], in addition to the original papers [7, 8, 97]), but rather (we hope) a gentle introduction for the beginner. We apply the general theory to a few more illustrative examples in §§3.4–3.5, where we also give a discussion of Thin Orbits, see in particular §3.5.

While the general theory is in principle "complete," in that the Brun sieve can now be executed on matrix orbits, the whole program is far from finished, if one wishes to produce actual primes or almost-primes with very few factors in specific settings. We wish to highlight here some instances in which one can go beyond the capabilities of the general theory. In certain special settings, one can now produce actual primes in Affine Sieve-type problems by applying a variety of methods, each completely different from the general framework; we review some of these in §3.6. Finally, we discuss in §3.7 other special settings in which, though primes cannot yet be produced, novel techniques have nevertheless given improved levels of distribution, in the end coming quite close to producing primes. We hope these give the reader some sense of the present landscape.

3.1. The Brun Sieve

As throughout, we give only the most basic ideas. The first sieving procedure for producing tables of primes is credited to the ancient Eratosthenes (∼200 BCE), whose method exposes a simple but important observation: if $n < x$ and n has no prime factors below $\sqrt{x}$, then n is prime. Thus to make a table of the primes up to 100, one needs only to strike out (sieve) numbers divisible by 2, 3, 5, and 7 (the primes below $10 = \sqrt{100}$). A very slight generalization of the above is that: if $n < x$ has no prime factors below $x^{1/(R+1)}$, then n is a product of at most R primes. We call such a number R-almost-prime, and let $\mathcal{P}_R$ denote the set of R-almost-primes.

As a warmup, let us try (and fail) to prove the PNT by sieving. To count the primes up to x, we first take the integers up to x (there are x of them), throw out those divisible by 2 (there are *roughly* $x/2$ of them), then 3 (there are roughly $x/3$ of them), and so on for all primes up to $\sqrt{x}$. But then we have twice thrown out multiplies of $2 \times 3 = 6$, so should add them back in (there are roughly $x/6$ of them), and so on goes the familiar inclusion-exclusion principle:

$$\pi(x) \stackrel{?}{\approx} x - \frac{x}{2} - \frac{x}{3} - \cdots + \frac{x}{2\times 3} + \frac{x}{2\times 5} + \cdots - \frac{x}{2\times 3\times 5} - \cdots$$

The problem with this approach is two-fold. First of all, the word "roughly" above is very dangerous; it hides the error $r_q = r_q(x)$ in

$$\#\{n < x : n \equiv 0(\mathrm{mod} q)\} \;=\; \frac{x}{q} + r_q, \qquad\qquad |r_q| < 1. \tag{3.1}$$

These remainder terms r_q, when added together with absolute values in the inclusion-exclusion procedure, very quickly swamp the main term. Perhaps we are simply too crude and a better estimation of these errors can make the above rigorous? Alas, were this the case, an elementary analysis (see, e.g., [45]) will predict that

$$\pi(x) \;\sim\; 2e^{-\gamma}\frac{x}{\log x},$$

where γ is the Euler-Mascheroni constant,

$$\gamma \;:=\; \lim_{n\to\infty}\left(1+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\cdots+\frac{1}{n}-\log n\right) \;\approx\; 0.577.$$

Since $2e^{-\gamma} \approx 1.12$, we would be off by a constant from the truth. So these error terms *must* be at least of the same order as the main term. This simple sieving procedure cannot by itself prove the PNT.

It was a technical tour-de-force when Brun [**?**] managed to push arguments of the above flavor, together with a heavy dose of combinatorics and other estimates, to prove a different type of approximation to the Twin Prime Conjecture, in some sense orthogonal to Zhang's Theorem.

Brun's Theorem (1919). — *There are infinitely many integers n so that both n and $n+2$ have at most nine prime factors.*

That is, infinitely often n and $n+2$ are simultaneously in $\mathcal{P}_R$ with $R = 9$. After much work by many people, the sieve was finally pushed to its limit[2] in [20]:

Chen's Theorem (1973). — *There are infinitely many primes p so that $p+2$ is either prime or the product of two primes.*

It was realized long ago that this sieving procedure applies to much more general problems. Suppose we have an infinite set of natural numbers

$$\mathcal{S} \subset \mathbb{N} \tag{3.2}$$

and wish to prove the existence and abundance of primes or R-almost-primes in $\mathcal{S}$. Roughly speaking, all that is needed is an appropriate analogue

(2) See the discussion of the "parity problem" in, e.g., [29].

of (3.1). In particular, suppose that $\mathcal{S}$ is fairly well distributed on average among multiples of q, in the sense that[3]

$$\#\{n \in \mathcal{S} \cap [1,x] : n \equiv 0(q)\} \;=\; \frac{1}{q}\#\{\mathcal{S} \cap [1,x]\} + r_q, \tag{3.3}$$

(or perhaps with $1/q$ in (3.3) replaced by some analytically similar function like $1/\phi(q)$), where the errors are controlled by

$$\sum_{q<Q} |r_q| \;=\; o\left(\#\{\mathcal{S} \cap [1,x]\}\right), \tag{3.4}$$

for some Q. Such an expression should look familiar; it is in some sense the generalization of (2.9), and Q is likewise called a *level of distribution* for $\mathcal{S}$. Then the sieve technology (again very roughly) tells us that if Q can be taken as large as a power of x, say

$$Q \;=\; x^{\vartheta-\varepsilon} \tag{3.5}$$

for some *exponent of distribution* $\vartheta > 0$, then $\mathcal{S}$ contains R-almost-primes, with

$$R \;=\; \left\lceil \frac{1}{\vartheta} + \varepsilon \right\rceil. \tag{3.6}$$

For example, if $\mathcal{S}$ is the set of shifted primes, $\mathcal{S} = \{p+2 : p \text{ prime}\}$, then the Bombieri-Vinogradov Theorem gives us an exponent of distribution $\vartheta = 1/2$, which gives R-almost-primes in $\mathcal{S}$ with $R = \lceil 2+\varepsilon \rceil = 3$. To obtain [thm:Chen]Chen's Theorem is much much harder.

3.2. Affine Sieve Warmup: Pythagorean Areas

Arguably the oldest "Affine Sieve" problem is the following. Let (x,y,z) be a Pythagorean triple, that is, an integer solution to the equation $x^2+y^2 = z^2$. What can one say about the number of prime factors of the area $\frac{1}{2}xy$ of a Pythagorean triple?

It was known to the ancients that Pythagorean triples $\mathbf{x} = (x,y,z)$ with coprime entries and x odd are parametrized by coprime pairs (c,d) of opposite parity with

$$x = c^2 - d^2, \qquad y = 2cd, \qquad z = c^2 + d^2. \tag{3.7}$$

---

(3) There are various ways the assumption (3.3) can (and should) be relaxed and generalized further, but for the purposes of our discussion, we will ignore all technicalities and stick with this simple-minded version.

In fact, it is easy to see that the area is always divisible by 6, so we can further remove unwanted prime factors by studying the function $f(\mathbf{x}) = \frac{1}{12}xy$. Observe that in the parametrization (3.7), we have

$$f(\mathbf{x}) = \frac{1}{12}xy = \frac{1}{6}cd(c+d)(c-d). \tag{3.8}$$

When does $f(\mathbf{x})$ have few prime factors? That is, for which R and triples $\mathbf{x}$ is $f(\mathbf{x}) \in \mathcal{P}_R$?

One can easily check that there are only finitely many pairs (c,d) so that (3.8) is the product of two primes. The largest such pair is $(c,d) = (7,6)$, which corresponds to the triple $\mathbf{x} = (13, 84, 85)$ of one-sixth area $f(\mathbf{x}) = \frac{1}{12}13 \times 84 = 91 = 7 \times 13$.

Allowing $R = 3$ primes, we could set, say, $d = 2$; then from (3.8), we are asking for many c's so that $\frac{1}{3}c(c-2)(c+2)$ is the product of three primes (a type of "triplet prime" problem). As the reader may surmise, it is expected that infinitely many such c's exist, but this seems far outside the range of what can be proved today. Nevertheless, it should be clear that for $f(\mathbf{x})$ to have three prime factors, $\mathbf{x}$ must be of some "special" form, so either c or d (or their sum or difference) must be "small"; see Figure 1.

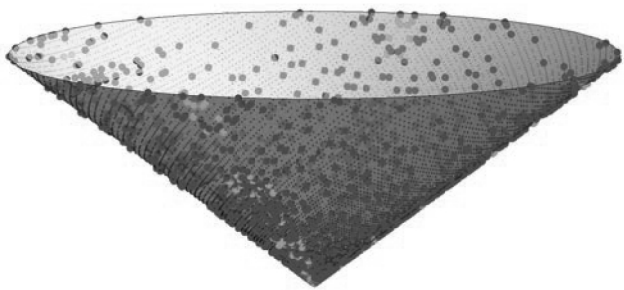

Figure 1. — A piece of the cone V in (3.9) with markings at the primitive Pythagorean triples $\mathbf{x}$. Points $\mathbf{x}$ are marked according to whether the "area" $f(\mathbf{x}) = \frac{1}{12}xy$ is in $\mathcal{P}_R$ with $R \leqslant 3$ (●), $R = 4$ (●), or $R \geqslant 5$ (●).

A better way of saying this is to use the Zariski topology. The ambient variety on which all Pythagorean triples live is the cone V given by

$$V \ : \ F(\mathbf{x}) = 0, \tag{3.9}$$

where F is the quadratic form

$$F(\mathbf{x}) \ = \ x^2 + y^2 - z^2. \tag{3.10}$$

Let $\mathcal{X}_R$ denote the set of integer Pythagorean triples $\mathbf{x}$ with $f(\mathbf{x}) \in \mathcal{P}_R$. Then a restatement of the "smallness" of points in $\mathcal{X}_3$ is that is a proper subvariety of V. That is, points in $\mathcal{X}_3$ have extra algebraic relations.

This "smallness" somehow fundamentally changes the nature of the problem; e.g. setting $d = 2$ as above, one is asking for a "triplet prime" type statement, rather than the original area problem. It seems natural, then, to exclude such small solutions. That is, we shall insist on finding an R so that the set $\mathcal{X}_R$ is Zariski dense in V; this means that any polynomial which vanishes on all of $\mathcal{X}_R$ must also vanish on V.

If we now allow $R = 4$ prime factors, then we see in Figure 1 that such points seem to spread out all over the cone. In fact, it was observed in [8] that Green-Tao's revolutionary work [47] on linear equations in primes rigorously establishes the Zariski density of $\mathcal{X}_4$ in V. This is because $f(\mathbf{x})$ in (3.8) is the product of four linear factors in two variables, which in the Green-Tao nomenclature is a system of "finite complexity" (we refer the reader to their paper for the definition, which is not needed here). Thus the problem of Pythagorean areas, at least if one insists on Zariski density, is completely solved.

3.2.1. *Reformulation*

What does this simple problem have to do with orbits? Let

$$G = \mathrm{SO}_F(\mathbb{R}) = \mathrm{SO}_{2,1}(\mathbb{R})$$

be the real special orthogonal group preserving the quadratic form F in (3.10); that is,

$$G \;=\; \{g \in \mathrm{SL}_3(\mathbb{R}) : F(g \cdot \mathbf{x}) = F(\mathbf{x}),\ \forall \mathbf{x}\}. \tag{3.11}$$

This is a nice algebraic (defined by polynomial equations) Lie group, and its integer subgroup

$$\Gamma \;:=\; \mathrm{SO}_F(\mathbb{Z}) \tag{3.12}$$

is a nice arithmetic (the set of integer points on an algebraic group) discrete group.

To make these groups slightly less mysterious, it is a well-known fact (see, e.g., the discussion in [63, §4]) that they can be parametrized, as follows. It can be checked that whenever

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{R}),$$

the matrix

$$g := \begin{pmatrix} \frac{1}{2}\left(a^2-b^2-c^2+d^2\right) & ac-bd & \frac{1}{2}\left(a^2-b^2+c^2-d^2\right) \\ ab-cd & bc+ad & ab+cd \\ \frac{1}{2}\left(a^2+b^2-c^2-d^2\right) & ac+bd & \frac{1}{2}\left(a^2+b^2+c^2+d^2\right) \end{pmatrix} \quad (3.13)$$

is in G. Likewise, Γ is essentially the image under the above morphism of the more familiar discrete group $\mathrm{SL}_2(\mathbb{Z})$.

The set of all primitive Pythagorean triples (up to symmetry) is then given by the orbit:

$$\mathcal{O} \ := \ \Gamma \cdot \mathbf{x}_0, \quad (3.14)$$

where $\mathbf{x}_0$ is any primitive base point $\mathbf{x}_0 \in V(\mathbb{Z})$, e.g.

$$\mathbf{x}_0 = (3,4,5).$$

To study the area, we again consider the function $f(\mathbf{x}) = \frac{1}{12}xy$, and ask for an R so that the set $\mathcal{X}_R$ of $\mathbf{x} \in \mathcal{O}$ with $f(\mathbf{x}) \in \mathcal{P}_R$ is Zariski dense in the cone V in (3.9), which is the Zariski closure of $\mathcal{O}$.

3.3. The General Procedure

We have taken a very simple problem and made it look very complicated. But now we have seen almost all of the essential features of the general

Affine Sieve: One takes

(1) a finitely generated subgroup Γ of $\mathrm{GL}_n(\mathbb{Q})$ (later we will want to relax this to allow semigroups),

(2) some base point $\mathbf{x}_0 \in \mathbb{Q}^n$, which then forms the orbit $\mathcal{O}$ as in (3.14), and

(3) a polynomial function f which takes integer values on $\mathcal{O}$.

With this data, one asks for an (or the smallest) integer $R < \infty$ so that the set

$$\mathcal{X}_R = \mathcal{X}_R(\mathcal{O}, f) \ := \ \{\mathbf{x} \in \mathcal{O} : f(\mathbf{x}) \in \mathcal{P}_R\}$$

is Zariski dense in the Zariski closure of $\mathcal{O}$. In practice, the Zariski density is not hard to establish, so we will simply say that $f(\mathcal{O})$ contains R-almost-primes (or that we have produced R-almost-primes) to mean the more precise statement.

Let us see now how the general Affine Sieve method proceeds. In the notation of (3.2), we wish to sift for R-almost-primes in the set

$$\mathcal{S} \ := \ f(\mathcal{O}).$$

As in (3.3), we must understand the distribution of $\mathcal{S} \cap [1, x]$ among the multiples of q up to some level Q. Roughly speaking, if $\gamma \in \Gamma$ is of size $\|\gamma\|$ about T, then so is the size of $\|\mathbf{x}\|$, where $\mathbf{x} = \gamma \cdot \mathbf{x}_0$, since the base point $\mathbf{x}_0$ is fixed.[4] If f is a polynomial of degree d, then generically $f(\mathbf{x})$ is of size T^d for such an $\mathbf{x}$. Hence restricting $\mathcal{S}$ to $f(\mathbf{x}) < x$ is roughly the same as restricting $\|\gamma\| < T$ with $T = x^{1/d}$. The left hand side of (3.3) may then be captured in essence by

$$\sum_{\substack{\gamma \in \Gamma \\ \|\gamma\| < x^{1/d}}} \mathbf{1}_{\{f(\gamma \cdot \mathbf{x}_0) \equiv 0(q)\}}. \tag{3.15}$$

We should first determine what happens if $q = 1$, that is, when the congruence condition is dropped. Say the group Γ has *exponent of growth*

$$\delta \ > \ 0, \tag{3.16}$$

which means roughly that the number of points in Γ of norm at most T is about T^δ, or

$$\sum_{\substack{\gamma \in \Gamma \\ \|\gamma\| < x^{1/d}}} 1 \ = \ x^{\delta/d + o(1)}. \tag{3.17}$$

Note that for general q, the condition $f(\gamma \cdot \mathbf{x}_0) \equiv 0(q)$ is only a restriction on γ mod q, so we can decompose the sum above into residue classes as

$$\sum_{\gamma_0 \in \Gamma (\mathrm{mod} q)} \mathbf{1}_{\{f(\gamma_0 \cdot \mathbf{x}_0) \equiv 0(q)\}} \left[\sum_{\substack{\gamma \in \Gamma \\ \|\gamma\| < x^{1/d}}} \mathbf{1}_{\{\gamma \equiv \gamma_0(q)\}} \right]. \tag{3.18}$$

The bracketed term above is the key to the whole game. What do we expect? If the euclidean ball in Γ of size $x^{1/d}$ is equidistributed among the possible residue classes mod q, then the bracketed term should be "roughly" equal to

$$\frac{1}{|\Gamma (\mathrm{mod} q)|} \sum_{\substack{\gamma \in \Gamma \\ \|\gamma\| < x^{1/d}}} 1.$$

In reality, one can prove today in some generality thanks to the work of many people (e.g. [94, 68, 69, 58, 2, 19, 21, 101, 37, 6, 8, 9, 52, 10, 82, 98])

(4) For simplicity, take all norms here to be Euclidean, though in many settings it is advantageous (or even necessary, since we do not yet know how to count with archimedean norms in full generality!) to use other norms, e.g., the wordlength metric in the generators of Γ.

an estimate of the form

$$\sum_{\substack{\gamma\in\Gamma\\ \|\gamma\|<x^{1/d}}} \mathbf{1}_{\{\gamma\equiv\gamma_0(q)\}} \;=\; \frac{1}{|\Gamma(\mathrm{mod}q)|}\sum_{\substack{\gamma\in\Gamma\\ \|\gamma\|<x^{1/d}}} 1+O\left(q^C\left[\sum_{\substack{\gamma\in\Gamma\\ \|\gamma\|<x^{1/d}}} 1\right]^{1-\Theta}\right). \tag{3.19}$$

Here $C<\infty$ and $\Theta\geqslant 0$ are some constants, and if

$$\Theta \;>\; 0, \tag{3.20}$$

then Θ is often referred to as a "spectral gap" for Γ.

Results of this type follow (with quite a bit of work in many separate cases) from theorems (or partial results towards conjectures) going under various guises; some of these "buzzwords" are: the Selberg 1/4-Conjecture, the generalized Ramanujan conjectures, mixing rates for homogeneous flows, temperedness of representations, resonance-free regions for transfer operators, expander graphs, among many others; see, e.g. [87, 88, 89, 54, 71, 3].

Now inserting (3.19) into (3.18), using (3.17), and assuming that the proportion of γ_0 in $\Gamma(\mathrm{mod}q)$ with

$$f(\gamma_0\cdot\mathbf{x}_0)\equiv 0(q)$$

is about $1/q$ (for example, f should not be identically zero), we obtain an estimate for (3.15) roughly of the form (3.3), with

$$|r_q| \;\ll\; q^C x^{\delta(1-\Theta)/d}. \tag{3.21}$$

(The value of the constant C may change from line to line.) Again using (3.17) as an approximation for $\#\mathcal{S}\cap[1,x]$, we obtain that (3.4) holds with

$$Q \;=\; X^{\delta\Theta/(Cd)-\varepsilon},$$

say, for any $\varepsilon>0$. Thus the set $\mathcal{S}$ has exponent of distribution

$$\vartheta \;=\; \frac{\delta\Theta}{Cd}, \tag{3.22}$$

and hence contains R-almost-primes with

$$R \;=\; \left\lceil \frac{Cd}{\delta\Theta}+\varepsilon \right\rceil. \tag{3.23}$$

So as long as $C<\infty$, that is, the dependence on q in the error term of (3.19) is at worst polynomial, and as long as the "spectral gap" Θ is strictly positive, this general sieving procedure produces R-almost-prime values in $\mathcal{S}$ for some $R<\infty$.

3.4. Applying the General Procedure

3.4.1. *Fibonacci Composites*

Again it is instructive to first see how the general method can fail to work. Let Γ be the semigroup generated by the square of the matrix $\left(\begin{smallmatrix} 0 & 1 \\ 1 & 1 \end{smallmatrix}\right)$, set $\mathbf{x}_0 = (0, 1)$, with orbit $\mathcal{O} = \Gamma \cdot \mathbf{x}_0$, and consider the function $f(x, y) = x$. It is elementary to check that here

$$\mathcal{S} \;=\; f(\mathcal{O}) \;=\; \{\mathfrak{f}_{2n}\}$$

is just the set of even-indexed Fibonacci numbers, $\mathfrak{f}_{2n}$. This set is much too thin for the above methods to apply, since the number of Fibonacci numbers up to x is about $\log x$; that is, the group Γ has exponent of growth δ in (3.16) equal to zero. In particular, a counting result of the type (3.19) is simply impossible, and one cannot establish a positive exponent of distribution ϑ as in (3.22).

In fact, there seems to be good reason for the sieve to fail in this context. While it is believed that infinitely many Fibonacci numbers are prime, there is some heuristic evidence that if n is composite, then the nth Fibonacci number $\mathfrak{f}_n$ has at least on the order of n prime factors. Assuming this heuristic, there should not exist a finite R for this setting; that is, the sieve does not work here because it must not. (Note that the Zariski closure of the group Γ here is a torus, $\mathbb{C}^\times$; as Sarnak likes to say, for the Affine Sieve, "the torus is the enemy!")

3.4.2. *Back to Pythagorean Areas*

What does the above procedure give for Pythagorean areas? The function $f(\mathbf{x}) = \frac{1}{12}xy$ is quadratic, so $d = 2$. It is not hard to see that Γ has growth exponent δ in (3.16) equal to 1. Selberg's 1/4-Conjecture, if true, would imply an estimate (in smooth form) for (3.19) with "spectral gap" $\Theta = 1/2$; this is again a square-root cancellation type phenomenon. Unconditionally, the best-known bound (due to Kim-Sarnak) proves (3.19) with $\Theta = \frac{1}{2} - \frac{7}{64}$. The value for C coming from (a slight variant of) the above procedure can be whittled down to 2. One small technicality is that our f in (3.8) is now the product of four irreducible factors, so the fraction $1/q$ on the right hand side of (3.3) should be replaced by $4/q$ (giving a sieve of "dimension" 4); the sieve still works in the same way, just with a worse dependence of R in (3.6) on the level of distribution in (3.5).

The above technicalities aside, all this machinery will in the end produce an exponent of distribution ϑ of about 1/10, and about

$$R \;=\; 30 \tag{3.24}$$

primes, falling far short of Green-Tao's optimal result $R = 4$. Of course the orbit $\mathcal{O}$ here is very simply described, making its study amenable to other means. In the following subsection, we give a sampling of problems in which more elementary descriptions do not seem advantageous (or even possible), yet where the Affine Sieve applies as just indicated. We hope these serve to illustrate some of the power and robustness of the Affine Sieve.

3.5. More Examples: Anisotropic and Thin Orbits

3.5.1. *Anisotropic "Areas"*

Keeping a nearly identical setup, let us change ever so slightly the quadratic form F from (3.10) to

$$F(\mathbf{x}) \;=\; x^2 + y^2 - 3z^2.$$

The salient features of this form are that, like (3.10), it is rational (the ratios of its coefficients are in $\mathbb{Q}$) and indefinite (it takes positive and negative values), but unlike (3.10), it is *anisotropic* over $\mathbb{Q}$. This means that it has no non-zero rational points on the cone $F = 0$. (Exercise.) So to have an integral orbit, we can change our variety V from (3.9) to, say,

$$V \;:\; F(\mathbf{x}) = 1,$$

which over $\mathbb{R}$ is a one-sheeted hyperboloid containing the integer base point $\mathbf{x}_0 = (1, 0, 0)$. Let $G = \mathrm{SO}_F(\mathbb{R})$ now be the real special orthogonal group preserving this new form, and let $\Gamma = \mathrm{SO}_F(\mathbb{Z})$ be the arithmetic group of integer matrices in G. Taking the orbit $\mathcal{O} = \Gamma \cdot \mathbf{x}_0$ and function

$$f(\mathbf{x}) = \frac{1}{2}xy$$

as an analogue of "area," one can compute (see [62]) that $\mathcal{S} = f(\mathcal{O})$ is essentially the set of all values of

$$(a^2 - b^2 + 3c^2 - 3d^2)(ab + 3cd), \tag{3.25}$$

where a, b, c, d range over all integers satisfying

$$a^2 + b^2 - 3c^2 - 3d^2 \;=\; 1. \tag{3.26}$$

(In fancier language, the spin group of Γ is isomorphic to the norm one elements of a particular quaternion division algebra.)

Needless to say, the Green-Tao technology of linear equations is not designed to handle this new set $\mathcal{S}$, while the Affine Sieve works in exactly the same way as previously described (in this setting, it was executed by Liu-Sarnak [70]), producing R-almost-primes[5] with $R = 16$.

3.5.1. *A Thin Group*

While the group Γ in §3.5 was more complicated, it was still arithmetic; in particular, any solution in the integers to the polynomial equation (3.26) gave (by a simple formula) an element in Γ. The situation is even more delicate if the group Γ is restricted to some *infinite* index subgroup of $\mathrm{SO}_F(\mathbb{Z})$. Here is a quintessential "thin" (see below for the definition) group.

Let us return again to the Pythagorean setting of (3.10) and the cone (3.9) with base point $\mathbf{x}_0 = (3, 4, 5)$ and the "area" function $f(\mathbf{x})$ in (3.8). For the sake of being explicit, let Γ be the group generated by the two matrices

$$M_1 := \begin{pmatrix} -7 & -4 & -8 \\ 4 & 1 & 4 \\ 8 & 4 & 9 \end{pmatrix} \quad \text{and} \quad M_2 := \begin{pmatrix} -1 & 2 & 2 \\ -2 & 1 & 2 \\ -2 & 2 & 3 \end{pmatrix}, \tag{3.27}$$

which one can check are the images under the morphism (3.13) of $\left(\begin{smallmatrix} 1 & 4 \\ 0 & 1 \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} 1 & 0 \\ 2 & 1 \end{smallmatrix}\right)$, respectively. The orbit

$$\mathcal{O} = \Gamma \cdot \mathbf{x}_0 \tag{3.28}$$

of $\mathbf{x}_0$ under this group Γ is illustrated in Figure 2; this is the picture one may keep in mind when thinking of thin orbits.

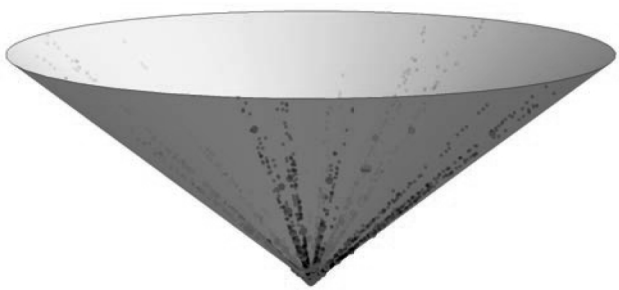

Figure 2. — A piece of the thin Pythagorean orbit $\mathcal{O}$ in (3.28). Points $\mathbf{x} \in \mathcal{O}$ are again marked according to whether the "area" $f(\mathbf{x}) = \frac{1}{12}xy$ is in $\mathcal{P}_R$ with $R \leqslant 3$ (●), $R = 4$ (●), or $R \geqslant 5$ (●).

(5) For the experts, this number is about half of that in (3.24), due to (3.25) being a two-dimensional sieve problem instead of (3.8) which is four-dimensional. In the anisotropic case, there are no "extra" parametrizations like (3.7), so the "area" is only a product of two irreducible factors, not four.

Note that, unlike Figure 1, there now seem to be only finitely many $\mathbf{x} \in \mathcal{O}$ with $f(\mathbf{x})$ having $R \leqslant 3$ prime factors; these are invisible at the scale drawn in Figure 2. The (presumably infinite number of) points of "special" form visible in Figure 1 seem to disappear for this thin orbit, again reinforcing our suggestion that Zariski density is the "right" demand for the general setting.

What do we mean by "thin"? There are a number of competing definitions of this word, and we will need to give a new one to suit our purposes. The meaning of thin typically involved in the Affine Sieve refers to "thin matrix groups" (not to be confused with "thin sets," as defined by Serre [95, §3.1]), which are finitely generated groups $\Gamma < \mathrm{GL}_n(\mathbb{Z})$ which have infinite index in the group of integer points of their Zariski closure. That is, let

$$G = \mathrm{Zcl}(\Gamma) < \mathrm{GL}_n$$

be the Zariski closure of Γ, and $G(\mathbb{Z})$ its integer points; then Γ is called a thin matrix group if the index

$$[G(\mathbb{Z}) : \Gamma] = \infty.$$

For our purposes, we will want to allow Γ to be a finitely generated semi-group of $\mathrm{GL}_n(\mathbb{Z})$, but not necessarily a group. In this case, we cannot speak of index, and need a different condition to characterize what should be considered thin. Moreover, we will want the flexibility to apply the adjective thin to either the (semi-)group Γ, or the resulting orbit $\mathcal{O}$, or the resulting set of integers $\mathcal{S} = F(\mathcal{O})$. Our characterization will simply be by an archimedean degeneracy in the algebro-geometric closure, as follows.

Definition: Thin Integer Set. — Let $\mathcal{Z} \subset \mathbb{Z}^n$ be a set of integer vectors, let $\mathrm{Zcl}(\mathcal{Z})$ be the Zariski closure of $\mathcal{Z}$, and let B_x be a ball of radius $x > 0$ (with respect to any fixed archimedean norm) about the origin in $\mathbb{R}^n$. We will call $\mathcal{Z}$ a *thin integer set* if

$$\#(\mathcal{Z} \cap B_x) \;=\; o\Big(\#(\mathrm{Zcl}(\mathcal{Z}) \cap \mathbb{Z}^n \cap B_x)\Big), \qquad \text{as } x \to \infty.$$

That is, $\mathcal{Z}$ has zero "density" inside the integer points of its Zariski closure.

It is an easy fact [6] that when $\Gamma < \mathrm{GL}_n(\mathbb{Z})$ is group, then it is a thin matrix group if and only if it is a thin integer set in $\mathbb{Z}^{n\times n} \cong \mathbb{Z}^{n^2}$. (Thanks

[6] A sketch for the experts: the trivial representation does not weakly occur in the regular action of $G = \mathrm{Zcl}(\Gamma)$ on $L^2(\Gamma\backslash G)$ if and only if $\mathrm{vol}(\Gamma\backslash G) = \infty$, in which case Howe-Moore gives the decay of matrix coefficients. On the other hand, the count for arithmetic groups is known already by methods of Duke-Rudnick-Sarnak and Eskin-McMullen.

to Peter Sarnak for insisting that we make our definition so that the two definitions would agree on their intersection.)

Our group $\Gamma = \langle M_1, M_2 \rangle$ from (3.27) has infinite index in $\mathrm{SO}_F(\mathbb{Z})$, so is thin. Its exponent of growth can be estimated as

$$\delta \approx 0.59\cdots,$$

which, it turns out, is also the Hausdorff dimension of the *limit set* of Γ. The latter is roughly speaking the Cantor-like fractal set seen at the boundary at infinity in Figure 2; that is, the set of directions in which the orbit $\mathcal{O}$ grows.

Now there is certainly no hope of a more direct approach to studying $\mathcal{S} = f(\mathcal{O})$, as we cannot even determine, given a matrix $M \in \mathrm{SO}_F(\mathbb{Z})$, whether it is in the group Γ. Unlike the arithmetic group case, it is not enough to check whether the entries of M satisfy some polynomial equations; instead one must determine whether M can be realized as some word in the generators (3.27). As the general membership problem in a group is undecidable [80], we had better avoid this issue. Luckily, the standard Affine Sieve procedure works just as described in §3.3. (In this setting, the details were worked out by the author [60, 61], and the author with Oh [59]).

A good question to ask at this point might be: Why would anyone care about these strange thin groups? Here are just two motivations: (1) thin groups are in some sense "generic" (see, e.g., [34, 36, 93], for a discussion into which we will not delve here), and (2) many naturally-arising and interesting problems *require* their study. Let us postpone our discussion of these natural problems for a moment, turning now to another topic.

3.6. The Affine Sieve Captures Primes

We have described the general procedure and explained how it works in a number of sample settings, but it is clear that without further ingredients, producing primes seems hopeless. Yet, as we have already seen in the case of Pythagorean areas, the Green-Tao theorem, using completely different tools, goes far beyond the present capabilities of the Affine Sieve. We give here but a sampling of four more settings in which other technologies prove more successful, producing a minimal number of prime factors.

3.6.1. *Matrix Ensembles with Prime Entries*

Now that we appreciate the utility of posing problems in terms of matrix orbits, why not ask the following even simpler Affine Sieve-type question:

Among the set of all $n \times n$ integer matrices of, say, fixed determinant $D \geqslant 1$, are there infinitely many with all entries prime? For example, here is a prime 3×3 matrix of determinant $D = 4$:

$$\det \begin{pmatrix} 3 & 5 & 7 \\ 11 & 13 & 17 \\ 5 & 13 & 19 \end{pmatrix} = 4. \tag{3.29}$$

How is this an Affine Sieve problem? Let $V_{n,D}(\mathbb{Z})$ be the set in question of all $n \times n$ integer matrices of determinant D. The full[7] group $\mathrm{SL}_n(\mathbb{Z})$ acts on $V_{n,D}(\mathbb{Z})$ on the left (determinant is preserved), and a theorem of Borel and Harish-Chandra tells us that $V_{n,D}(\mathbb{Z})$ breaks up into finitely many such orbits. Thus we may as well just take one fixed matrix $M_0 \in V_{n,D}(\mathbb{Z}) \subset \mathbb{Z}^{n^2}$ and consider the orbit

$$\mathcal{O} = \mathrm{SL}_n(\mathbb{Z}) \cdot M_0.$$

For an $n \times n$ integer matrix $M = (m_{ij}) \in \mathcal{O}$, our function f is now the product of all coordinates,

$$f(M) = \prod_{i,j} m_{ij},$$

which, being a product of n^2 terms, we would like to make R-almost-prime with $R = n^2$.

First let us consider the case $n = 2$, that is, for a given D, we want primes a, b, c, d with

$$ad - bc = D. \tag{3.30}$$

The set of solutions in which at least one of the entries is the even prime 2 is again of "special form," and may be discarded without affecting Zariski density. Thus restricting to odd primes, we see immediately that there is a local obstruction to solving (3.30), namely D had better be even. (In fact, it is not hard to convince oneself that in the $n \times n$ case, there is again a local obstruction unless $D \equiv 0 (\mathrm{mod} 2^{n-1})$, which is why we chose $D = 4$ in (3.29).)

But now (3.30) looks like a "twin prime" type question: When can an even number D be written as the difference, not of two primes, but two E_2's? (An "E_2" is a number which is the product of exactly two primes.) Miraculously, the GPY technology, extended to this setting by Goldston-Graham-Pintz-Yıldırım [38], is able to settle the "Bounded Gaps for E_2's

(7) We will sometimes use "full" as the negation of "thin."

Problem," proving that E_2's differ by at most 6 infinitely often. Thus there are many solutions to (3.30) in the primes for at least one value of D in $\{2, 4, 6\}$, but we do not know which!

Turning now to the higher rank setting of $n \geqslant 3$, the following clever observation was made by Nevo-Sarnak [81]. One can first populate all but the last row with primes, writing

$$M = \begin{pmatrix} * & \cdots & * & * \\ \vdots & * & * & * \\ * & * & * & * \\ m_{n,1} & \cdots & m_{n,n-1} & m_{n,n} \end{pmatrix},$$

say, where each $*$ is a prime and the $m_{n,j}$'s are variables. Then the equation $\det M = D$ is a *linear* equation to be solved in $n \geqslant 3$ prime unknowns. For example, we found (3.29) by setting $D = 4$ and finding the solution

$$(a, b, c) \ = \ (5, 13, 19)$$

to

$$4 \ = \ \det \begin{pmatrix} 3 & 5 & 7 \\ 11 & 13 & 17 \\ a & b & c \end{pmatrix} \ = \ -6a + 26b - 16c.$$

It goes back to I. M. Vinogradov (1937) that linear equations in at least three unknowns can be solved in primes, and thus (overcoming many technicalities to get this simple description to actually work) Nevo-Sarnak are able to completely resolve the higher rank problem.

3.6.2. *Prime Norms in* $\mathrm{SL}_2(\mathbb{Z})$

Here is another problem of Affine Sieve type: Instead of restricting the entries to be prime as above, let us look at the full group $\mathrm{SL}_2(\mathbb{Z})$, say, and consider its set of square-norms. That is, consider the set $\mathcal{S}$ of values of

$$a^2 + b^2 + c^2 + d^2,$$

where $ad - bc = 1$. Does the set $\mathcal{S}$ contain an infinitude of primes?

Again one can apply the general Affine Sieve procedure, but Friedlander-Iwaniec [30] found a more profitable approach. After a linear change of variables, the problem can be converted into solving the system

$$\begin{cases} x^2 + y^2 \ = \ p + 2 \\ z^2 + w^2 \ = \ p - 2 \end{cases} \tag{3.31}$$

for primes p and integers x, y, z, w; that is, we must write both $p+2$ and $p-2$ as sums of two squares. Using a "half-dimensional" sieve and assuming the Elliott-Halberstam Conjecture, Friedlander-Iwaniec are able to solve the system (3.31), thereby (conditionally) resolving the problem in this setting.

3.6.2. *Pseudorandom Primes*

The oldest (and arguably simplest) pseudorandom number generator is the map

$$x \ \mapsto \ gx(\mathrm{mod} p),$$

where p is a prime and g is a primitive root mod p, that is, a generator of $(\mathbb{Z}/p\mathbb{Z})^\times$. For optimal equidistribution (and many other applications; see, e.g., the discussion in [63, §2]), one needs the continued fraction expansion

$$\frac{g}{p} \ = \ [a_1, a_2, \dots, a_k] \ = \ \cfrac{1}{a_1 + \cfrac{1}{a_2 + \ddots}} \tag{3.32}$$

to have only "small" partial quotients, $a_j \leqslant A$, say, for some constant $A > 0$. Does there exist an absolute constant $A > 0$ so that infinitely many such fractions g/p can be found with partial quotients bounded by A?

To turn this into an Affine Sieve problem, observe that (3.32) is equivalent to

$$\begin{pmatrix} 0 & 1 \\ 1 & a_1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & a_2 \end{pmatrix} \cdots \begin{pmatrix} 0 & 1 \\ 1 & a_k \end{pmatrix} \ = \ \begin{pmatrix} * & g \\ * & p \end{pmatrix}.$$

Hence to find such pairs (g, p), one should look at the set of second columns in the *semi-group*

$$\Gamma \ := \ \left\langle \begin{pmatrix} 0 & 1 \\ 1 & a \end{pmatrix} : a \leqslant A \right\rangle^+ \cap \ \mathrm{SL}_2. \tag{3.33}$$

For $A \geqslant 2$, this semigroup is Zariski dense in SL_2, but it is known to be thin (and it is here that we wish to extend the definition of thinness beyond the realm of groups). Instead of using the Affine Sieve, Bourgain and the author [11, 15] developed a version of the Hardy-Littlewood circle method to attack this problem, giving an affirmative answer to the above question: There are infinitely many primes p and primitive roots $g(\mathrm{mod} p)$ so that g/p has all partial quotients bounded by $A = 51$.[8] In fact, they proved a

(8) Added in print: Shinnyih Huang [55] has recently reduced this number to $A = 7$, using refinements due to Frolenkov-Kan [33].

"density" version of Zaremba's Conjecture: Almost every natural number occurs in the set $\mathcal{S}$ of bottom right entries of a matrices in Γ (see [15] or [63, §2] for a precise statement). Thus while Γ is thin, the set $\mathcal{S}$ is not, and no sifting is needed to produce primes in $\mathcal{S}$.

3.6.7. *Prime Apollonian Curvatures*

It seems these days no discourse on thin groups is complete without mention of Apollonian gaskets. Lest we bore the reader, we will not yet again repeat the definitions and pictures, which are readily available elsewhere, e.g., [63, §3]. Nevertheless, the following question is quintessential Affine Sieve: Given a primitive Apollonian gasket $\mathscr{G}$, which primes arise as curvatures in $\mathscr{G}$?

It was proved by Sarnak [90] that infinitely many prime curvatures arise, by finding primitive values of shifted binary quadratic quadratic forms among the curvatures and applying Iwaniec's "half-dimensional" sieve. In this way, he proved that the number of primes up to x which are curvatures in $\mathscr{G}$ is at least of order $x(\log x)^{-3/2}$. Bourgain [17] sharpened the lower bound to $x(\log x)^{-1}$, that is, a positive proportion of the primes arise. Finally, Bourgain and the author [12], again using the circle method instead of the Affine Sieve, obtained an asymptotic formula for this number.

As in §3.6, this is an easy consequence of the stronger theorem that an asymptotic "local-global" principle holds for such curvatures (see [12] and [63, §3] for details). So while the group and orbit in this context are again thin, the set of all curvatures is not, and the primes are obtained as a byproduct.

3.7. Improving Levels of Distribution in the Affine Sieve

We conclude our discussion with two final examples in which one can go beyond the general theory. In these, one is currently not able to produce primes, but instead can improve on the exponent of distribution over that in (3.22), without making new progress on spectral gaps as in (3.19). The idea is to avoid putting the individual estimate (3.21) into the sum (3.4), and instead to try to exploit cancellation from the sum on q up to Q, in some analogy with the Elliott-Halberstam Conjecture. It is not known how to do this in the general Affine Sieve, but for the specific examples below, such estimates have recently been obtained by Bourgain and the author [14, 13].

3.7.1. *McMullen's Arithmetic Chaos Conjecture*

We will not describe the origins and implications of McMullen's (Classical) Arithmetic Chaos Conjecture, referring the reader to his fascinating paper [75] and online lecture notes [76]. The conjecture is implied by an analogue of Zaremba's Conjecture, that, for some $A > 1$, every sufficiently large integer arises (with the "right" multiplicity) in the set $\mathcal{S}$ of traces of matrices in the semigroup Γ in (3.33). At the moment, even a "density" version of this statement, as in §3.6, seems out of reach, but one can ask instead if infinitely many primes appear in $\mathcal{S}$. Not surprisingly, the standard Affine Sieve procedure applies here just as well (now requiring the work of Bourgain-Gamburd-Sarnak [9] to prove a statement functionally as strong as (3.19)). But, if applied directly, this produces a terribly poor exponent of distribution ϑ in (3.22), owing to the terribly poor "spectral gap" Θ. Using different tools in this setting , Bourgain and the author [13] have produced in this context an unconditional exponent of distribution $\vartheta = 1/4$, thus showing that $\mathcal{S}$ contains R-almost-primes with $R = 5$.

3.7.2. *Thin Pythagorean Hypotenuses*

Finally, let us return again to the Pythagorean setting of the quadratic form F in (3.10), the cone $F = 0$, the base point $\mathbf{x}_0 = (3, 4, 5)$, and a thin group Γ as in §3.5. Instead of studying areas, let us now take as our function f the "hypotenuse," $f(\mathbf{x}) = z$. Do infinitely many primes arise in $\mathcal{S} = f(\mathcal{O})$?

If $\mathcal{O}$ were the full orbit of all Pythagorean triples, then, through the parametrization (3.7), we would essentially asking whether primes can be represented as sums of two squares. As is very well-known, Fermat answered in the affirmative almost 400 years ago, namely all primes $\equiv 1 (\mathrm{mod} 4)$ are hypotenuses.

But in the thin setting, it seems quite difficult to produce primes at this time. One new difficulty here is that, unlike other problems described above, we now have not only a thin orbit $\mathcal{O}$, but the set $\mathcal{S}$ of hypotenuses is itself thin! The number of integers in $\mathcal{S}$ up to x, even with multiplicity, is about x^δ, where $\delta < 1$ is the growth exponent of Γ as in (3.16).

What does the Affine Sieve process give? Returning to the exponent of distribution ϑ in (3.22), we see that the degree of the hypotenuse function is $d = 1$, and the value of C can be whittled down to 2 as in §3.4. Moreover, to try to optimize ϑ, we can restrict our attention to thin groups Γ whose growth exponent δ is almost as large as possible, $\delta = 1 - \varepsilon$. Then, even assuming a "square-root" version of (3.19), that is, assuming the "spectral gap" can be set to $\Theta = 1/2$, we obtain a (very conditional) exponent of

distribution $\vartheta = 1/4 - \varepsilon$, producing R-almost-primes in $\mathcal{S}$ with $R = 5$. In [14], Bourgain and the author obtained, again for Γ having growth exponent δ sufficiently close to 1, the exponent of distribution $\vartheta = 7/24 - \varepsilon$ unconditionally, thereby producing R-almost-primes with $R = 4$ in this thin setting. The methods (bilinear forms, exponential sums, and dispersion) are outside the scope of this survey.

Added in proof. — The explicit values of R in §§3.4–3.5 are now outdated; recent work of the author and Jiuzu Hong [53] gives an improvement on the general Affine Sieve procedure which differs slightly from that given here (we will not go into the technicalities). Still, the problem of going beyond these values in specific cases remains, and in these settings (that is, in §3.7), the reported R values are still the best known.

Acknowledgements. — The author is grateful to Dick Gross for the invitation to visit Harvard University, during which time these lecture notes were written. Many thanks to John Friedlander, Andrew Granville, Curt McMullen, Sam Payne, Peter Sarnak, Yitang Zhang, and the referee for comments and suggestions on an earlier draft. Thanks also to the organizers of the "Hyperbolic Geometry and Arithmetic" workshop in Toulouse, November 2012, especially Cyril Lecuire; the second half of these notes was conceived at this meeting (the first half was not yet a theorem). The author is also indebted to Jean-Pierre Otal, without whose persistence these notes would not have materialized.

Bibliography

[1] Barban (M. B.). — The sieve" method and its application to number theory. Uspehi Mat. Nauk, 21, p. 51-102 (1966).

[2] Blomer (V.) and Brumley (F.). — On the Ramanujan conjecture over number fields. Ann. of Math. (2), 174(1), p. 581-605 (2011).

[3] Blomer (V.) and Brumley (F.). — The role of the Ramanujan conjecture in analytic number theory. Bull. Amer. Math. Soc. (N.S.), 50(2), p. 267-320 (2013).

[4] Bombieri (E.) and Davenport (H.). — Small differences between consecutive prime numbers. Proc. Roy. Soc. Ser. A, p. 1-18 (1966).

[5] Bombieri (E.), Friedlander (J.), and Iwaniec (H.). — Primes in arithmetic progressions to large moduli. Acta Math., 156, p. 203-251 (1986).

[6] Bourgain (J.) and Gamburd (A.). — Uniform expansion bounds for Cayley graphs of SL2(Fp). Ann. of Math. (2), 167(2), p. 625-642 (2008).

[7] Bourgain (J.), Gamburd (A.), and Sarnak (P.). — Sieving and expanders. C. R. Math. Acad. Sci. Paris, 343(3), p. 155-159 (2006).

[8] Bourgain (J.), Gamburd (A.), and Sarnak (P.). — Affine linear sieve, expanders, and sum-product. Invent. Math., 179(3), p. 559-644 (2010).

[9] Bourgain (J.), Gamburd (A.), and Sarnak (P.). — Generalization of Selberg's 3/16th theorem and affine sieve. Acta Math, 207, p. 255-290 (2011).

[10] Breuillard (E.), Green (B.), and Tao (T.). — Approximate subgroups of linear groups. Geom. Funct. Anal., 21(4), p. 774-819 (2011).
[11] Bourgain (J.) and Kontorovich (A.). — On Zaremba's conjecture. — Comptes Rendus Mathematique, 349(9), p. 493-495 (2011).
[12] Bourgain (J.) and Kontorovich (A.). — On the local-global conjecture for integral Apollonian gaskets (2012). To appear, Invent. Math., arXiv:1205.4416v1, 63 p. 27.
[13] Bourgain (J.) and Kontorovich (A.). — The affine sieve beyond expansion I: thin hypotenuses (2013). Preprint, arXiv:1307.3535.
[14] Bourgain (J.) and Kontorovich (A.). — Beyond expansion II: Traces of thin semigroups (2013). Preprint, arXiv:1310.7190.
[15] Bourgain (J.) and Kontorovich (A.). — On Zaremba's conjecture. Annals Math., 180(1), p. 137-196 (2014).
[16] Bombieri (E.). — On the large sieve. Mathematika, 12, p. 201-225 (1965).
[17] Bourgain (J.). — Integral Apollonian circle packings and prime curvatures. J. Anal. Math., 118(1), p. 221-249 (2012).
[18] Brun (V.). — Le crible d'Eratosthéne et le theoréme de Goldbach. C. R. Acad. Sci. Paris, 168, p. 544-546 (1919).
[19] Burger (M.) and Sarnak (P.). — Ramanujan duals II. Invent. Math, 106, p. 1-11 (1991).
[20] Chen (J. R.). — On the representation of a larger even integer as the sum of a prime and the product of at most two primes. Sci. Sinica, 16, p. 157-176 (1973).
[21] Clozel (L.). — Démonstration de la conjecture τ. Invent. Math., 151(2), p. 297-328 (2003).
[22] Daveport (H.). — Multiplicative Number Theory, volume 74 of Grad. Texts Math. Springer-Verlag, New York (1980).
[23] Deuring (M.). — Imaginäre quadratische Zahlkörper mit der Klassenzahl 1. Math. Z., 37(1), p. 405-415 (1933).
[24] de la Vallée-Poussin (Ch.J.). — Recherches analytiques sur la théorie des nombers premiers. Ann. Soc. Sci. Bruxelles, 20, p. 183-256 (1896).
[25] Elliott (P.D.T.A.) and Halberstam (H.). — A conjecture in prime number theory. Symp. Math. IV (Rome 1968/69), p. 59-72 (1968).
[26] Erdös (P.). — The difference between consecutive primes. Duke Math J., 6, p. 438-441 (1940).
[27] Friedlander (J.) and Granville (A.). — Limitations to the equidistribution of primes. I. Ann. of Math. (2), 129(2), p. 363-382 (1989).
[28] Fouvry (E.) and Iwaniec (H.). — Primes in arithmetic progressions. Acta Arith., 42, p. 197-218 (1983).
[29] Friedlander (J.) and Iwaniec (H.). — What is ... the parity phenomenon? Notices Amer. Math. Soc., 56(7), p. 817-818 (2009).
[30] Friedlander (J.) and Iwaniec (H.). — Hyperbolic prime number theorem. Acta Math., 202(1), p. 1-19 (2009).
[31] Friedlander (J.) and Iwaniec (H.). — Opera de cribro, volume 57 of American Mathematical Society Colloquium Publications. American Mathematical Society, Providence, RI (2010).
[32] Friedlander (J.) and Iwaniec (H.). — Close encounters among the primes (2014). arXiv:1312.2926.
[33] Frolenkov (D.) and Kan (I. D.). — A reinforcement of the Bourgain-Kontorovich's theorem by elementary methods II (2013). Preprint, arXiv:1303.3968.

[34] Fuchs (E.), Meiri (C.), and Sarnak (P.). — Hyperbolic monodromy groups for the hypergeometric equation and Cartan involutions, 2012. To appear, JEMS.

[35] Fouvry (E.). — Autour du théorème de Bombieri-Vinogradov. Acta Math, 152, p. 219-244 (1984).

[36] Fuchs (E.). — The ubiquity of thin groups (2012). To appear, MSRI Proceedings.

[37] Gamburd (A.). — On the spectral gap for infinite index "congruence" subgroups of SL2(Z). Israel J. Math., 127, p. 157-200 (2002).

[38] Goldston (D. A.), Graham (S. W.), Pintz (J.), and Yildirim (C. Y.). — Small gaps between products of two primes. Proc. London Math. Soc., 98(3), p. 741-774 (2009).

[39] Goldfeld (D.). — The class number of quadratic fields and the conjectures of Birch and Swinnerton-Dyer. Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4), 3(4), p. 624-663 (1976).

[40] Goldfeld (D.). — Gauss's class number problem for imaginary quadratic fields. Bull. Amer. Math. Soc. (N.S.), 13(1), p. 23-37 (1985).

[41] Goldston (D. A.). — On Bombieri and Davenport's theorem concerning small gaps between primes. Mathematika, 39(1), p. 10-17 (1992).

[42] Goldston (D. A.), Pintz (J.), and Yildirim (C. Y.). — The path to recent progress on small gaps between primes. In Analytic number theory, volume 7 of Clay Math. Proc., pages 129-139. Amer. Math. Soc., Providence, RI (2007).

[43] Goldston (D. A.), Pintz (J.), and Yildirim (C. Y.). — Primes in tuples I. Ann. of Math. (2), 170(2), p. 819-862 (2009).

[44] Goldston (D. A.), Pintz (J.), and Yildirim (C. Y.). — Primes in tuples II. Acta Math., 204, p. 1-47 (2010).

[45] Granville (A.). — Harald Cramér and the distribution of prime numbers. Scand. Actuar. J., (1), p. 12-28 (1995). Harald Cramér Symposium (Stockholm, 1993).

[46] Green (B.). — Approximate groups and their applications: work of Bourgain, Gamburd, Helfgott, and Sarnak. Current Events Bulletin, AMS (2010).

[47] Green (B.) and Tao (T.). — Linear equations in primes. Ann. of Math. (2), 171(3), p. 1753-1850 (2010).

[48] Gross (B. H.) and Zagier (D. B). — Heegner points and derivatives of L-series. Invent. Math., 84(2), p. 225-320 (1986).

[49] Hadamard (J.). — Sur la distribution des zéros de la fonction $\zeta(s)$ et ses conséquences arithmétiques. Bull. Soc. Math. France, 24, p. 199-220 (1896).

[50] Heath-Brown (D. R.). — Prime twins and Siegel zeros. Proc. London Math. Soc. (3), 47(2), p. 193-224 (1983).

[51] Heilbronn (H.). — On the class number in imaginary quadratic elds. Quarterly J. of Math., 5, p. 150-160 (1934).

[52] Helfgott (H. A.). — Growth and generation in $SL_2(\mathbb{Z}/p\mathbb{Z})$. Ann. of Math. (2), 167(2), p. 601-623 (2008).

[53] Hong (J.) and Kontorovich (A.). — Almost prime coordinates for anisotropic and thin Pythagorean orbits (2014). To appear, Israel J. Math. arXiv:1401.4701.

[54] Hoory (S.), Linial (N.), and Wigderson (A.). — Expander graphs and their applications. Bull. Amer. Math. Soc. (N.S.), 43(4), p. 439-561 (electronic), 2006.

[55] Huang (S.). — An improvement on Zaremba's conjecture (2013). Preprint, arXiv:1310.3772.

[56] Huxley (M. N.). — Small differences between consecutive primes. II. Mathematika, 24, p. 142-152 (1977).

[57] Iwaniec (H.) and Kowalski (E.). — Analytic number theory, volume 53 of American Mathematical Society Colloquium Publications. American Mathematical Society, Providence, RI, 2004.

[58] Kim (H. H.). — Functoriality for the exterior square of GL4 and the symmetric fourth of GL2. J. Amer. Math. Soc., 16(1), p. 139-183 (electronic) (2003). With appendix 1 by Dinakar Ramakrishnan and appendix 2 by Kim and Sarnak (P.).

[59] Kontorovich (A.) and Oh (H.). — Almost prime Pythagorean triples in thin orbits. J. reine angew. Math., 667, p. 89-131 (2012). arXiv:1001.0370.

[60] Kontorovich (A. V.). — The Hyperbolic Lattice Point Count in Infinite Volume with Applications to Sieves. Columbia University Thesis (2007).

[61] Kontorovich (A.). — The hyperbolic lattice point count in infinite volume with applications to sieves. Duke J. Math., 149(1), p. 1-36 (2009). arXiv:0712.1391.

[62] Kontorovich (A.). — Expository note: an arithmetic surface (2011). Unpublished note, http://math.yale.edu/ avk23/files/UniformLattice. pdf.

[63] Kontorovich (A.). — From Apollonius to Zaremba: local-global phenomena in thin orbits. Bull. Amer. Math. Soc. (N.S.), 50(2), p. 187-228 (2013).

[64] Kowalski (E.). — Sieve in expansion. Séminaire Bourbaki, 63(1028), p. 1-35 (2011).

[65] Landau (E.). — Über die Klassenzahl imaginär-quadratischer Zahlkörper. Nachr. Ges. Wiss. Gottingen, p. 285-295 (1918).

[66] Landau (E.). — Bemerkungen zum Heilbronnschen Satz. Acta Arith, p. 1-18 (1935).

[67] Linnik (U. V.). — The large sieve. C. R. (Doklady) Acad. Sci. URSS (N.S.), 30, p. 292-294 (1941).

[68] Lax (P.D.) and Phillips (R.S.). — The asymptotic distribution of lattice points in Euclidean and non-Euclidean space. Journal of Functional Analysis, 46, p. 280-350 (1982).

[69] Luo (W.), Rudnick (Z.), and Sarnak (P.). — On Selberg's eigenvalue conjecture. Geom. Funct. Anal., 5(2), p. 387-401 (1995).

[70] Liu (J.) and Sarnak (P.). — Integral points on quadrics in three variables whose coordinates have few prime factors. Israel J. Math, 178, p. 393-426 (2010).

[71] Lubotzky (A.). — Expander graphs in pure and applied mathematics. Bull. Amer. Math. Soc., 49, p. 113-162 (2012).

[72] Maier (H.). — Primes in short intervals. Michigan Math. J., 32(2), p. 221-225 (1985).

[73] Maier (H.). — Small differences between prime numbers. Michigan Math J., 35, p. 323-344 (1988).

[74] Maynard (J.). — Small gaps between primes (2013). Preprint, arXiv:1311.4600.

[75] McMullen (C. T.). — Uniformly Diophantine numbers in a fixed real quadratic field. Compos. Math., 145(4), p. 827-844 (2009).

[76] McMullen (C. T.). — Dynamics of units and packing constants of ideals, 2012. Online lecture notes, http://www.math.harvard.edu/ ctm/ expositions/home/text/papers/cf/slides/slides.pdf.

[77] Montgomery (H. L.). — Topics in Multiplicative Number Theory, volume 227 of Lecture Notes in Math. Springer, New York (1971).

[78] Mordell (L. J.). — On the riemann hypothesis and imaginary quadratic fields with a given class number. J. London Math. Soc., 9, p. 289-298 (1934).

[79] Motohashi (Y) and Pintz (J.). — A smoothed GPY sieve. Bull. Lond. Math. Soc., 40(2), p. 298-310 (2008).

[80] Novikov (P. S.). — Ob algoritmičeskoĭ nerazrešimosti problemy toždestva slov v teorii grupp. Trudy Mat. Inst. im. Steklov. no. 44. Izdat. Akad. Nauk SSSR, Moscow (1955).

[81] Nevo (A.) and Sarnak (P.). — Prime and almost prime integral points on principal homogeneous spaces (2009).

[82] Pyber (L.) and Szabo (E.). — Growth in finite simple groups of lie type of bounded rank, 2010. Preprint arXiv:1005.1858.

[83] Rényi (A.). — On the representation of an even number as the sum of a single prime and single almost-prime number. Izvestiya Akad. Nauk SSSR. Ser. Mat., 12, p. 57-78 (1948).

[84] Rankin (R. A.). — The difference between consecutive prime numbers. II. Proc. Cambridge Philos. Soc., 36, p. 255-266 (1940).

[85] Riemann (B.). — Ueber die Anzahl der Primzahlen unter einer gegebenen Grösse. Monatsberichte der Berliner Akademie (1859).

[86] Roth (K.F.). — On the large sieves of Linnik and Rényi. Mathematika, 12, p. 1-9 (1965).

[87] Sarnak (P.). — Selberg's eigenvalue conjecture. Notices Amer. Math. Soc., 42(11), p. 1272-1277 (1995).

[88] Sarnak (P.). — What is... an expander? Notices Amer. Math. Soc., 51(7), p. 762-763 (2004).

[89] Sarnak (P.). — Notes on the generalized Ramanujan conjectures. In Harmonic analysis, the trace formula, and Shimura varieties, volume 4 of Clay Math. Proc., pages 659-685. Amer. Math. Soc., Providence, RI (2005).

[90] Sarnak (P.). — Letter to J. Lagarias (2007). http://web.math.princeton.edu/sarnak/AppolonianPackings.pdf.

[91] Sarnak (P.). — Equidistribution and primes. Astérisque, (322), p. 225-240 (2008). Géométrie différentielle, physique mathématique, mathématiques et société. II.

[92] Sarnak (P.). — Affine sieve (2010). Slides from lectures, http://www.math.princeton.edu/sarnak/Affinesievesummer2010.pdf.

[93] Sarnak (P.). — Notes on thin matrix groups. In Thin Groups and Superstrong Approximation, volume 61 of Mathematical Sciences Research Institute Publications, p. 343-362. Cambridge University Press (2014).

[94] Selberg (A.). — On the estimation of Fourier coefficients of modular forms. Proc. of Symposia in Pure Math., VII, p. 1-15 (1965).

[95] Serre (J.-P.). — Topics in Galois theory, volume 1 of Res. Notes in Math. A.K. Peters (2008).

[96] Salehi Golsefidy (A.). — Affine sieve and expanders (2012). To appear, Proceedings of MSRI.

[97] Salehi Golsefidy (A.) and Sarnak (P.). — Affine sieve (2011). To appear, JAMS.

[98] Salehi Golsefidy (A.) and Varjú (P. P.). — Expansion in perfect groups. Geom. Funct. Anal., 22(6), p. 1832-1891 (2012).

[99] Siegel (C. L.). — Über die Classenzahl quadratischer Zahlkörper. Acta Arith, 1, p. 83-86 (1935).

[100] Soundararajan (K.). — Small gaps between prime numbers: the work of Goldston-Pintz-Yildirim. Bull. Amer. Math. Soc. (N.S.), 44(1), p. 1-18 (2007).

[101] Sarnak (P.) and Xue (X.). — Bounds for multiplicities of automorphic representations. Duke J. Math., 64(1), p. 207-227 (1991).

[102] Vinogradov (A. I.). — The density hypothesis for Dirichet L-series. Izv. Akad. Nauk SSSR Ser. Mat., 29, p. 903-934, 1965.
[103] Walfisz (A.). — Zur additiven Zahlentheorie. II. Math. Z., 40(1), p. 592-607 (1936).
[104] Zhang (Y.). — Bounded gaps between primes (2013). To appear, Annals Math. 2