Contents

1 Convergence results for Fourier series

1.1 A historical perspective

Fourier series made their first appearance in mathematics through the way of physics, as is the norm with a lot of classical analysis. In particular, they turned up in the analysis of the following differential equation

$$\frac{\partial^2 u}{\partial r^2} + \frac{1}{r}\frac{\partial u}{\partial r} + \frac{1}{r^2}\frac{\partial^2 u}{\partial \theta^2} = 0 \tag{1}$$

on the unit disc, with the following boundary conditions

$$u(1, \theta) = f(\theta)$$

where f is an arbitrary continuous function on $[-\pi, \pi]$ and $f(-\pi) = f(\pi)$.

Equation 1 is what is called the Laplace equation, which describes the temperature u on a conducting surface at steady state. This naturally raises the question: What would the temperature profile on a disk look like at steady state if the boundary was maintained at some temperature $f(\theta)$. That is precisely what the described differential equation along with the boundary conditions tries to answer.

Once one makes the simplifying assumption that the solution is separable, i.e.

$$u(r, \theta) = a(r) \cdot b(\theta)$$

equation 1 is easy to solve, and one sees that it has solutions of the form

$$u_n(r, \theta) = r^n e^{in\theta}$$

for all integers n. And finite linear combinations of the solutions u_n are also solutions to the differential equation 1. But there is one problem. Not every continuous function f on $[-\pi, \pi]$ is a finite linear combination of $e^{in\theta}$, e.g. $|x|$. Leaving that problem aside for now, one notes that if f is of the form

$$f(\theta) = \sum_{n=-N}^{N} \widehat{f}(n)e^{in\theta}$$

then $\widehat{f}(n)$ is can be determined in the following manner

$$\widehat{f}(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)e^{-int}dt \tag{2}$$

Here $\widehat{f}(n)$ is called the nth Fourier coefficient of f. Since we assumed that f is a finite linear combination of $e^{-in\theta}$, that means it has only finitely many non-zero coefficients. Contrariwise, continuous functions that are not finite linear combinations of $e^{in\theta}$ must have infinitely many non-zero Fourier coefficients. Could it perhaps be that f is an *infinite* linear combination of $e^{in\theta}$? To be more precise, does the following sequence converge to f (in an appropriate sense) as N goes to ∞?

$$S_N(\theta) = \sum_{n=-N}^{N} \widehat{f}(n)e^{in\theta}$$

The best we can hope for is for the sequence S_N to converge uniformly to the function f, but as we shall see, that is not actually true, i.e. there are continuous functions whose Fourier series diverges for some point in the domain. The next step would be either weakening the mode of convergence, i.e. instead of expecting convergence in the L^∞ norm, one could expect convergence in weaker norms such as L^1 or L^2, or even Cesàro convergence. Another possible way to go about would be to strengthen the conditions on the function, i.e. forcing the function to be C^1, or even absolutely continuous ensures that the sequence S_N converges uniformly to f.

1.2 Fourier series of continuous functions are Cesàro summable

Before we go on to show that Fourier series converge under weaker notions of convergence, we'll develop a technique that can be used more generally to show certain sequences of functions converge uniformly. Furthermore, in all the sections that are to follow, we'll be working with continuous complex valued functions on the circle T, which is defined as the space $\mathbb{R}$ quotiented with the equivalence relation $\sim$ defined as

$$a \sim b \iff (a - b) = 2n\pi, \ n \in \mathbb{Z}$$

1.2.1 Convolution

For two continuous functions f and g on T, we can define a binary operation $*$

$$(f * g)(x) = \int_{-\pi}^{\pi} f(x - t)g(t)dt$$

It follows that the operation is commutative (substitution of variables), associative (again substitution of variable), and distributive (integration is a linear operator). It also follows that $f * g$ is continuous since f is uniformly continuous. It's not too hard to show that this operation has no identity element (It can be done by showing that if such a function existed, it would take on the value 0 for every point in its domain except 0. But then its integral would also be 0. Contradiction!).

1.2.2 Dirac sequences

We saw that there is no continuous function which acts as an identity for $*$ operation. However, there do exist sequences of functions, such that the sequence $\{f * g_n\}$ converges uniformly to f as n goes to infinity. Such a sequence $\{g_n\}$ is called a Dirac sequence. The formal definition of a Dirac sequence is the following

Definition 1.1. A sequence of a continuous functions $\{g_n\}$ is called a Dirac sequence if it satisfies the following conditions

1. $g_n(x) \geq 0$ for all $n \in \mathbb{N}$ and all $x \in T$.

2. $g_n(x) = g_n(-x)$ for all $n \in \mathbb{N}$ and all $x \in T$.

3. $\int_{-\pi}^{\pi} g_n(t)dt = 1$ for all $n \in \mathbb{N}$.

4. For all $\varepsilon > 0$ and $\delta > 0$, there exists an N such that for all $n > N$,

$$\int_{-\pi}^{-\delta} g_n(t)dt + \int_{\delta}^{\pi} g_n(t)dt < \varepsilon$$

Following from the definition, one gets a useful theorem

Theorem 1.1. *If $\{g_n\}$ is a Dirac sequence, then for all continuous functions f, the sequence $\{f * g_n\}$ converges uniformly to f.*

Proof. Pick any $\varepsilon > 0$. We need to show there exists an N such that for all $n > N$, $\|(f * g_n) - f\|_\infty < \varepsilon$. Pick a particular continuous function f. Let M be the maximum of $|f|$ on T. Pick $\varepsilon_1 > 0$ and $\varepsilon_2 > 0$ such that

$$\varepsilon_1\varepsilon_2 + \varepsilon_2 + 2M\varepsilon_1 < \varepsilon \tag{3}$$

Since f is uniformly continuous on T, pick a δ such that $|x - y| < \delta$ implies $|f(x) - f(y)| < \varepsilon_2$.

Now pick an N such that for all $n > N$

$$\int_{-\pi}^{-\delta} g_n(t)dt + \int_{\delta}^{\pi} g_n(t)dt < \varepsilon_1$$

This would imply for all $n > N$

$$\int_{-\delta}^{\delta} g_n(t)dt > 1 - \varepsilon_1$$

Now consider $(f * g_n)(x_0)$ for some $x \in T$.

$$(f * g_n)(x_0) = \int_{-\pi}^{\pi} f(x_0 - t)g_n(t)dt$$
$$= \left(\int_{-\pi}^{-\delta} f(x_0 - t)g_n(t)dt + \int_{\delta}^{\pi} f(x_0 - t)g_n(t)dt \right) + \left(\int_{-\delta}^{\delta} f(x_0 - t)g_n(t)dt \right)$$

Let's analyse the two terms separately. Since $-M \le f(x_0 - t) \le M$, we can bound the first term as

$$-M\varepsilon_1 \le \int_{-\pi}^{-\delta} f(x_0 - t)g_n(t)dt + \int_{\delta}^{\pi} f(x_0 - t)g_n(t)dt \le M\varepsilon_1$$

Similarly, for all x in the interval $(x_0 - \delta, x_0 + \delta)$

$$f(x_0) - \varepsilon_2 \le f(x) \le f(x_0) + \varepsilon_2$$

This lets us bound the second term in the following manner

$$(1 - \varepsilon_1)(f(x_0) - \varepsilon_2) \le \int_{-\delta}^{\delta} f(x_0 - t)g_n(t)dt \le f(x_0) + \varepsilon_2$$

4

This gives us a complete bound on $(f * g_n)(x_0)$.

$$-M\varepsilon_1 + (1 - \varepsilon_1)(f(x_0) - \varepsilon_2) \leq (f * g_n)(x_0) \leq M\varepsilon_1 + (f(x_0) + \varepsilon_2)$$

Using the triangle inequality and the fact that $|f(x_0)| < M$, we get

$$f(x_0) - (\varepsilon_1\varepsilon_2 + \varepsilon_2 + 2M\varepsilon_1) \leq (f * g_n)(x_0) \leq f(x_0) + (\varepsilon_1\varepsilon_2 + \varepsilon_2 + 2M\varepsilon_1)$$

But from inequality 3, we get

$$f(x_0) - \varepsilon \leq (f * g_n)(x_0) \leq f(x_0) + \varepsilon$$

This shows that $\{f * g_n\}$ converges uniformly to f for all continuous f. $\qquad\square$

This is a useful result because this will let us deal with the question of convergence of the partial Fourier series of some continuous function f. One can write the nth partial Fourier series as the convolution of f with some function g_n, called a kernel, and if one shows that the kernels g_n form a Dirac sequence, then the Fourier series also converges.

1.2.3 Cesàro summability

We mentioned that for general continuous functions, their Fourier series need not converge uniformly, or even pointwise, to the function; weakening the notion of convergence lets the conjecture go through. A weaker notion of convergence for infinite sums is the notion of Cesàro summability.

Definition 1.2. A sequence $\{x_n\}$ is said to be Cesàro summable if the following sequence $\{\sigma_n\}$ converges

$$\sigma_n = \frac{\sum_{k=1}^{n} s_k}{n}$$

where s_k is the kth partial sum of the sequence $\{x_n\}$. If the sequence $\{\sigma_n\}$ converges to L, then L is called the Cesàro sum of $\{x_n\}$.

It's easy to see that if the series $\sum_{k=1}^{n} x_n$ converges, then the Cesàro sums also converge to the same limit. To show it is a strictly weaker notion of convergence, consider the sequence $x_n = (-1)^n$. Clearly, the partial sums of x_n do not converge, but the Cesàro sums do converge to $\frac{1}{2}$.

1.2.4 The Fejér kernel

As outlined in a previous subsection, one would like to find functions g_N such that

$$(f * g_N)(\theta) = \sum_{n=-N}^{N} \widehat{f}(n)e^{in\theta}$$

where $\widehat{f}(n)$ is the nth Fourier coefficient of f. Rewriting the above sum, we get

$$\sum_{n=-N}^{N} \widehat{f}(n)e^{in\theta} = \sum_{n=-N}^{N} \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} f(t)e^{-int}dt \right) e^{in\theta}$$

Since this is a finite sum, we can exchange the sum and the integral to get

$$\sum_{n=-N}^{N} \widehat{f}(n)e^{in\theta} = \int_{-\pi}^{\pi} \left(\frac{1}{2\pi} \sum_{n=-N}^{N} e^{in(\theta-t)} \right) f(t)dt$$

Summing up the geometric series, we get

$$\frac{1}{2\pi} \sum_{n=-N}^{N} e^{int} = \frac{1}{2\pi} \frac{\sin\left(Nt + \frac{t}{2}\right)}{\sin\left(\frac{t}{2}\right)}$$

The function g_N we wanted was is this function

$$g_N(t) = \frac{1}{2\pi} \frac{\sin\left(Nt + \frac{t}{2}\right)}{\sin\left(\frac{t}{2}\right)} \tag{4}$$

g_N certainly is even, and its integral over T is 1, but it is not non-negative everywhere, hence it fails to form a Dirac sequence. Let's look at the Cesàro partial sums instead, and concentrates on the kernel corresponding to those, which are called Fejér kernels. Clearly, the kernel corresponding to the Nth Cesàro sum would be the following

$$F_N = \frac{\sum_{n=1}^{N} g_n}{N}$$

Summing up the geometric progression once again, we get the following closed form expression for F_N

$$F_N(t) = \frac{1}{2N\pi} \frac{\sin^2\left(\frac{Nt}{2}\right)}{\sin^2\left(\frac{t}{2}\right)}$$
$$= \frac{1}{2N\pi} \frac{1 - \cos(nt)}{1 - \cos(t)}$$

It follows that F_N is even and its integral over T is 1. Furthermore, it is non-negative, and finally we have the following inequality for $0 < \delta < |t| \le \pi$

$$\frac{1}{2N\pi} \frac{1 - \cos(nt)}{1 - \cos(t)} \le \frac{1}{2N\pi} \frac{2}{1 - \cos(t)}$$
$$\le \frac{1}{N\pi} \frac{1}{1 - \cos(\delta)}$$

This shows that $\{F_N\}$ satisfies all the conditions required to be a Dirac sequence. This easily leads to the following result.

Theorem 1.2 (Fejér's Theorem). *The Fourier series of a continuous function is Cesàro summable, and the partial Cesàro sums converge uniformly to f.*

Proof. The Fejér kernels form a Dirac sequence. Now use the result of theorem 1.1. $\square$

This result gives us two things: it tells us that the exponential polynomials (linear combinations of $e^{in\theta}$) are dense in $C(T)$, and for a given f in $C(T)$, it gives us an *explicit* sequence of polynomials which converge uniformly to f. The first result is not that impressive, since it also follows from Stone-Weierstrass theorem (exponential polynomials form a subalgebra of $C(T)$ containing a non-zero constant function and it separates points), but the Stone-Weierstrass theorem does not give an explicit sequence, which this result does.

1.3 Sufficient conditions for convergence of Fourier series

Now that we've established the Fourier series of continuous functions are Cesàro summable, which is a notion of convergence strictly weaker than uniform convergence, we'll proceed in the other direction, which involves strengthening conditions on the function f.

Clearly, if the Fourier coefficients of f are absolutely convergent, then that would mean the Fourier series also converges uniformly (triangle inequality). While this is rather strong condition, this does give us an idea to look at the size of Fourier coefficients in order to determine convergence. This leads us to a lemma:

Lemma 1.3. *If* $f \in C^1(T)$, *then* $\widehat{f}(n)$ *is in* $o\left(\frac{1}{n}\right)$.

Proof. Consider the Fourier coefficient of f', which is a continuous function.

$$\widehat{f'}(n) = \int_{-\pi}^{\pi} f'(t)e^{-int}dt$$

$$= \left[e^{-int}f(t)\right]_{-\pi}^{\pi} + in\int_{-\pi}^{\pi} f(t)e^{-int}dt \qquad \text{(Integration by parts)}$$

$$= in\widehat{f}(n)$$

This means

$$\left|\widehat{f}(n)\right| = \left|\frac{\widehat{f'}(n)}{n}\right|$$

Since $\widehat{f'}(n)$ is in $o(1)$, that means $\widehat{f}(n)$ is in $o\left(\frac{1}{n}\right)$. $\qquad\square$

This leads to the following corollary.

Corollary 1.4. *If* $f \in C^k(T)$, *then* $\widehat{f}(n)$ *is in* $o\left(\frac{1}{n^k}\right)$.

From the above result, we get that if $f \in C^2(T)$, then $\widehat{f}(n)$ is in $o\left(\frac{1}{n^2}\right)$, which means the Fourier sums are absolutely convergent, and the Fourier series converges uniformly to f.

The final result of this section will be that if f is continuous and $\widehat{f}(n)$ is $O\left(\frac{1}{n}\right)$, the Fourier series converges uniformly to f. In particular, this will give us that the Fourier series of C^1 converge uniformly since the Fourier coefficients are in $o\left(\frac{1}{n}\right)$.

Theorem 1.5. *If* f *is a continuous function and* $\widehat{f}(n)$ *is* $O\left(\frac{1}{n}\right)$, *then the partial Fourier sums* $S_n(f)$ *converge uniformly to* f *on* T. *[2]*

Proof. What we will do in this proof is create an intermediate series $I_{(n,k)}$, between S_n and σ_n (the partial Cesàro sums), such that $I_{(n,k)}(f)$ converges uniformly to f as n goes to ∞ for large enough k, and the distance between $I_{(n,k)}(f)$ and $S_m(f)$, where $kn \le m < (k+1)n$, is less than $\frac{A}{k}$ for some constant A.

Define the series $I_{(n,k)}$ as

$$I_{n,k}(f)(\theta) = \left(k + 1 + \frac{1}{n}\right)\sigma_{(k+1)n+1}(f)(\theta) - \left(k + \frac{1}{n}\right)\sigma_{kn+1}(f)(\theta) \qquad (5)$$

Since we know that $\sigma_n(f)$ converges uniformly to f, expression 5 will also converge to f as n goes to ∞, for all k, although the rate of convergence will depend on k.

Now consider another way of writing $I_{(n,k)}(f)(\theta)$ (one needs to write out all the individual terms to see this)

$$I_{(n,k)}(f)(\theta) = \sum_{q=-kn}^{kn} \widehat{f}(q)e^{iq\theta} + \sum_{kn<|q|\leq(k+1)n} \frac{(k+1)n+1-|q|}{n}\widehat{f}(q)(n)e^{iq\theta}$$

Since $\widehat{f}(k)$ is $O\left(\frac{1}{k}\right)$, it is bounded above by $\frac{A}{k}$ for some constant A.

Writing $S_m(f)(\theta)$ similarly

$$S_m(f)(\theta) = \sum_{q=-kn}^{kn} \widehat{f}(q)e^{iq\theta} + \sum_{kn<|q|\leq m} \widehat{f}(q)(n)e^{iq\theta}$$

where $kn \leq m < (k+1)n$. Taking their difference, we get

$$|I_{(n,k)}(f)(\theta) - S_m(f)(\theta)| \leq \sum_{kn<|q|\leq(k+1)n} 2|\widehat{f}(q)|$$

$$\leq 4n \cdot \frac{A}{kn}$$

$$= \frac{4A}{k}$$

For $\varepsilon > 0$, pick a k such that $\frac{2A}{k} < \frac{\varepsilon}{2}$. Pick an $N > k$ such that for all $n > N$,

$$\left\|I_{(n,k)}(f) - f\right\|_\infty < \frac{\varepsilon}{2} \tag{6}$$

Now pick any $m > kN$. Since $N > k$, we have the following inequality

$$k(n+1) < (k+1)n$$

where $n > N$. That means for every $m > kN$, there exists $n > N$, such that

$$kn \leq m < (k+1)n$$

That means for all $m > kN$

$$\left\|S_m(f) - I_{(n,k)}(f)\right\|_\infty < \frac{\varepsilon}{2} \tag{7}$$

Combining inequalities 6 and 7, we get

$$\left\|S_m(f) - f\right\|_\infty < \varepsilon$$

This shows that the partial Fourier sums converge uniformly the function f. $\qquad\square$

1.4 A continuous function whose Fourier series diverges at a point [1]

The last section showed that the Fourier series of a C^1 function does converge; this section aims to show that that's almost the best result one can get, i.e. there does exist a continuous function whose Fourier series not only does not converge to the function uniformly, it does not even converge pointwise.

Consider the sequence of kernels described in equation 4

$$g_N(t) = \frac{1}{2\pi} \frac{\sin\left(Nt + \frac{t}{2}\right)}{\sin\left(\frac{t}{2}\right)}$$

where $(f * g_N)$ corresponds to the Nth partial Fourier sum of f. The first thing we will do is bound the L^1 norm of g_N from below.

$$
\begin{aligned}
\|g_N\|_1 &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \frac{\sin\left(Nt + \frac{t}{2}\right)}{\sin\left(\frac{t}{2}\right)} \right| dt \\
&\geq \frac{1}{\pi} \int_{-\pi}^{\pi} \left| \frac{\sin\left(Nt + \frac{t}{2}\right)}{t} \right| dt && (|x| \geq |\sin(x)|) \\
&= \frac{1}{\pi} \int_{-\pi\left(N+\frac{1}{2}\right)}^{\pi\left(N+\frac{1}{2}\right)} \left| \frac{\sin(y)}{y} \right| dy \\
&> \frac{1}{\pi} \sum_{k=1}^{\frac{N}{2}} \int_{-\pi(k+1)}^{-\pi(k-1)} \left| \frac{\sin(y)}{y} \right| dy \\
&\geq \frac{1}{\pi} \sum_{k=1}^{\frac{N}{2}} \frac{\int_{-\pi}^{\pi} |\sin(x)|}{-\pi\left(k+\frac{1}{2}\right)} \\
&\geq k \log(N) && (\text{For some } k > 0)
\end{aligned}
$$

Now consider the function $q_N(x) = \text{sign}(g_N(-x))$. In that case

$$S_N(q_N)(0) = \|g_N\|_1$$

And we can approximate q_N with a continuous function whose absolute value is bounded by 1 as closely as possible; this gives us a sequence of functions j_N such that

$$s_N(j_N)(0) \geq \|g_N\|_1 - 1$$
$$\|j_N\|_\infty \leq 1$$

But $S_N(\cdot)(0)$ is a continuous operator from the Banach space $C(T)$ with the uniform norm to the space $\mathbb{C}$. Since $\|g_N\|_1$ is unbounded, that means the operator norm of $S_N(\cdot)(0)$ is also not uniformly bounded. But from the uniform boundedness principle, if a continuous operator from a Banach space is not uniformly bounded, it's not pointwise bounded either for some point x in the domain. This means the sequence $\|S_N(f)(0)\|$ is not bounded for some $f \in C(T)$. For that continuous function f, the Fourier series diverges.

2   Weyl's equidistribution theorem

In the previous section, we saw how the idea of Fourier series emerged from a physical problem; the rest of the section was devoted to establishing firm foundations for the notion of Fourier series, especially those of convergence. With the foundational work done, one can look at the applications of the theory. In this and the subsequent section, we'll look at two theorems that use Fourier analysis; the surprising thing is that both these theorems are in number theory, quite distant from physics, where Fourier analysis first emerged. The fact that one can use tools that emerged in the analysis of a physical system for solving problems in number theory is a testament to the power and versatility of the tool.

In this section, we'll look at Weyl's equidistribution theorem, which informally states that a sequence of multiples of an irrational number modulo 1 is equidistributed in the interval $[0, 1]$.

2.1   Equidistributed sequences

We'll begin by formally defining what equidistributed sequences are.

Definition 2.1. A sequence $\{q_n\}$ is said to be equidistributed in the interval $[a, b]$ if

- $q_n \in [a, b]$ for all $n \in \mathbb{N}$.

- For all subintervals $[c, d]$ of $[a, b]$

$$\lim_{N \to \infty} \frac{1}{N} \sum_{k=1}^{N} \mathbb{1}_{[c,d]}(q_k) = \frac{d - c}{b - a}$$

where $\mathbb{1}_{[c,d]}$ is the indicator function for the interval $[c, d]$.

For now, let's just focus on the interval $[0, 2\pi]$, and look for equidistributed sequences in it. Observe that the second part of the above definition can be rewritten as

$$\lim_{N \to \infty} \frac{1}{N} \sum_{k=1}^{N} \mathbb{1}_{[c,d]}(q_k) = \frac{1}{2\pi} \int_{0}^{2\pi} \mathbb{1}_{[c,d]}(x)dx \tag{8}$$

2.2   Weyl's criterion for equidistribution

The above definition for equidistributed sequences makes sense (in the manner that agrees with our intuition), it is a little cumbersome to work with, since one needs to sum the indicator function of every subinterval of $[0, 2\pi]$ over the entire sequence to ensure it really is equidistributed. It would be nicer if one had a simpler, but equivalent notion for equidistribution that was easier to work with. It turns out there is one.

Lemma 2.1. *A subsequence $\{q_k\}$ of $[0, 2\pi]$ is equidistributed iff for every continuous function f on $[0, 2\pi]$*

$$\lim_{N \to \infty} \frac{1}{N} \sum_{k=1}^{N} f(q_k) = \frac{1}{2\pi} \int_{0}^{2\pi} f(x)dx$$

Proof. First we'll show that if the sequence is equidistributed then the above equation is correct. Consider *simple functions*[1] on $[0, 2\pi]$. Since these functions are linear combinations of indicator functions, and the indicator functions satisfy equation 8, the simple functions also satisfy that equation, i.e.

$$\lim_{N \to \infty} \frac{1}{N} \sum_{k=1}^{N} s(q_k) = \frac{1}{2\pi} \int_0^{2\pi} s(x)dx$$

where s is a simple function. Also, note the fact that in the compact interval $[0, 2\pi]$, for every $\varepsilon > 0$ and for every continuous function f, there exists a simple function s, such that $\|f - s\|_\infty < \varepsilon$.

Now we'll show that for every $\varepsilon > 0$, and for every continuous function f, there exists an $N \in \mathbb{N}$ such that for all $n > N$

$$\left| \frac{1}{n} \sum_{k=1}^{n} f(q_k) - \frac{1}{2\pi} \int_0^{2\pi} f(x)dx \right| < \varepsilon$$

Pick a simple function s such that $\|f - s\|_\infty < \frac{\varepsilon}{3}$. We'll have the following inequalities

$$\frac{1}{2\pi} \left| \int_0^{2\pi} f(x)dx - \int_0^{2\pi} s(x)dx \right| \leq \frac{1}{2\pi} \left| \int_0^{2\pi} (f(x) - s(x))dx \right| \tag{9}$$

$$\leq \frac{\varepsilon}{3} \tag{10}$$

And for all $n \in \mathbb{N}$

$$\left| \frac{1}{n} \sum_{k=1}^{n} f(q_k) - \frac{1}{n} \sum_{k=1}^{n} s(q_k) \right| \leq \left| \frac{1}{n} \sum_{k=1}^{n} f(q_k) - s(q_k) \right| \tag{11}$$

$$\leq \frac{\varepsilon}{3} \tag{12}$$

Now pick an N large enough such that for all $n > N$

$$\left| \frac{1}{n} \sum_{k=1}^{n} s(q_k) - \frac{1}{2\pi} \int_0^{2\pi} s(x)dx \right| < \frac{\varepsilon}{3} \tag{13}$$

Adding up inequalities 10, 12, and 13, we have the inequality we wanted.

Now for the converse statement. The key idea here is to approximate the indicator function $\mathbb{1}_{[c,d]}$ with an appropriate continuous function. For a given $\varepsilon > 0$, consider the following continuous approximation for the function $\mathbb{1}_{[c,d]}$

$$f(x) = \begin{cases} 1 & x \in (c, d) \\ \frac{8x - 8c + \varepsilon}{\varepsilon} & x \in \left[c - \frac{\varepsilon}{8}, c\right] \\ \frac{8d - 8x + \varepsilon}{\varepsilon} & x \in \left[d, d + \frac{\varepsilon}{8}\right] \\ 0 & \text{otherwise} \end{cases}$$

---

[1] Simple functions are functions whose range is finite. These functions can be written as linear combinations of indicator functions.

For all $x \in [0, 2\pi]$, we have the following inequality for the function f

$$\mathbb{1}_{[c,d]}(x) \leq f(x) \leq \mathbb{1}_{\left[c-\frac{\varepsilon}{8}, d+\frac{\varepsilon}{8}\right]}(x)$$

Rewriting the above inequality,

$$\left| f(x) - \mathbb{1}_{[c,d]}(x) \right| \leq \mathbb{1}_{\left[c-\frac{\varepsilon}{8}, c\right]}(x) + \mathbb{1}_{\left[d, d+\frac{\varepsilon}{8}\right]}(x)$$

This means there exists a large enough N_1, such that for all $n > N_1$

$$\frac{1}{n} \sum_{k=1}^{n} \left(\mathbb{1}_{\left[c-\frac{\varepsilon}{8}, c\right]}(q_k) + \mathbb{1}_{\left[d, d+\frac{\varepsilon}{8}\right]}(q_k) \right) \leq \frac{\varepsilon}{3}$$

For that same N_1,

$$\left| \frac{1}{n} \sum_{k=1}^{n} \mathbb{1}_{[c,d]}(q_k) - \frac{1}{n} \sum_{k=1}^{n} f(q_k) \right| \leq \frac{\varepsilon}{3} \tag{14}$$

Similarly,

$$\frac{1}{2\pi} \left| \int_0^{2\pi} f(x)dx - \int_0^{2\pi} \mathbb{1}_{[c,d]}(x)dx \right| \leq \frac{\varepsilon}{3} \tag{15}$$

Finally, because of our hypothesis, we have a large enough N_2 such that for all $n > N_2$,

$$\left| \frac{1}{n} \sum_{k=1}^{n} f(q_k) - \frac{1}{2\pi} \int_0^{2\pi} f(x)dx \right| \leq \frac{\varepsilon}{3} \tag{16}$$

Let $N = \max(N_1, N_2)$, and adding up inequalities 14, 15, and 16, we get our result. This completes the proof of equivalence. $\qquad\square$

This result shows how to equidistribution in terms of continuous functions. The next result will use Fejér's theorem to break up the result about continuous functions to exponential sums, giving us a much simpler formalism to work with.

Theorem 2.2. *A subsequence $\{q_k\}$ of $[0, 2\pi]$ is equidistributed iff for all non-zero integers a*

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \exp(iaq_k) = 0$$

Proof. First we'll show that if the sequence is equidistributed, then the limit is 0. Since $\exp(iax)$ is a continuous function for all a, from lemma 2.1, we have

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \exp(iaq_k) = \frac{1}{2\pi} \int_0^{2\pi} \exp(iax)dx$$

$$= 0 \qquad\qquad (a \text{ is non-zero})$$

12

For the converse, we will use Fejér's theorem. Given a continuous f and $\varepsilon > 0$, use Fejér's theorem to find an exponential polynomial p such that $\|p - f\|_\infty < \frac{\varepsilon}{3}$. In that case, we have

$$\frac{1}{2\pi}\left| \int_0^{2\pi} f(x)dx - \int_0^{2\pi} p(x)dx \right| < \frac{\varepsilon}{3} \tag{17}$$

and

$$\left| \frac{1}{n}\sum_{k=1}^n f(q_k) - \frac{1}{n}\sum_{k=1}^n p(q_k) \right| < \frac{\varepsilon}{3} \tag{18}$$

And because of our hypothesis, we have a large enough N, such that for all $n > N$

$$\left| \frac{1}{n}\sum_{k=1}^n p(q_k) - \int_0^{2\pi} p(x)dx \right| \leq \frac{\varepsilon}{3} \tag{19}$$

Adding up inequalities 17, 18, and 19, we get our result. $\qquad\square$

2.3 Weyl's equidistribution theorem: The weak version

The weak version of Weyl's equidistribution theorem states that the sequence

$$q_k = kz \mod 2\pi$$

is equidistributed in the interval $[0, 2\pi]$ if z is an irrational multiple of π. This can be shown really easily using the earlier formalisation, by looking at the exponential sums.

$$\sum_{k=1}^n \exp(iakz) = e^{iaz}\frac{1 - e^{ianz}}{1 - e^{iaz}}$$

Since z is an irrational multiple of π, the denominator is non-zero for all non-zero integers a. That means for a given a, the sum is bounded, hence

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^n \exp(iakz) = 0$$

and as a consequence, the sequence is equidistributed.

2.4 Weyl's equidistribution theorem: The stronger version

Note: Since it will get a little cumbersome to keep working with integral and irrational multiples of 2π, we will instead deal with sequences in the interval $[0, 1]$. The only change we'll need to keep in mind is that we'll be dealing the following sum

$$\sum_{k=1}^n \exp(2\pi iaq_k)$$

The stronger version of Weyl's equidistribution theorem states that the sequence

$$q_k = \alpha k^2 + \beta k + \gamma \mod 1$$

is equidistributed if α is irrational.

This will involve showing that the sequence

$$S_n = \sum_{k=1}^{n} \exp(2\pi i a q_k) \tag{20}$$

is $o(n)$ for all integers n, but since α is irrational, it suffices to show it for all irrational α. In a nutshell, we have reduced the problem of equidistribution to a problem of bounding the given sum.

2.4.1 A preliminary bound for another sum

Notation: We will define the function $e(x)$ to be $e(2\pi i x)$.

Lemma 2.3. *For all irrational θ*

$$\left| \sum_{k=1}^{N} e(n\theta) \right| \leq \min\left(N, \frac{1}{\|\theta\|} \right) \tag{21}$$

where $\|\theta\|$ is distance of θ to its closest integer.

Proof. By triangle inequality,

$$\left| \sum_{k=1}^{N} e(n\theta) \right| \leq \sum_{k=1}^{N} |e(n\theta)|$$
$$= N$$

And summing up the geometric series, we have

$$\left| \sum_{k=1}^{N} e(n\theta) \right| \leq \frac{2}{|1 - e(\theta)|}$$
$$= \frac{1}{|\sin(\pi\theta)|}$$
$$\leq \frac{1}{\|\theta\|}$$

$\square$

2.4.2 Weyl differencing

Since we want to show the sum S_N (see equation 20) is in $o(N)$, it will suffice to show that $|S_N|^2$ is in $o(N^2)$.

$$\left| \sum_{k=1}^{N} e(f(k)) \right|^2 = \left(\sum_{k=1}^{N} e(f(k)) \right) \left(\sum_{k=1}^{N} e(-f(k)) \right)$$

$$= \sum_{1 \le j,l \le N} e(f(j) - f(l))$$

$$= N + \sum_{d=1}^{N-1} \sum_{k=1}^{N-d} \left(e(f(k+d) - f(k)) + e(f(k) - f(k+d)) \right)$$

$$= N + \sum_{d=1}^{N-1} \sum_{k=1}^{N-d} 2\mathrm{Re} \left(e(f(k+d) - f(k)) \right)$$

$$= N + 2 \sum_{d=1}^{N-1} \mathrm{Re} \left(\sum_{k=1}^{N-d} e(f(k+d) - f(k)) \right)$$

$$\le N + 2 \sum_{d=1}^{N-1} \left| \sum_{k=1}^{N-d} e(f(k+d) - f(k)) \right|$$

Plugging in the quadratic polynomial in f, we get

$$\left| \sum_{k=1}^{N} e(q_k) \right|^2 \le N + 2 \sum_{d=1}^{N-1} \left| \sum_{k=1}^{N-d} e((2\alpha d)k) \right|$$

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

Since N is in $o(N^2)$, all we need to show is that $\sum_{d=1}^{N} \min \left(N, \frac{1}{\|2\alpha d\|} \right)$ is also in $o(N^2)$.

2.4.3 Rational approximation of irrational numbers

The main result of this section will show that given an irrational number α, it's possible to find infinitely many rational numbers $\frac{p}{q}$ ($\gcd(p,q) = 1$) such that $|\alpha - \frac{p}{q}| < \frac{1}{q^2}$. We will need this result to bound the main sum.

Proposition 2.4. *Given an irrational number α and $N \in \mathbb{N}$, there exists a rational number $\frac{p}{q}$ such that*

$$\left| \alpha - \frac{p}{q} \right| < \frac{1}{q(N+1)}$$

and $|q| \le N$.

Proof. Consider the following sequence modulo 1:

$$\alpha, 2\alpha, 3\alpha, \ldots, N\alpha$$

Divide $[0, 1]$ into $N + 1$ equally sized intervals. Clearly, one of the elements of that sequence must either lie in the subinterval $\left[0, \frac{1}{N+1}\right]$, or $\left[\frac{N}{N+1}, 1\right]$. Call that element $q\alpha$. In that case, the following inequality holds:

$$\|q\alpha\| \leq \frac{1}{N + 1}$$

Let the integer closest to $q\alpha$ be p. In that case

$$|p - q\alpha| \leq \frac{1}{N + 1}$$
$$\left|\frac{p}{q} - \alpha\right| \leq \frac{1}{q(N + 1)}$$

And since $q < N + 1$

$$\left|\alpha - \frac{p}{q}\right| < \frac{1}{q^2}$$

$\square$

Corollary 2.5. *If α is irrational, there are infinitely many rational numbers $\frac{p}{q}$ such that*

$$\left|\alpha - \frac{p}{q}\right| < \frac{1}{q^2}$$

Proof. Suppose there were finitely many. Take the minimal distance d of α with the rational approximations. Pick a large enough N such that $\frac{1}{N+1} < d$. Use that N in proposition 2.4 and get a contradiction. $\square$

2.4.4 Bounding the complete sum[4]

At the end of section 2.4.2, we saw that all we need to do is to show that the following sum

$$W_N = \sum_{d=1}^{N} \min\left(N, \frac{1}{\|2\alpha d\|}\right)$$

is in $o(N^2)$.

Lemma 2.6. *If $\alpha_1, \alpha_2, \ldots, \alpha_N$ are real numbers such that $\|\alpha_i - \alpha_j\| \geq \frac{1}{r}$ for some natural number r when $i \neq j$. Then*

$$\sum_{i=1}^{N} \min\left(\frac{1}{\|\alpha_i\|}, N\right) \leq 2N + 2r(\log(N) + 2)$$

Proof. Without loss of generality, assume all the α_i lie within $\left[-\frac{1}{2}, \frac{1}{2}\right]$. Furthermore, at least half the sum will either come from the positive or the negative α_i. WLOG, assume it's the positive α_i. Then we can just double the sum to get an upper bound. Assume that $\alpha_1 < \alpha_2 < \cdots < \alpha_n$, where α_1 to α_n are the positive α_i. In that case

$$\sum_{i=1}^{n} \min\left(\frac{1}{\|\alpha_i\|}, N\right) \leq N + \left\lfloor \frac{r}{N} \right\rfloor N + \sum_{i > \left\lfloor \frac{r}{N} \right\rfloor} \frac{r}{i} \qquad \left(\text{Because } \|\alpha_i - \alpha_j\| > \frac{1}{r}\right)$$

The last term can have at most n terms, hence it is bounded by $r(\log(n) + 1)$. But n is smaller than or equal to N, since n is the number of positive α_i. Hence the last term is bounded by $r(\log(N) + 1)$. The second term is bounded by r, and the first term is N. Hence the complete bound is

$$\sum_{i=1}^{n} \min\left(\frac{1}{\|\alpha_i\|}, N\right) \leq N + r + r(\log(n) + 1)$$

Since this is at least half the total sum, we get our bound. $\qquad\square$

Lemma 2.7. *If $\frac{p}{q}$ is a rational approximation for α such that $\left|\alpha - \frac{p}{q}\right| < \frac{1}{q^2}$, then*

$$\sum_{d=1}^{N} \min\left(\frac{1}{\|\alpha d\|}, N\right) \leq \left(\frac{2N}{q}\right)(2N + 4q(\log(N) + 2))$$

Proof. We have the following inequality, which is not too hard to prove:

$$\|i\alpha - j\alpha\| \geq \left\|(i-j)\frac{p}{q}\right\| - \frac{|i-j|}{q^2} \qquad \left(\text{Taylor expand around } \frac{p}{q}\right)$$

When $|i - j| \leq \frac{q}{2}$, then $\frac{(i-j)p}{q}$ is not an integer since p and q are co-prime. That means

$$\left\|(i-j)\frac{p}{q}\right\| \geq \frac{1}{q}$$

and

$$\frac{|i-j|}{q^2} \leq \frac{1}{2q}$$

which means

$$\|i\alpha - j\alpha\| \geq \frac{1}{2q}$$

Now we split up the range of summation, i.e. the interval $[1, N]$ into intervals of length $\frac{q}{2}$. To compute the sum over each of these intervals, we'll use lemma 2.6 with $r = 2q$. We'll have $\frac{2N}{q}$ such intervals, over each of which, the sum will be be bounded by $(2N + 4q(\log(N) + 2))$. The complete bound is hence

$$W_N \leq \left(\frac{2N}{q}\right)(2N + 4q(\log(N) + 2))$$

$\qquad\square$

As N goes to infinity, the $\frac{W_N}{N^2}$ becomes $\frac{2}{q}$. But since α is irrational, q can be as large as possible. This shows that W_N is in $o(N^2)$.

3   Roth's theorem

The fact that the integral

$$\int_{-\pi}^{\pi} e^{inx} dx$$

is 1 only when n is 0 and for other integral values of n, it is 0, can be very useful. It is this very observation that forms the backbone of a general class of techniques in number theory called the *circle method*. In a nutshell, this method lets one *count*, by integrating an appropriately picked function over the unit circle, hence the name.

Roth's theorem is question about existence of three term APs, but like many existence problems, it can be resolved by counting, and making sure your count doesn't end at 0. Formally, the theorem states that given a number $\delta > 0$, called the density, there exists a large enough N such that any subset of $\{0, 1, 2, 3, \ldots, N-1\}$ with a size greater than δN contains a three term arithmetic progression. The key idea in the proof is to count the number of triples x, y, z such that $x + z - 2y$ is 0. And this is where the circle method comes in, transforming a combinatorial problem into a problem in analysis.

3.1   Sketch of proof

The key idea in the proof of Roth's theorem is that for a given $0 < \delta < 1$, and a large enough N, if a subset A of $[N]^2$, which has size δN does not have a three term AP, then there exists a progression B_1 in $[N]$ such that the density of A in B_1 is greater than δ. Clearly, $A \cap B_1$ won't contain a three term AP either, so one iterates the argument again and again to get a sequence of nested sub progressions $B_1, B_2, \ldots B_k$ such that the density of A in B_k is at least 1. But that would mean $B_k \subset A$, and if $|B_k| \geq 3$, we'll have a contradiction. So one picks a large enough N such that $|B_k| \geq 3$, and that leads to the contradiction. Which means A must have had a three term AP.

3.2   Fourier analysis on finite cyclic groups

Let's begin by endowing the set $[N]$ some additional structure, namely identifying it with the group $\mathbb{Z}/N\mathbb{Z}$. Furthermore, with the counting measure $\mathbb{Z}/N\mathbb{Z}$, we can integrate functions from $\mathbb{Z}/N\mathbb{Z}$ to $\mathbb{C}$. And with the discrete metric, we have the topology on the space; we can finally do some analysis.

Consider the set of functions from $\mathbb{Z}/N\mathbb{Z}$ to $\mathbb{C}$. This clearly forms a Hilbert space, with the inner product being

$$\langle f, g \rangle = \frac{1}{N} \int_{\mathbb{Z}/N\mathbb{Z}} f(x) \overline{g(x)} dx$$
$$= \frac{1}{N} \sum_{k=0}^{N-1} f(k) \overline{g(k)}$$

The next step would be to determine an orthonormal basis for the Hilbert space. It turns out that the set of homomorphisms from $\mathbb{Z}/N\mathbb{Z}$ to $\mathbb{C}$ does form an orthonormal

---

[2] $[N]$ is shorthand for the set $\{0, 1, 2, 3 \ldots, N-1\}$

basis for the space. Define h_r to be the following homomorphism

$$h_r(k) = e^{\frac{2\pi i}{N}rk}$$

Notice that

$$\frac{1}{N} \int_{\mathbb{Z}/N\mathbb{Z}} h_r(x)dx = \begin{cases} 1 & \text{if } r = 0 \\ 0 & \text{if } r \neq 0 \end{cases}$$

The rth Fourier coefficient of f is defined as

$$\widehat{f}(r) = \langle f, h_r \rangle$$
$$= \frac{1}{N} \sum_{k=0}^{N-1} f(k)e^{-\frac{2\pi i}{N}rk}$$

Hence, for any function f

$$f(x) = \widehat{f}(0)h_0(x) + \widehat{f}(1)h_1(x) + \ldots + \widehat{f}(N-1)h_{N-1}(x)$$

3.3   Counting progressions in $[N][3]$

Note: We'll call (x, y, z) a progression in $\mathbb{Z}/N\mathbb{Z}$ when $x + z = 2y$ in the group $\mathbb{Z}/N\mathbb{Z}$, and we'll call it a progression in $[N]$ when $x + y = 2z$ in $\mathbb{Z}$.

Consider a subset A of $[N]$ of size δN. Identify A as a subset of $\mathbb{Z}/N\mathbb{Z}$ as well. We are looking for three term APs, or triples (x, y, z) such that $x + z - 2y = 0$ in $\mathbb{Z}/N\mathbb{Z}$. The number of such triples will be given by

$$S_0 = \sum_{x,y,z \in A} \frac{1}{N} \int_{\mathbb{Z}/N\mathbb{Z}} h_{x+z-2y}(x)dx$$
$$= \sum_{x,y,z \in [N]} \frac{1}{N} \sum_{k=0}^{N-1} \mathbb{1}_A(x)\mathbb{1}_A(y)\mathbb{1}_A(z)e^{-\frac{2\pi i}{N}k(x+z-2y)}$$
$$= N^2 \sum_{k=0}^{N-1} \widehat{\mathbb{1}_A}(k)^2 \widehat{\mathbb{1}_A}(-2k)$$

where $\mathbb{1}_A$ is the indicator function of A. This counts the number of three term APs in the group $\mathbb{Z}/N\mathbb{Z}$. But all the progressions in $\mathbb{Z}/N\mathbb{Z}$ need not be progressions in $[N]$. But if we know that x and y belong to $\left[\frac{N}{3}, \frac{2N}{3}\right]$, and $x + z - 2y = 0$ in the group, then x, y, z form an AP in $[N]$. Define the set M_A to be $\left[\frac{N}{3}, \frac{2N}{3}\right] \cap A$. Then, a lower bound

for the number of three term APs in $[N]$ is given by

$$S = \sum_{x,y \in M_A} \sum_{z \in A} \frac{1}{N} \int_{\mathbb{Z}/N\mathbb{Z}} h_{x+z-2y}(x)dx$$

$$= N^2 \sum_{k=0}^{N-1} \widehat{\mathbb{1}_A}(k) \widehat{\mathbb{1}_{M_A}}(k) \widehat{\mathbb{1}_{M_A}}(-2k)$$

$$= N^2 \left(\widehat{\mathbb{1}_A}(0) \widehat{\mathbb{1}_{M_A}}(0) \widehat{\mathbb{1}_{M_A}}(0) + \sum_{k=1}^{N-1} \widehat{\mathbb{1}_A}(k) \widehat{\mathbb{1}_{M_A}}(k) \widehat{\mathbb{1}_{M_A}}(-2k) \right)$$

$$= \delta |M_A|^2 + N^2 \sum_{k=1}^{N-1} \widehat{\mathbb{1}_A}(k) \widehat{\mathbb{1}_{M_A}}(k) \widehat{\mathbb{1}_{M_A}}(-2k)$$

We want to show that $[N]$ contains at least one three term AP. But the expression for S also counts triples like (x, x, x), which we don't want to count as APs. A contains $|A|$ such triples. Hence we'd like S to be greater than $|A|$ to show the existence of an AP. We have the following lemma

Lemma 3.1. *If N is odd, $\widehat{\mathbb{1}_A}(k) < \varepsilon$ for all $k \neq 0$, where $\varepsilon = \frac{\delta^2}{8}$, and $|M_A| \geq \frac{\delta N}{4}$, then $S \geq \frac{\delta^3 N^2}{32}$.*

Proof. We know that

$$S = \delta |M_A|^2 + N^2 \sum_{k=1}^{N-1} \widehat{\mathbb{1}_A}(k) \widehat{\mathbb{1}_{M_A}}(k) \widehat{\mathbb{1}_{M_A}}(-2k)$$

We can bound the second term in the sum using Cauchy-Schwarz inequality and Plancherel's identity.

$$\left| \sum_{k=1}^{N-1} \widehat{\mathbb{1}_A}(k) \widehat{\mathbb{1}_{M_A}}(k) \widehat{\mathbb{1}_{M_A}}(-2k) \right| \leq \varepsilon \sum_{k=1}^{N-1} \left| \widehat{\mathbb{1}_{M_A}}(k) \widehat{\mathbb{1}_{M_A}}(-2k) \right|$$

$$\leq \varepsilon \left(\sum_{k=1}^{N-1} \left| \widehat{\mathbb{1}_A}(k) \right|^2 \right)^{\frac{1}{2}} \left(\sum_{k=1}^{N-1} \left| \widehat{\mathbb{1}_A}(-2k) \right|^2 \right)^{\frac{1}{2}}$$

$$= \varepsilon \left(\sum_{k=1}^{N-1} \left| \widehat{\mathbb{1}_A}(k) \right|^2 \right)$$

$$= \varepsilon \left(\sum_{k=1}^{N-1} |\mathbb{1}_A(k)|^2 \right)$$

$$= \varepsilon |M_A|$$

From this, we get that

$$S \geq \delta |M_A|^2 - \varepsilon N^2 |M_A|$$

Since $|M_A| \geq \frac{\delta N}{4}$, and $\varepsilon = \frac{\delta^2}{8}$, we get

$$S \geq \frac{\delta^3 N^2}{32}$$

$\square$

This lemma shows that if we pick an $N > \frac{32}{\delta^2}$, any subset A which satisfies the hypotheses of the lemma will contain a 3-AP.

3.4   Density incrementation

Now consider the contrapositive of lemma 3.1. It says that if a subset A of $[N]$ does not contain a 3-AP, then one of the following conditions must hold

1. For some non-zero k, $\widehat{\mathbb{1}_A}(k) > \varepsilon$, where $\varepsilon = \frac{\delta^2}{8}$.

2. $|M_A| < \frac{\delta N}{4}$.

We don't consider the third condition where N could be less than $\frac{32}{\delta^2}$ because we can make N as large as we want. Consider the second condition. If the density of A in $\left[\frac{N}{3}, \frac{2N}{3}\right]$ is less than $\frac{\delta N}{4}$, then by the pigeonhole principle, A has a density greater than $\delta + \frac{\delta}{8}$ in either $\left[0, \frac{N}{3}\right]$ or $\left[\frac{2N}{3}, N\right]$. To put it more generally, there exists an AP Z in $[N]$ of length greater than or equal to $\frac{N}{3}$ such that the density of A in Z is $\delta + \frac{\delta}{8}$, which is greater than the original density. Specifically, the progression Z in this case is either the interval $\left[0, \frac{N}{3}\right]$ or $\left[\frac{2N}{3}, N\right]$.

We will show even when $\widehat{\mathbb{1}_A}(k) > \varepsilon$ for some non-zero k, there exists a sufficiently long progression in $[N]$ such that the density of A in that subprogression is greater than δ.

Lemma 3.2. *If $\widehat{\mathbb{1}_A}(r) > \gamma$ for some $r \neq 0$, then there exists a $\mathbb{Z}/N\mathbb{Z}$ subprogression B of length at least $\frac{\sqrt{N}}{8}$ such that $|A \cap B| \geq \left(\delta + \frac{\gamma}{4}\right)|B|$.*

Proof. Consider the pairs of points

$$(0,0), (1,r), (2,2r), \ldots, (N-1, (N-1)r)$$

These points lie in the square $[0, N] \times [0, N]$. Divide this square into $\lfloor\sqrt{N}\rfloor^2$ squares (where $\lfloor x \rfloor$ is the greatest integer *less* than x) of side length $l = \frac{N}{\lfloor\sqrt{N}\rfloor}$. Since there are less than N squares, two of the points must lie in the same square. That means for some $d \leq l$

$$rd \leq l \pmod{N}$$

Let B' be an AP of length $\left\lfloor\frac{\lfloor N \rfloor}{2\pi}\right\rfloor$ in $\mathbb{Z}/N\mathbb{Z}$

$$\ldots, -3d, -2d, -d, 0, d, 2d, 3d, \ldots$$

Since $|B'|d$ is less than $|B'|l$, which in turn is less than $\frac{N}{2\pi}$, which means that it can be written as a union of two disjoint APs in $[N]$.

Consider the following inequality:

$$\left|N\widehat{\mathbb{1}}_{B'}(r) - |B'|\right| = \left|\sum_{\mathbb{Z}/N\mathbb{Z}} \mathbb{1}_{B'}(x)\left(e^{-\frac{2\pi i}{N}rx} - 1\right)\right|$$

$$= \left|\sum_{|x|\leq\frac{1}{2}|B'|}\left(e^{-\frac{2\pi i}{N}rdx} - 1\right)\right|$$

$$\leq \sum_{|x|\leq\frac{1}{2}|B'|}\left|e^{-\frac{2\pi i}{N}rdx} - 1\right|$$

$$\leq 2\sum_{x=0}^{\frac{|B'|}{2}}\frac{2\pi}{N}rdx$$

$$\leq \frac{4\pi l}{N}\sum_{x=0}^{\frac{|B'|}{2}}x$$

$$\leq |B'|\frac{|B'|\pi l}{N}$$

$$\leq \frac{|B'|}{2}$$

This means

$$\left|N\widehat{\mathbb{1}}_{B}(r)\right| \geq \frac{|B'|}{2}$$

The required progression, in which the density of A will be greater than $\left(\delta + \frac{\gamma}{4}\right)$, will be a translation of B', i.e. $B = B' + c$ for some constant c.

Define f_A to be the balanced indicator of A, i.e.

$$f_A(k) = \mathbb{1}_A(k) - \delta$$

Notice that the mean of f_A over $[N]$ is 0. Furthermore, if

$$\sum_{k=0}^{N-1} f_A(k)\mathbb{1}_B(k) \geq \kappa|B|$$

then

$$|A \cap B| \geq (\delta + \kappa)|B|$$

and vice versa. Here's why.

$$\sum_{k=0}^{N-1} f_A(k)\mathbb{1}_B(k) = \sum_{k\in B} f_A(k)$$

$$= (1 - \delta)|A \cap B| - \delta(|B| - |A \cap B|)$$

$$= |A \cap B| - \delta|B|$$

$|A \cap B| - \delta|B|$ will be greater than $\kappa|B|$ iff $|A \cap B| > (\delta + \kappa)|B|$. Now our problem reduces to determining for what c, $\sum_{k=0}^{N-1} f_A(k)\mathbb{1}_{B'}(k - c)$ is greater than or equal to $\frac{\gamma}{4}|B'|$. Define $G(c)$ to be

$$G(c) = \sum_{k=0}^{N-1} f_A(k)\mathbb{1}_{B'}(k - c)$$

The Fourier transform of G is the following:

$$\widehat{G}(r) = N\widehat{f_A}(r)\widehat{\mathbb{1}_{B'}}(r)$$

And we know that

$$\sum_{[N]} |G(c)| \geq \left|\widehat{G}(r)\right| \geq \frac{1}{2}\gamma N|B'|$$

And since the mean of G over $[N]$ is 0

$$\sum_{[N]} G(c) + |G(c)| \geq \frac{1}{2}\gamma N|B'|$$

For some c_0

$$G(c_0) + |G(c_0)| \geq \frac{1}{2}\gamma|B'|$$

Hence, $G(c_0) \geq \frac{1}{4}\gamma$

The progression $B = B' + c_0$ is the one that satisfies our requirements. $\qquad\square$

Lemma 3.3. *If $[N]$ has a $\mathbb{Z}/N\mathbb{Z}$ progression B of length $\frac{\sqrt{N}}{8}$, such that the density of A in B is greater than or equal to $\delta + \frac{\gamma}{4}$, and it is a union of two disjoint $[N]$ progressions, then there exists a progression P in $[N]$ such that $|P| \geq \frac{\gamma}{8}|B|$ such that A has density $\delta + \frac{\gamma}{8}$ in P.*

Proof. Write B as $P_1 \sqcup P_2$, where P_1 and P_2 are disjoint $[N]$ progressions, and $|P_1| \leq |P_2|$. If $|P_1| \leq \frac{\gamma}{8}|B|$, then

$$\begin{aligned}
|A \cap P_2| &\geq |A \cap B| - |P_1| \\
&\geq \left(\delta + \frac{\gamma}{4}\right)|B| - P_1 \\
&\geq \left(\delta + \frac{\gamma}{8}\right)|B| \\
&\geq \left(\delta + \frac{\gamma}{8}\right)|P_2|
\end{aligned}$$

One the other hand, if $|P_1| > \frac{\gamma}{8}$, then one of A must have density $\delta + \frac{\gamma}{4}$ on one of P_1 or P_2. $\qquad\square$

With the previous lemmas, we can finally state in full the density incrementation theorem.

Theorem 3.4. *Given $0 < \delta < 1$, and $N > \frac{32}{\delta^2}$, and a subset A of $[N]$ of size δN, if for some non-zero k, $\widehat{\mathbb{1}_A}(k) > \frac{\delta^2}{8}$ or $\left|\left[\frac{N}{3}, \frac{2N}{3}\right] \cap A\right| < \frac{\delta N}{4}$, then there exists a progression in $[N]$ of length at least $\frac{\delta^2\sqrt{N}}{512}$ such that the density of A in the progression is at least $\delta + \frac{\delta^2}{64}$.*

Proof. Follows from lemmas 3.2 and 3.3. $\qquad\square$

3.5   Iterating the density incrementation argument

The final step in the proof of Roth's theorem is to increment the density until one reaches a sub progression in which A has density 1. If the length of that sub progression is greater than 3, then we're done. After each iteration, the density grows by at least $\frac{\delta^2}{64}$, hence after $k = \frac{64}{\delta^2}(1 - \delta)$ steps, the density will be at least 1. We just need to pick a large enough N such that the sub progression after k steps has at least 3 elements. We want

$$\frac{\delta^{2k} N^{\frac{1}{2^k}}}{512^k} \geq 3$$

$$N \geq 3^{2^k} \left(\frac{512}{\delta^2}\right)^{2^k \cdot k}$$

Substituting the given value of k, we get our lower bound on N, and that completes the proof.

References

[1] *Joni's math notes: Divergence of fourier series*, `http://jonismathnotes.blogspot.in/2014/07/divergence-of-fourier-series.html`.

[2] Rajendra Bhatia, *Fourier series*, Mathematical Association of America, 2005.

[3] Mustazee Rahman, *Roth's theorem on 3-term arithmetic progressions*.

[4] Jacques Verstraëte, *Notes on equidistribution*.