APPLICATIONS OF COMMUTATIVE HARMONIC ANALYSIS

TOM SANDERS

1. Introduction

In this set of notes we shall investigate applications of commutative harmonic analysis, which for us will mean the Fourier transform on commutative groups. The focus will be on examples, and these will come from a range of settings including functional analysis, number theory, and probability.

Most of what we do will focus on quantitative estimates related to finite groups, and in light of this there is no one specific reference to be recommended. There are a number which give a flavour of different aspects of what we are interested in, and we shall try give references as we proceed.

Before we introduce the Fourier transform it is worth providing some motivation for its definition. For us this motivation will come from convolution. Convolution is something which will have come up in a variety of settings and is an exceptionally useful mathematical tool. We shall begin with some examples and then extend it to the more general setting in which we are interested.

Definition 1.1 (Convolution on $\mathbb{Z}$)**.** Given two functions $f, g \in \ell^1(\mathbb{Z})$, their *convolution* is the function $f * g$ defined by

$$f * g(x) := \sum_{y \in \mathbb{Z}} f(y)g(x-y) \text{ for all } x \in \mathbb{Z}.$$

It is immediate from the triangle inequality that this is well-defined in the sense that the left hand side is finite.

1.2. **Convolution and sumsets.** There are various reasons that convolution on the integers is useful. One is that it can be used to study sumsets: given two sets of integers A and B we write $A + B$ for their *sumset* defined by

$$A + B := \{a + b : a \in A, b \in B\}.$$

A number of questions in number theory can be phrased in terms of sumsets.

Example 1.3 (Goldbach's conjecture and Lagrange's theorem)**.** If we write P for the set of prime numbers then Goldbach's conjecture is simply the statement that

$$P + P \supset 2\mathbb{N}\backslash\{2\}.$$

Last updated: 27th August, 2013.

Similarly, if we write Q for the set of (non-negative) squares then Lagrange's theorem is simply the statement that

$$Q+Q+Q+Q=\mathbb{N}_0.$$

Convolution can be used to help understand sumsets (and so called iterated sumsets, $A+A+A$, &c.) because of the relationship

$$A+B=\operatorname{supp} 1_A * 1_B, \tag{1.1}$$

where 1_A and 1_B are the indicator functions on A and B respectively. Roughly, if we want to show that some element is contained in $A+B$, then it is sufficient to show that $1_A * 1_B$ is non-zero on this element.

1.4. **Convolution and representation functions.** Since sets are particularly important to us, it is also instructive to think about what the convolution of their indicator functions is in words. Suppose that A and B are finite sets of integers. Then

$$\begin{aligned} 1_A * 1_B(x) &= |\{(a,b)\in A\times B : a+b=x\}| \\ &= \text{`The number of ways of writing } x=a+b \text{ with } a\in A, b\in B.\text{'} \end{aligned}$$

We shall be interested in showing the existence of certain structures in sets or sumsets. As an example of a structure we might be interested in, a quadruple of integers (x_1,x_2,x_3,x_4) is called an *additive quadruple* if

$$x_1+x_2=x_3+x_4.$$

One of the basic ways to show that a such a structure is present in a set is to try to count copies of it, and the next proposition does exactly that.

Proposition 1.5. *Suppose that $A\subset\{1,\dots,N\}$ is such that there are no four distinct elements $x_1,x_2,x_3,x_4\in A$ forming an additive quadruple. Then*

$$|A|=O(N^{1/2}).$$

Proof. First note that if $x=(x_1,x_2,x_3,x_4)$ is an additive quadruple in which at least two elements are the same then either $x_1=x_2$, $x_1=x_3$, $x_1=x_4$, or $x_3=x_4$. (Note that if $x_1=x_3$ then $x_2=x_4$ since x is an additive quadruple, and similarly if $x_1=x_4$.) There are therefore at most $4|A|^2$ such quadruples in A.

On the other hand there is an exact formula for Q, the number of additive quadruples in A, in terms of convolution. As this is the first time we are doing such a calculation we shall go through it in somewhat more detail than we shall in future.

$$\begin{aligned} Q &= \sum_{x_1+x_2=x_3+x_4} 1_A(x_1)1_A(x_2)1_A(x_3)1_A(x_4) \\ &= \sum_{y\in\mathbb{Z}}\left(\sum_{x_1+x_2=y} 1_A(x_1)1_A(x_2)\right)\left(\sum_{x_3+x_4=y} 1_A(x_3)1_A(x_4)\right) \\ &= \sum_{y\in\mathbb{Z}}\left(\sum_{x+z=y} 1_A(x)1_A(z)\right)^2 = \sum_{y\in\mathbb{Z}} 1_A*1_A(y)^2 = \|1_A*1_A\|_{\ell^2(\mathbb{Z})}^2. \end{aligned}$$

At this point we want a lower bound on the right hand side and to get this we apply the Cauchy-Schwarz inequality[1]:

$$|A+A|\|1_A*1_A\|_{\ell^2(\mathbb{Z})}^2 = \|1_{A+A}\|_{\ell^2(\mathbb{Z})}^2\|1_A*1_A\|_{\ell^2(\mathbb{Z})}^2 \geqslant \|1_A*1_A\|_{\ell^1(\mathbb{Z})}^2 = |A|^4.$$

Since $A\subset\{1,\dots,N\}$ we have that

$$|A+A| \leqslant |\{1,\dots,N\}+\{1,\dots,N\}| = 2N-1,$$

and so

$$(2N-1).(4|A|^2) \geqslant Q \geqslant |A|^4,$$

and the proposition follows. □

Up to the implied constant this proposition is best possible, but it is worth thinking about what happens if we replace the equation $x_1+x_2=x_3+x_4$ by $x_1+x_2=2x_3$. That is to say, we ask what happens if $A\subset\{1,\dots,N\}$ contains no three-term arithmetic progressions with all elements distinct. It is still true that $|A|=o(N)$ but it is much harder to show. We shall investigate this question more thoroughly later in the notes.

Returning to convolution in general, it has also come up in another setting: on the reals.

Definition 1.6 (Convolution on $\mathbb{R}$)**.** Given two functions $f,g\in L^1(\mathbb{R})$, their *convolution* is the function $f*g$ defined by

$$f*g(x) := \int f(y)g(x-y)dy \text{ for almost all } x\in\mathbb{R}.$$

Here, of course, the definition is almost everywhere rather than point-wise and this follows from a special case of Young's inequality: by Tonelli's theorem we have that

$$\begin{aligned}\|f*g\|_{L^1(\mathbb{R})} = \int |f*g(x)|dx &\leqslant \int\int |f(y)g(x-y)|dydx\\ &= \int |f(y)|\int |g(x-y)|dxdy = \|f\|_{L^1(\mathbb{R})}\|g\|_{L^1(\mathbb{R})}.\end{aligned}$$

In light of this $L^1(\mathbb{R})$ forms an associative (commutative) algebra[2] under convolution.

We happen to be taking care here by applying Tonelli's theorem. The notes will move to make things largely finite quite soon at which point these concerns will take a different form.

[1] Inequalities should only be applied when we expect to be close to the case of equality; here we are in this situation. We are applying Cauchy-Schwarz to the inner product of 1_{A+A} with 1_A*1_A to get

$$\langle 1_{A+A},1_A*1_A\rangle_{\ell^2(\mathbb{Z})}^2 \leqslant \|1_{A+A}\|_{\ell^2(\mathbb{Z})}^2\|1_A*1_A\|_{\ell^2(\mathbb{Z})}^2,$$

and this is close to equality when 1_{A+A} is 'close to' a (scalar multiple of) 1_A*1_A. Since A has no additive quadruples with all element distinct one sees that $1_A*1_A(x)$ only takes the values 0, 1, 2 or 3. This means that we have the inequality

$$1_{A+A} \leqslant 1_A*1_A \leqslant 3.1_{A+A},$$

which shows that 1_{A+A} is 'close to' 1_A*1_A as desired, and hence that our application of Cauchy-Schwarz was sensible.

[2] Recall that an *associative algebra* is just a ring whose additive group is a vector space over a field $\mathbb{K}$ and for which the ring multiplication is $\mathbb{K}$-bilinear.

It may be instructive to have a concrete example of convolution in action.

Example 1.7 (Convolution of intervals)**.** Suppose that $L \geqslant 1$, and consider the convolution of $1_{[0,L]}$ and $1_{[0,1]}$. We get that

$$1_{[0,L]} * 1_{[0,1]}(x) = \mu([0,L] \cap (x - [0,1])) = \begin{cases} x & \text{if } 0 \leqslant x \leqslant 1 \\ 1 & \text{if } 1 \leqslant x \leqslant L \\ L+1-x & \text{if } L \leqslant x \leqslant L+1 \\ 0 & \text{otherwise.} \end{cases}$$

It is probably most helpful to draw this. A useful quick check on the particular parameters can be made by integrating the function and verifying the identity

$$\int f * g(x)dx = \int f(y)dy \int g(z)dz.$$

In this case the right hand side is $\mu([0,1])\mu([0,L]) = L$; the left can be checked from the picture.

At this point we should properly introduce measures on $\mathbb{R}$. Were we to do this we should proceed via the Riesz representation theorem but since we shall not need much of the theory we shall content ourselves with probability density functions.

Definition 1.8 (Probability density functions on $\mathbb{R}$)**.** Given an open bounded set $S \subset \mathbb{R}$ we write f_S for the function $\mu(S)^{-1}1_S$; this is the uniform probability density function supported on S.

1.9. **Convolution and smoothing.** Convolving has the effect of 'smoothing' functions on $\mathbb{R}$. A small illustration of this can be seen in Example 1.7 where the convolution of two step functions produced a continuous function. It turns out that more is true and that each time we convolve we 'increase the differentiability class' of a function.

Recall that $C^k(\mathbb{R})$ is the space of k-times continuously differentiable complex-valued functions on $\mathbb{R}$; in particular $C^0(\mathbb{R}) = C(\mathbb{R})$, the space of continuous complex-valued functions on $\mathbb{R}$.

The following proposition is just a simple application of (the proof of) the Fundamental Theorem of Calculus.

Proposition 1.10. *Suppose that $g \in C^k(\mathbb{R}) \cap L^1(\mathbb{R})$ and $I \subset \mathbb{R}$ is an open bounded interval. Then $g * f_I \in C^{k+1}(\mathbb{R}) \cap L^1(\mathbb{R})$.*

Proof. The fact that $g * f_I \in L^1(\mathbb{R})$ follows from the trivial instance of Young's inequality proved in Definition 1.6, so the content of the proposition is in establishing the differentiability.

We put $I = (a, b)$ for some $a, b \in \mathbb{R}$ and note that[3]

$$\begin{aligned} g * f_I(x+h) &= \frac{1}{\mu(I)} \int g(y) 1_I(x+h-y) dy \\ &= \frac{1}{\mu(I)} \int g(y)(1_I(x+h-y) - 1_I(x-y)) dy + g * f_I(x) \\ &= \frac{1}{\mu(I)} \int_{x-a}^{x+h-a} g(y) dy - \frac{1}{\mu(I)} \int_{x-b}^{x+h-b} g(y) dy + g * f_I(x). \end{aligned}$$

Since g is continuous, for any $z \in \mathbb{R}$ we have[4]

$$\int_z^{z+h} g(y) dy = h(g(z) + o_{g,z;h \to 0}(1)),$$

and it follows that

$$(g * f_I)'(x) = \frac{1}{\mu(I)}(g(x-a) - g(x-b)).$$

Since $C^k(\mathbb{R})$ is closed under translation, addition and scalar multiplication we conclude that the right hand side is an element of $C^k(\mathbb{R})$ and the result follows. □

As with much of what we do in these notes, the above proposition has not been proved in maximal generality. In fact we shall tend to prove results in the least generality illustrating the main idea(s).

As we have already seen (in, for example, Definition 1.1) we can convolve in settings where there is no non-trivial differential structure and in light of the above proposition we shall import the intuitive notion of smoothness into other settings through convolution.

One application of Proposition 1.10 is to show that convolutions can be used to make bump functions. Bump functions are incredibly useful so it is worth a short detour to construct them.

Example 1.11 (Bump functions)**.** A *bump function* is a non-zero complex-valued, compactly supported, infinitely differentiable function on $\mathbb{R}$. The existence of bump functions is not immediately obvious and often they are constructed by the introduction of an auxiliary function such as $\exp(-1/x)1_{[0,\infty)}(x)$. It turns out that we can use convolution to do this quite naturally.

The basic idea is that by Proposition 1.10 each time we convolve with an interval we increase the differentiability class of a function, so we should like to convolve an 'infinite' number of times. Of course, each time we convolve the support of the result is likely to grow by (the real analogue of) (1.1). However, if we take an infinite sequence of intervals

[3] Since we are (essentially) using the Fundamental Theorem of Calculus the integrals are morally with respect to differential forms rather than measures. In particular $\int_u^v k(z)dz = -\int_v^u k(z)dz$, which we mention since h may be negative.

[4] As a quick notational remark we recall that $o_{c_1,\ldots,c_n;x\to z}(1)$ denotes a quantity which tends to 0 as $x \to z$ in a way which may depend on $c_1, \ldots, c_n$, and similarly for big-O.

where the sum of their widths is bounded then the support will remain compact; this, then, is the plan.[5]

Suppose that $(w_n)_{n\geqslant 1}$ is a sequence of positive reals such that $\sum_{n\geqslant 1} w_n = 1$ (for example take $w_n = 2^{-n}$), and let $I_n = (0, w_n)$. For $m \leqslant n$ we consider[6] the functions

$$g_{m,n} := 1_{[0,2]} * 1_{[0,1]} * f_{I_m} * \cdots * f_{I_n}.$$

Since $\sum_{n\geqslant 1} w_n = 1$ we have (in light of (the real analogue of) (1.1)) that

$$\operatorname{supp} f_{I_m} * \cdots * f_{I_n} \subset (0, w_m) + \cdots + (0, w_n) = (0, \sum_{i=m}^{n} w_i) \subset (0,1).$$

It then follows from Example 1.7 (and the fact that convolution is associative) that

$$\begin{gathered} g_{m,n}(2) = \int 1_{[0,2]} * 1_{[0,1]}(x) f_{I_m} * \cdots * f_{I_n}(2-x)dx = 1, \\ \operatorname{supp} g_{m,n} \subset [0,2] + [0,1] + \operatorname{supp} f_{I_m} * \cdots * f_{I_n} \subset (0,4), \end{gathered}$$

and (since convolution is commutative) that

$$g_{1,n} = g_{m,n} * f_{I_1} * \cdots * f_{I_{m-1}}. \tag{1.3}$$

We now have a couple of key claims.

Claim. *The functions $g_{m,n}$ are 1-Lipschitz.*

Proof of Claim. From Example 1.7 it is clear that $h := 1_{[0,2]} * 1_{[0,1]}$ is 1-Lipschitz. But then

$$\begin{aligned} |g_{m,n}(x) - g_{m,n}(y)| &= \left| \int (h(x-z) - h(y-z)) f_{I_m} * \cdots * f_{I_n}(z)dz \right| \\ &\leqslant \sup_{z\in\mathbb{R}} |(x-z) - (y-z)| \int |f_{I_m} * \cdots * f_{I_n}(z)|dz = |x-y|, \end{aligned}$$

since h is 1-Lipschitz and the claim follows. □

[5] A reasonable concern might arise at this point: the derivative in Proposition 1.10 could go up by the reciprocal of the interval length with each application. But then if we apply it with narrower and narrower intervals it seems that it might get very large. Indeed, this is true, and a little thought will convince us that actually this *must* be true: suppose that $f : \mathbb{R} \to \mathbb{R}$ is infinitely differentiable in some neighbourhood N of a point x_0 and

$$\sup\{|f^{(k)}(x)| : x \in N\} = O_{x_0}(k!). \tag{1.2}$$

Then by Taylor's theorem we have

$$f(x) = \sum_{k=0}^{n-1} \frac{f^{(k)}(x_0)(x-x_0)^k}{k!} + O_{x_0}(|x-x_0|^n) \text{ whenever } x \in N.$$

It follows that the power series for f converges to f in some (possibly smaller) neighbourhood of x_0, and so f is real analytic at x_0. Thus, if f were infinitely differentiable and satisfied (1.2) everywhere then it would be analytic, and by the identity theorem if an analytic function has compact support then it is identically 0. It follows that for bump functions (1.2) cannot hold.

[6] On a first pass one might wonder why we consider $(g_{m,n})_{n\geqslant m}$ for all m rather than just $(g_{1,n})_{n\geqslant 1}$. This is so that we can get (1.4), but the fact that this might be useful only becomes clear somewhat later.

Claim. *The sequence $(g_{m,n})_{n\geqslant m}$ is Cauchy in the L^∞-norm.*

Proof of Claim. Suppose that $\epsilon > 0$. Then there is some $n_0 \geqslant m$ such that $\sum_{n>n_0} w_n < \epsilon/2$, whence for any $n > n_0$ and $x \in \mathbb{R}$ we have

$$\begin{aligned}|g_{m,n}(x) - g_{m,n_0}(x)| &= \left|\int g_{m,n_0}(x-y) f_{I_{n_0+1}} * \cdots * f_{I_n}(y)dy - g_{m,n_0}(x)\right| \\ &= \left|\int (g_{m,n_0}(x-y) - g_{m,n_0}(x)) f_{I_{n_0+1}} * \cdots * f_{I_n}(y)dy\right| \\ &\leqslant \sup\{|(x-y)-x| : y \in \operatorname{supp} f_{I_{n_0+1}} * \cdots * f_{I_n}\} \\ &= \sup\{|y| : y \in I_{n_0+1} + \cdots + I_n\},\end{aligned}$$

since $f_{I_{n_0+1}} * \cdots * f_{I_n}$ is a probability measure and g_{m,n_0} is 1-Lipschitz. On the other hand, by design

$$I_{n_0+1} + \cdots + I_n \subset (0, \sum_{n>n_0} w_n) \subset (0, \epsilon/2).$$

It follows that for $n, n' > n_0$ we have

$$|g_{m,n}(x) - g_{m,n'}(x)| \leqslant |g_{m,n}(x) - g_{m,n_0}(x)| + |g_{m,n'}(x) - g_{m,n_0}(x)| < \epsilon,$$

and the claim is proved. □

In light of the claims the sequence $(g_{m,n})_{n\geqslant m}$ is a sequence of continuous functions supported on $[0,4]$ which converges to some continuous function g_m on $[0,4]$ with $g_m(2) = 1$.

Now, write $k_m := f_{I_1} * \cdots * f_{I_{m-1}}$ and note that it is a probability density function so that $k \geqslant 0$ and $\int k(x)dx = 1$. By the triangle inequality we have

$$\begin{aligned}|g_{m,n} * k_m(x) - g_m * k_m(x)| &= \left|\int (g_{m,n}(y) - g_m(y))k_m(x-y)dy\right| \\ &\leqslant \|g_{m,n} - g_m\|_{L^\infty(\mathbb{R})} \int |k(x-y)|dy = \|g_{m,n} - g_m\|_{L^\infty(\mathbb{R})},\end{aligned}$$

whence

$$\lim_{n\to\infty} g_{m,n} * f_{I_1} * \cdots * f_{I_{m-1}} = g_m * f_{I_1} * \cdots * f_{I_{m-1}}.$$

Combining this with (1.3) we get that

$$g_1 = g_m * f_{I_1} * \cdots * f_{I_{m-1}} \text{ for all } m \in \mathbb{N}. \tag{1.4}$$

It follows from Proposition 1.10 that g_1 is infinitely differentiable. Of course, we have already seen that g_1 is non-zero, and compactly supported, so we conclude that it is a bump function.

It may be worth remarking that much as measures can be defined as continuous linear functionals on the space of continuous functions (vanishing at infinity), distributions can be defined on the space of bump functions.

1.12. **Convolution and random variables.** Suppose that Z is an integer-valued random variable. We write p_Z for the *probability mass function* of Z, so that $p_Z(z) = \mathbb{P}(Z = z)$ and consider it as an element of $\ell^1(\mathbb{Z})$.

Now, suppose that X and Y are independent integer-valued random variables. Then the probability mass function of $X + Y$ is given by convolution:

$$p_{X+Y}(z) = \sum_{x\in\mathbb{Z}} p_X(x)p_Y(z-x) = p_X * p_Y(z).$$

In fact X and Y could be real-valued instead and this relation becomes the statement that the law of $X + Y$ is the convolution of the laws of X and Y. As we have not defined measures[7] we are not in a position to formally state this yet.

Once we have seen the very natural relationship between convolution and the Fourier transform it will be of no surprise that the Fourier transform comes up in proofs of results such as the Central Limit Theorem.

1.13. **Convolution with respect to multiplication.** One final example to illustrate the variety of applications of convolution comes from 'the other' group structure on the reals: multiplication. Convolution with respect to this structure is used a lot in number theory and a good reference for our discussion here is the book [Ten95] of Tenenbaum.

Definition 1.14 (Convolution on $\mathbb{Q}_{>0}$)**.** Given two functions $f, g \in \ell^1(\mathbb{Q}_{>0})$, their *convolution* (sometimes called Dirichlet convolution) is the function $f * g$ defined by

$$f * g(x) = \sum_{y\in\mathbb{Q}_{>0}} f(y)g(x/y) \text{ for all } x \in \mathbb{Q}_{>0}.$$

As with convolution on $\mathbb{Z}$ this is easily seen to be well-defined by the triangle inequality.

⚠**Warning** ⚠ *The indicator function of a set implicitly depends on the superset from which the set was taken. This means that convolutions of what seem like the same function can be very different: write $f^\times$ for the indicator function of $\{1, \ldots, N\}$ considered as a subset of $\mathbb{Q}_{>0}$, and f^+ for the indicator function of $\{1, \ldots, N\}$ considered as a subset of $\mathbb{Z}$. Since $f^\times \in \ell^1(\mathbb{Q}_{>0})$ we convolve it with itself using Definition 1.14 and see, for example, that*

$$\operatorname{supp} f^\times * f^\times = \{ab : a, b \in \{1, \ldots, N\}\}$$

which is a subset of $\{1, \ldots, N^2\}$ of size $\Omega(N^{2-o(1)})$. On the other hand $f^+ \in \ell^1(\mathbb{Z})$ so we convolve it with itself using Definition 1.1 and see that

$$\operatorname{supp} f^+ * f^+ = \{a + b : a, b \in \{1, \ldots, N\}\} = \{2, 3, \ldots, 2N\}$$

which is a very different set.

[7] If X and Y are absolutely continuous the statement about measures yields the fact that the probability density function of $X + Y$ is the convolution of the probability density functions of X and Y in the sense of convolution over $\mathbb{R}$ as we have defined it.

One function of interest in number theory is the *divisor function*, denoted τ, and defined on the naturals by setting $\tau(x)$ to be the number of natural divisors of x. This is related to convolution by the following identity:

$$\tau(x) = 1_{\{1,\dots,N\}} * 1_{\{1,\dots,N\}}(x) \text{ whenever } x \leqslant N.$$

The convolution here means that τ is sufficiently smooth that we can compute its average value relatively accurately using a method called Dirichlet's hyperbola method.

Proposition 1.15. *We have the estimate*

$$\sum_{x \leqslant N} \tau(x) = N \log N + (2\gamma - 1)N + O(\sqrt{N}),$$

where $\gamma := \int_1^\infty \frac{\{x\}}{x\lfloor x \rfloor} dx$ *is* Euler's constant.

To prove this we shall use the following lemma which is the context in which Euler's constant was first discovered.

Lemma 1.16. *We have the estimate*

$$\sum_{x \leqslant N} \frac{1}{x} = \log N + \gamma + O(1/N)$$

Proof. Given the definition of γ this is essentially immediate:

$$\begin{aligned} \sum_{x \leqslant N} \frac{1}{x} - \log N &= \sum_{x \leqslant N} \frac{1}{x} - \int_1^{N+1} \frac{1}{x} dx + O(1/N) \\ &= \int_1^{N+1} \left(\frac{1}{\lfloor x \rfloor} - \frac{1}{x} \right) dx + O(1/N) \\ &= \int_1^{N+1} \frac{\{x\}}{x\lfloor x \rfloor} dx + O(1/N) \\ &= \gamma + \int_{N+1}^\infty O(1/x^2) dx + O(1/N) = \gamma + O(1/N). \end{aligned}$$

The result is proved. □

Proof of Proposition 1.15. An obvious start is to note that

$$\sum_{x \leqslant N} \tau(x) = \sum_{ab \leqslant N} 1 = \sum_{a \leqslant N} \left\lfloor \frac{N}{a} \right\rfloor = \sum_{a \leqslant N} \left(\frac{N}{a} + O(1) \right) = N \log N + O(N)$$

by Lemma 1.16. The weakness of this argument is that the approximation

$$\left\lfloor \frac{N}{a} \right\rfloor = \frac{N}{a} + O(1)$$

is not a strong statement when a is close to N – the error term is of comparable size to the main term. However, since $ab \leqslant N$ we certainly have that at least one of a and b is always at most $\sqrt{N}$. It follows that

$$\sum_{ab\leqslant N}1=\sum_{a\leqslant\sqrt{N}}\sum_{b\leqslant N/a}1+\sum_{b\leqslant\sqrt{N}}\sum_{a\leqslant N/b}1-\sum_{a,b\leqslant\sqrt{N}}1=2\sum_{a\leqslant\sqrt{N}}\left\lfloor\frac{N}{a}\right\rfloor-\sum_{a,b\leqslant\sqrt{N}}1.$$

This is called the hyperbola method because it is a way of counting lattice points below the hyperbola $xy=N$. Now, as before we have that

$$\sum_{x\leqslant N}\tau(x)=2\sum_{a\leqslant\sqrt{N}}\left(\frac{N}{a}+O(1)\right)-(\sqrt{N}+O(1))^2.$$

On the other hand by Lemma 1.16 we have that

$$\sum_{a\leqslant\sqrt{N}}\frac{N}{a}=N(\log\sqrt{N}+\gamma+O(1/\sqrt{N}))=\frac{1}{2}N\log N+\gamma N+O(\sqrt{N}),$$

and the result follows on rearranging. □

Voronï used bump functions amongst other things to show that the error term is bounded by $O(N^{1/3+o(1)})$ and this has since been improved to $O(N^{\alpha})$ for some $\alpha<1/3$. In the other direction Hardy and Landau showed that the error is $\Omega(N^{1/4})$, but the true order is not known.

2. Convolution on finite Abelian groups

After all the examples of the first chapter we are now in a position to define the convolution in the setting in which we are most interested. The book [Kat04] of Katznelson gives a flavour of things which is more examples focused, while the book [Rud90] of Rudin gives a good level of generality but may be a little harder to read.

To begin with we shall need a few definitions concerning functions spaces.

2.1. **The space of measures and Lebesgue space.**

Suppose that X is a finite (non-empty) set. We write $M(X)$ for the space of complex-valued measures on X. Since X is finite, a measure is essentially just a (complex-valued) way of weighting elements of X. In particular, if $f:G\rightarrow\mathbb{C}$ we have

$$\int fd\nu=\sum_{x\in X}f(x)\nu(\{x\}).$$

The space $M(X)$ is a complex vector space and it can be normed in a natural way: for each $\nu\in M(X)$ we define the norm of ν to be

$$\|\nu\|=\sup\{\int fd\nu:\|f\|_{L^\infty(\nu)}\leqslant 1\}=\sum_{x\in X}|\nu(\{x\})|,\tag{2.1}$$

and this makes $M(X)$ into a normed space. More generally X is a locally compact topological space and $M(X)$ is the space of regular Borel measures on X. We mention this for the purposes of intuition – if it does not help it can be safely ignored.

Example 2.2 (Counting measure and uniform probability measure)**.** There are two particularly important measures supported on X. First, there is *counting measure* denoted δ_X, which we use when X is behaving like a discrete topological space. This assigns mass 1 to each element of X. When dealing with counting measure we tend to talk about the size of sets and use summation instead of integration so that

$$\delta_X(A) = |A| \text{ and } \int f(x)d\delta_X(x) = \sum_{x\in X} f(x).$$

Secondly, there is *uniform probability measure* denoted $\mathbb{P}_X$, which we use when X is behaving like a compact topological space. This assigns mass $|X|^{-1}$ to each element of X. When dealing uniform probability measure we tend to talk about the density of sets and use expectation instead of integration so that

$$\mathbb{P}_X(A) = |A|/|X| \text{ and } \int f(x)d\mathbb{P}_X(x) = \mathbb{E}_{x\in X} f(x).$$

A *positive measure* is a measure taking only non-negative values; suppose that $\nu \in M(X)$ is such, and $p \in [1,\infty]$. We define the Lebesgue space $L^p(\nu)$ to be the vector space of complex-valued functions on X endowed with the norm defined on $f : X \to \mathbb{C}$ by

$$\|f\|_{L^p(\nu)} := \left(\int |f(x)|^p dx\right)^{1/p},$$

with the usual convention when $p = \infty$, that is to say

$$\|f\|_{L^\infty(\nu)} = \max\{|f(x)| : x \in \operatorname{supp} \nu\}.$$

Technically these norms may only be a semi-norms (if the support of ν is not the whole of X), but this will make no difference to us and we can import essentially all our intuition about normed spaces.

Example 2.3 ($\ell^p(X)$ and $L^p(X)$)**.** The measures defined in Example 2.2 also give rise to two special classes of Lebesgue space: we write

$$\ell^p(X) := L^p(\delta_X) \text{ and } L^p(X) := L^p(\mathbb{P}_X),$$

where this is equality in the sense of normed spaces not just vector spaces. These Lebesgue spaces have a useful nesting of norms property which is an immediate application of Hölder's inequality:

$$\|f\|_{\ell^p(X)} \leqslant \|f\|_{\ell^q(X)} \text{ and } \|f\|_{L^q(X)} \leqslant \|f\|_{L^p(X)} \text{ whenever } p \geqslant q \text{ and } p, q \in [1,\infty].$$

Remark. The nesting of norm inequalities above have p and q one way for the $\ell^p(G)$-spaces and the other for the $L^p(G)$-spaces. In general when recalling inequalities it can be difficult to remember which way round they are. To help with this it can be useful simply to test them against some functions. For the nesting of norm inequalities above one can simply test the inequalities against the function $1_{\{x\}}$ (for some element $x \in X$) in $L^p(X)$ and 1_X in $\ell^p(X)$, where we get

$$\|1_{\{x\}}\|_{L^p(X)} = |X|^{-1/p} \text{ and } \|1_X\|_{\ell^p(X)} = |X|^{1/p}.$$

The first of these is visibly increasing in p and the second decreasing.

2.4. **Haar measure and convolution.** Convolution makes sense on any locally compact group, and there are arguments for establishing it in this generality. We shall concentrate on finite Abelian groups which the reader may note does not even include the examples of the introduction. It turns out that all the phenomena we are interested in can be seen in finite groups so this is not the loss which it might at first appear to be. Readers interested in work concentrating on finite groups might like to consult the book [Ter99] of Terras, although she does proceed in the more general setting including non-Abelian groups.

Definition 2.5 (Haar measure on G)**.** For any locally compact group there is a unique (up to scale) translation invariant regular Borel measure on the group called Haar measure. In particular, this is true for G: we say that a measure μ on G is a *Haar measure* if it is strictly positive and translation invariant. It is easy to see that this means that μ gives each element of G the same mass.

On G there are two particularly important Haar measures coming from Example 2.2: δ_G, counting measure on G, and $\mathbb{P}_G$, uniform probability measure on G.

With Haar measure defined we can make sense of convolution.

Definition 2.6 (Convolution)**.** Given a Haar measure μ and two functions $f, g \in L^1(\mu)$ we define their convolution to be the function $f * g$ determined point-wise by

$$f * g(x) = \int f(y)g(x-y)dy = \int f(y)g(x-y)d\mu(y) \text{ for all } x \in G.$$

We shall tend to write dx instead of $d\mu(x)$ in integration when the Haar measure is clear. Much as in Definition 1.6 we have that

$$\begin{aligned} \|f * g\|_{L^1(\mu)} &= \int |\int f(y)g(x-y)dy|dx \\ &\leqslant \int |f(y)| \int |g(x-y)|dxdy = \|f\|_{L^1(\mu)}\|g\|_{L^1(\mu)} \end{aligned}$$

which is a special case of Young's inequality.

We can also convolve functions with measures: indeed, suppose that $f \in L^1(\mu)$ and $\nu \in M(G)$. Then we define $f * \nu$ and $\nu * f$ by

$$f * \nu(x) = \nu * f(x) := \int f(x-y)d\nu(y) \text{ for all } x \in G,$$

and in a similar way to the above we have that $\|f * \nu\|_{L^1(\mu)} \leqslant \|f\|_{L^1(\mu)}\|\nu\|$.

Finally, given two measures $\nu, \rho \in M(G)$ (*not* necessarily Haar measures) we define their *convolution* $\nu * \rho$ to be the measure determined by

$$\nu * \rho(A) = \int 1_A(y+z)d\nu(y)d\rho(z).$$

Since G is finite, it may be helpful to note that

$$\begin{aligned} \nu * \rho(\{x\}) &= \int 1_{\{x\}}(y+z)d\nu(y)d\rho(z) \\ &= \sum_{y,z \in G} 1_{\{x\}}(y+z)\nu(\{y\})\rho(\{z\}) = \sum_{y \in G} \nu(\{y\})\rho(\{x-y\}). \end{aligned}$$

Again, it is fairly straightforward to show that $\|\nu * \rho\| \leqslant \|\nu\|\|\rho\|$, and so $M(G)$ forms an associative (commutative) algebra.

⚠**Warning** ⚠ *The convolution of two functions depends on the particular choice of Haar measure. Given $f, g \in L^1(G)$ we have that*

$$f * g(x) = \mathbb{E}_{y \in G} f(y)g(x-y).$$

On the other hand, if we write f' for the function f considered as an element of $\ell^1(G)$ and g' for the function g considered as an element of $\ell^1(G)$ then

$$f' * g'(x) = \sum_{y \in G} f(y)g(x-y).$$

In particular, as functions we have $f' * g' = |G|.f * g$. This may seem rather worrying at first, but in practice if one finds oneself out by a power of $|G|$ in a calculation it can usually be traced back to a normalisation error.

At this point it is useful to record a few simple facts about convolution. As always checking this sort of thing is important, but in this instance we shall leave it as an exercise.

Lemma 2.7 (Basic facts)**.** *Suppose that G is endowed with a Haar measure μ and $f, g, h \in L^1(\mu) \cup M(G)$. Then*

(i) (Linearity) *$(\lambda f + \mu g) * h = \lambda(f * h) + \mu(g * h)$ for all $\lambda, \mu \in \mathbb{C}$;*
(ii) (Associativity) *$f * (g * h) = (f * g) * h$;*
(iii) (Commutativity) *$f * g = g * f$.*

We should remark that while we usually reserve the letters f, g and h for functions, in this instance they can be either functions or measures. We have done this simply to avoid stating essentially the same thing several different times.

The point of Lemma 2.7 is that it shows that $L^1(\mu)$ and $M(G)$ form commutative normed algebras, and in fact there is a rather close relationship between the two.

2.8. **Embedding $L^1(\nu)$ in $M(X)$.** Returning, briefly, to the setting where X is any finite non-empty set, then given a positive measure $\nu \in M(X)$ we can embed $L^1(\nu)$ into $M(X)$ via the map taking $f \in L^1(\nu)$ to the measure $fd\nu$ defined by

$$(fd\nu)(A) = \int 1_A f d\nu \text{ for all } A \subset X.$$

The map $f \mapsto fd\nu$ is then an isometric linear embedding of $L^1(\nu)$ into $M(X)$. More than this, if $X = G$ and μ is a Haar measure on G then the embedding is also an algebra homomorphism so that

$$(f d\mu) * (g d\mu) = (f * g) d\mu \text{ for all } f, g \in L^1(\mu).$$

2.9. **Relative p.d.fs and convolution as a measure of relative density.** Given a non-empty set $S \subset G$ we write μ_S for the uniform probability measure supported on S, in $M(G)$. This differs slightly from $\mathbb{P}_S$ which is defined only on subsets of S, but there is nevertheless a relationship between them (which one could take as defining for μ_S):

$$\mu_S(A) = \mathbb{P}_S(A \cap S) \text{ for all } A \subset G.$$

Now, suppose that $X, A \subset G$. Then it turns out that $1_A * \mu_X(x)$ is the relative density of A on the set $x - X$, that is to say it is the number of points in $A \cap (x - X)$ divided by the number of points in $x - X$ (which is the same as the number of points in X). To see this note that

$$\begin{aligned} 1_A * \mu_X(x) &= \int 1_A(y) d\mu_X(x-y) \\ &= \frac{1}{|X|} \sum_{y \in G} 1_A(y) 1_X(x-y) = \frac{|A \cap (x-X)|}{|X|}. \end{aligned}$$

We shall frequently use this in the case when $X = V$ for some subgroup $V \leqslant G$. In this case $1_A * \mu_V(x)$ is the relative density of A on the coset $x - V = x + V$. More than this, if $v \in V$ then $x + v + V = x + V$, so we see that $1_A * \mu_V(x+v) = 1_A * \mu_V(x)$, and hence $1_A * \mu_V$ is constant on cosets of V. (The reader may wish to compare this with the later Example 2.16.)

At this point it may be instructive to consider a couple of examples of convolutions.

Example 2.10 (Convolution of subgroups)**.** Suppose that $V, W \leqslant G$. Then

$$\mu_V * \mu_W = \mu_{V+W}.$$

To see this first note, by (the analogue of) (1.1), that the support of the left hand side is equal to $V + W$. On the other hand

$$\mu_V * \mu_W(\{x\}) = \frac{1}{|V||W|} \sum_{y \in G} 1_V(y) 1_W(x-y) = \frac{1}{|V||W|} |V \cap (x - W)|,$$

so if $x \in V + W$ then $x = v + w$ for some $v \in V$ and $w \in W$, whence

$$|V \cap (x-W)| = |V \cap (v + w - W)| = |(V - v) \cap (w - W)| = |V \cap W|,$$

so $\mu_V * \mu_W$ is supported on $V + W$, constant on $V + W$, and by integrating we see it is a probability measure whence we get the claimed equality.

Example 2.11 (Convolution of random sets)**.** Suppose that each $x \in G$ is placed in the set A independently with probability $\alpha = \Omega(1)$. Then $\mathbb{E}\mathbb{P}_G(A) = \alpha$ and

$$\mathbb{E} 1_A * 1_{-A}(x) = \mathbb{E}_{y \in G} \mathbb{E} 1_A(y) 1_A(y - x) = \begin{cases} \alpha^2 & \text{if } x \neq 0_G \\ \alpha & \text{otherwise.} \end{cases}$$

Indeed, we expect it to be very likely that $A-A$ is essentially the whole of G, and to see this we present an informal variance calculation: first note that

$$\begin{aligned}\mathbb{E}1_A * 1_{-A}(x)^2 &= \frac{1}{|G|^2}\sum_{y,z}\mathbb{E}1_A(y)1_A(z)1_A(y-x)1_A(z-x)\\ &= \begin{cases}\alpha^2+O(\alpha/|G|) & \text{if } x=0_G\\ \alpha^4+O(\alpha^2/|G|) & \text{otherwise.}\end{cases}\end{aligned}$$

It follows from this that[8] $\mathrm{Var}(1_A * 1_{-A}(x)) = O(1/|G|)$, and so by Chebychev's inequality we have that

$$\begin{aligned}\mathbb{P}(1_A * 1_{-A}(x)=0) &= \mathbb{P}(1_A * 1_{-A}(x)-\mathbb{E}1_A * 1_{-A}(x) \leqslant -\mathbb{E}1_A * 1_{-A}(x))\\ &\leqslant \frac{\mathrm{Var}(1_A * 1_{-A}(x))}{(\mathbb{E}1_A * 1_{-A}(x))^2} = O(1/|G|).\end{aligned}$$

We conclude that the expected number of x with $x \notin A-A$ is $O(1)$. In fact, by Chernoff-type tail estimates of a sort we shall cover later on, one can see that it is much less than 1 and that with high probability $A-A=G$.

When we defined convolution in Definition 2.6 we established a simple algebra inequality: given μ a Haar measure on G we had

$$\|f * g\|_{L^1(\mu)} \leqslant \|f\|_{L^1(\mu)}\|g\|_{L^1(\mu)} \text{ for all } f,g \in L^1(\mu).$$

This is an important fact which it turns out is part of a wider range of inequalities.

Proposition 2.12 (Young's inequality)**.** *Suppose that μ is a Haar measure on G, $f \in L^p(\mu)$, $g \in L^q(\mu)$ and $1+1/r = 1/p+1/q$. Then*

$$\|f * g\|_{L^r(\mu)} \leqslant \|f\|_{L^p(\mu)}\|g\|_{L^q(\mu)}.$$

Proof. Since G is finite no analysis is required and we simply have a calculation:

$$\begin{aligned}\|f * g\|_{L^r(\mu)}^r &\leqslant \int\left(\int |f(y)||g(x-y)|dy\right)^r dx\\ &= \int\left(\int |f(y)|^{p/r}|g(x-y)|^{q/r}.|f(y)|^{p(1/p-1/r)}.|g(x-y)|^{q(1/q-1/r)}dy\right)^r dx\\ &\leqslant \int\left(\left(\int |f(y)|^p|g(x-y)|^q dy\right)^{1/r}\left(\int |f(y)|^p dy\right)^{1/p-1/r}\right.\\ &\quad \times\left.\left(\int |g(x-y)|^q dy\right)^{1/q-1/r}\right)^r dx = \|f\|_{L^p(\mu)}^r\|g\|_{L^q(\mu)}^r,\end{aligned}$$

where the passage between the second and third line is via the three-variable Hölder inequality which applies since $(1/r)+(1/p-1/r)+(1/q-1/r)=1$. Taking rth roots the result is proved. □

[8] A more accurate calculation of the variance shows that it takes three different values depending on whether $x=0_G$, $2x=0_G$ and $x \neq 0_G$, or $2x \neq 0_G$. We leave this calculation to the interested reader.

It is worth making a few remarks on the quality of this inequality. Since G is finite we can take $f = g = 1_G$ and we see that the inequality is tight for any (admissible) triple of indices (p, q, r). For $p = 1$, $q = 1$ or $r = \infty$ the inequality is (true and) tight for any locally compact group, but for other, so called *internal* triples of indices (that is triples with $p > 1$, $q > 1$ and $r < \infty$) it is not.

When G is finite the tightness for internal triples is a result of the fact that G has an 'open and compact' subgroup. In groups such as $\mathbb{R}$ without any open compact subgroup, given an internal triple of indices (p, q, r) there is a constant $c_{p,q} < 1$ such that

$$\|f * g\|_{L^r(\mathbb{R})} \leqslant c_{p,q}\|f\|_{L^p(\mathbb{R})}\|g\|_{L^q(\mathbb{R})} \text{ for all } f \in L^p(\mathbb{R}), g \in L^q(\mathbb{R}).$$

The existence of such a constant in locally compact groups was shown by Fournier [Fou77], while the best possible constant for locally compact Abelian groups was found by Beckner in the classic paper [Bec75].

When we defined convolution in Definition 2.6 we also defined it between functions and measures and there is a hybrid of Young's inequality and the algebra inequality for measures (that is $\|\nu * \rho\| \leqslant \|\nu\|\|\rho\|$) in that case. The proof is left as an exercise.

Proposition 2.13 (Young's inequality for measures)**.** *Suppose that μ is a Haar measure on G, $\nu \in M(G)$ and $f \in L^p(\mu)$. Then*

$$\|f * \nu\|_{L^p(\mu)} \leqslant \|f\|_{L^p(\mu)}\|\nu\|.$$

We now come to a crucial definition.

Definition 2.14 (Convolution operators)**.** Suppose that G is endowed with a Haar measure μ. Then to each $\nu \in M(G)$ we can associate a linear map

$$M_\nu : L^p(\mu) \to L^p(\mu); f \mapsto \nu * f.$$

Moreover, in light of Young's inequality for measures, the map

$$M(G) \to B(L^p(\mu)); \nu \mapsto M_\nu$$

is an injective algebra homomorphism of norm 1, where multiplication in $M(G)$ is convolution and in $B(L^p(\mu))$ is composition.

We could make a similar definition for functions in $L^1(\mu)$, but in light of the embedding in §2.8 we shall not bother since $L^1(\mu)$ sits isometrically as a sub-algebra of $M(G)$. Were it needed we should put $M_f := M_{fd\mu}$ for all $f \in L^1(\mu)$.

Definition 2.15 (Adjoints)**.** Given a measure $\nu \in M(G)$ we write $\tilde{\nu}$ for its *adjoint* measure defined by

$$\tilde{\nu}(A) := \overline{\nu(-A)} \text{ for all } A \subset .$$

The reason for the name is that if μ is a Haar measure on G, then $M_\nu^* = M_{\tilde{\nu}}$ when M_ν is considered as an operator on $L^2(\mu)$. Indeed, we have

$$\begin{aligned}\langle M_\nu f, g\rangle_{L^2(\mu)} &= \int\int f(y)d\nu(x-y)\overline{g(x)}d\mu(x)\\ &= \int\int f(y)\overline{g(y-z)}d\mu(y-z)d\nu(-z)\\ &= \int\int f(y)\overline{g(y-z)}d\mu(y)d\nu(-z) = \langle f, M_\nu g\rangle_{L^2(\mu)}\end{aligned}$$

for all $f, g \in L^2(\mu)$ since μ is translation invariant (so that $d\mu(y-z) = d\mu(y)$).

Example 2.16 (Convolution operators as projections)**.** Suppose that G is endowed with a Haar measure μ. A special case of Example 2.10 is that when $V \leqslant G$ we have $\mu_V * \mu_V = \mu_V$, in which case the operator M_{μ_V} is a projection since $M_{\mu_V}^2 = M_{\mu_V * \mu_V} = M_{\mu_V}$.

More than this, since μ_V is positive we see that the operator norm of M_{μ_V}, denoted $\|M_{\mu_V}\|$, is 1, so in a sense the projection is 'as good as can be'. In the case $p = 2$ this means that M_{μ_V} is an *orthogonal* projection on $L^2(\mu)$. Indeed, looking back at §2.9 we see that

$$M_{\mu_V}(f) = \mathbb{E}(f|\sigma(G/V))$$

where $\sigma(G/V)$ is the σ-algebra generated by the partition G/V.

3. The Fourier transform on finite Abelian groups

In the first chapter we saw that convolution is a useful and varied tool; in the second we set it up on general finite Abelian groups. More than this we defined convolution operators which give a slightly different language for expressing quantities of interest in terms of convolution. Whenever we have an operator we ask if there is a natural basis in which to represent it, and in this case there is and it is the Fourier basis.

Throughout this chapter μ will be a Haar measure on G. The set of convolution operators naturally acts on the Hilbert space $L^2(\mu)$; it is commuting as a result of the basic facts in Lemma 2.7:

$$M_\nu M_\rho = M_{\nu*\rho} = M_{\rho*\nu} = M_\rho M_\nu \text{ for all } \nu, \rho \in M(G);$$

and it is adjoint-closed as a result of the comments in Definition 2.15:

$$M_\nu^* = M_{\tilde{\nu}} \text{ for all } \nu \in M(G).$$

We now recall a basic theorem from linear algebra.

Theorem 3.1 (Spectral theorem)**.** *Suppose that H is a finite dimensional Hilbert space and $\mathcal{M}$ is an adjoint-closed set of commuting operators on H. Then there is an orthonormal basis of H with respect to which every operator in $\mathcal{M}$ is diagonal*[9]*.*

[9] Note that throughout this set of notes we shall be interested in diagonalisation with respect to unitary matrices *not* invertible matrices.

We shall be interested in the above theorem applied to the collection of convolution operators, in which case we get a basis in which every element is an eigenvector for every convolution operator. It turns out that such vectors have a very special structure, and to illicit this it will be useful to have a convenient basis for the convolution operators.

3.2. **δ-functions: a basis for the algebra of convolution operators.** The set of convolution operators forms a space and hence, itself, has a basis[10]. Write δ_x for the probability measure assigning mass 1 to x (and hence 0 everywhere else). Then

$$M_{\delta_x}(f)(y) = \int f(y-z)d\delta_x(z) = f(y-x) \text{ for all } x,y \in G. \tag{3.1}$$

Since G is finite the set $\{\delta_x\}_{x\in G}$ forms a basis for $M(G)$. Indeed, the measures are visibly independent and

$$\nu = \int \delta_x d\nu(x) \text{ for all } \nu \in M(G).$$

Now the map $\nu \mapsto M_\nu$ is linear so it follows that

$$M_\nu = \int M_{\delta_x} d\nu(x) \text{ for all } \nu \in M(G);$$

the operators $(M_{\delta_x})_{x\in G}$ form a basis for the algebra of convolution operators.

We now use this basis to study the eigenvectors afforded by the Spectral Theorem (Theorem 3.1).

Proposition 3.3. *Suppose that $v \in L^2(\mu)$ is an eigenvector for every convolution operator. Then there is a homomorphism $\lambda_v : G \to S^1$ and $\sigma \in \mathbb{C}$ such that $v = \sigma\lambda_v$ and*

$$M_\nu v = \left(\int \overline{\lambda_v(x)} d\nu(x)\right) v \text{ for all } \nu \in M(G). \tag{3.2}$$

Conversely, suppose that v is a scalar multiple of a homomorphism $\lambda_v : G \to S^1$. Then we have (3.2).

Proof. We let $\lambda_v(x)$ be the eigenvalue of the operator $M_{\delta_{-x}}$ corresponding to v. Note that $M_{\delta_{-0_G}}$ is the identity so $\lambda_v(0_G) = 1$, while

$$\lambda_v(x+y)v = M_{\delta_{-(x+y)}}v = M_{\delta_{-x}}M_{\delta_{-y}}v = M_{\delta_{-x}}(\lambda_v(y)v) = \lambda_v(x)\lambda_v(y)v$$

for all $x, y \in G$, so $\lambda_v(x+y) = \lambda_v(x)\lambda_v(y)$ for all $x, y \in G$. It follows that λ_v is a homomorphism. To see that λ_v maps into S^1 we note by (3.1) and the translation invariance of μ that

$$\begin{aligned} \|v\|_{L^2(\mu)}^2 = \int |v(y)|^2 dy = \int |v(y+x)|^2 dy &= \int |M_{\delta_{-x}} v(y)|^2 dy \\ &= \int |\lambda_v(x)|^2 |v(y)|^2 dy = |\lambda_v(x)|^2 \|v\|_{L^2(\mu)}^2, \end{aligned}$$

[10] Note that this is a basis of a subspace of $B(L^2(\mu))$, *not* of $L^2(\mu)$

whence $|\lambda_v(x)| = 1$ as required. Now, in light of (3.1) we see that

$$v(x) = M_{\delta_{-x}}v = \lambda_v(x)v(0_G) \text{ for all } x \in G,$$

from which is follows that $v = \sigma\lambda_v$ with $\sigma = v(0_G)$, and finally to get (3.2) we note that

$$\begin{aligned} M_\nu v = \left(\int M_{\delta_x} d\nu(x)\right) v &= \int (M_{\delta_x} v) d\nu(x) \\ &= \int (\lambda_v(-x)v) d\nu(x) = \left(\int \overline{\lambda_v(x)} d\nu(x)\right) v \end{aligned}$$

for all $\nu \in M(G)$ as required.

In the other direction we have an easy calculation. Suppose that $v = \sigma\lambda_v$ for some scalar $\sigma \in \mathbb{C}$ and homomorphism $\lambda_v : G \to S^1$. Then

$$M_\nu v(y) = \int \sigma\lambda_v(y - x) d\nu(x) = \sigma\lambda_v(y) \int \overline{\lambda_v(x)} d\nu(x) = \left(\int \overline{\lambda_v(x)} d\nu(x)\right) v(y)$$

for all $y \in G$ and $\nu \in M(G)$, and the result is proved. □

A homomorphism from G to S^1 is called a *character*[11] of G, and from this point on we write $\widehat{G}$ for the set of characters on G.

The Spectral Theorem applied to the class of convolution operators gives us an orthonormal basis of vectors which, by Proposition 3.3, are scalar multiples of some characters. It turns out that they are the same scalar multiple for all characters, but we have yet to prove this. This basis suggests two main questions:

(i) What do vectors (functions) $f \in L^2(\mu)$ 'look like' with respect to this new basis?
(ii) What do the operators M_ν for $\nu \in M(G)$ 'look like' with respect to this new basis?

To address the first of these questions we make the following definition.

Definition 3.4 (Fourier transform)**.** Given $f \in L^2(\mu)$ we define the *Fourier transform* of f to be the function $\widehat{f} : \widehat{G} \to \mathbb{C}$ determined by

$$\widehat{f}(\gamma) = \langle f, \gamma\rangle = \int f(x)\overline{\gamma(x)} d\mu(x).$$

The idea is that $\widehat{f}(\gamma)$ is the projection of f onto the vector γ, so that $\widehat{f}$ is f written with respect to the orthogonal basis $\widehat{G}$, although we have not yet shown that $\widehat{G}$ is such.

The basis afforded by the Spectral theorem is an orthonormal basis, but the set $\widehat{G}$ we have chosen will only turn out to be orthogonal – the characters can, in general, have large norm. It follows that we shall need to weight $\widehat{G}$ and we do so by defining a measure μ^* on $\widehat{G}$, called the *dual measure*, assigning mass $\mu(G)^{-1}$ to each element of $\widehat{G}$. The reason for this choice is that it is what comes out of the proof of the following theorem.

[11] More generally for any finite group a character is the trace of a representation. For extensions in this direction the reader may wish to consult a book on representation theory *e.g.* James and Liebeck [JL93].

Theorem 3.5 (Parseval's theorem)**.** *The Fourier transform*

$$\widehat{\cdot}: L^2(\mu) \to L^2(\mu^*); f \mapsto \widehat{f}$$

is an isometric linear map.

Proof. The particular choice of μ^* comes out of the proof. We start by noting that $f \mapsto \widehat{f}$ is linear, so it suffices to prove that it is isometric.

The Spectral Theorem gives us an orthonormal basis $v_1, \dots, v_N$ of $L^2(\mu)$ such that each v_i is an eigenvector for every convolution operator. It follows from Proposition 3.3 that for each i there is a character γ_i and a scalar σ_i such that $v_i = \sigma_i\gamma_i$, and by Pythagoras' Theorem (sometimes called the generalised Parseval identity) we have that

$$\|f\|_{L^2(\mu)}^2 = \sum_{i=1}^N |\langle f_i, v_i\rangle_{L^2(\mu)}|^2 = \sum_{i=1}^N |\sigma_i|^2|\widehat{f}(\gamma_i)|^2 = \sum_{\gamma\in\widehat{G}} |\widehat{f}(\gamma)|^2 m(\gamma)$$

where $m(\gamma) = \sum_{i:\gamma_i=\gamma} |\sigma_i|^2$ is non-negative. We compute the values of m by testing the above equality against characters. Indeed, suppose that $\lambda \in \widehat{G}$. Since μ is translation invariant, for any $z \in G$ we have

$$\begin{aligned} \widehat{\lambda}(\gamma) = \int \lambda(x)\overline{\gamma(x)}d\mu(x) &= \int \lambda(x+z)\overline{\gamma(x+z)}d\mu(x) \\ &= \lambda(z)\overline{\gamma(z)}\int \lambda(x)\overline{\gamma(x)}d\mu(x). \end{aligned} \tag{3.3}$$

It follows that if $\gamma \neq \lambda$ then $\widehat{\lambda}(\gamma) = 0$ since in that case there is some $z \in G$ such that $\gamma(z) \neq \lambda(z)$.[12] On the other hand $\widehat{\lambda}(\lambda) = \mu(G)$ so that

$$\mu(G) = \int |\lambda(x)|^2 d\mu(x) = \sum_{\gamma\in\widehat{G}} |\widehat{\lambda}(\gamma)|^2 m(\gamma) = \mu(G)^2 m(\lambda).$$

and hence $m(\lambda) = \mu^*(\{\lambda\})$; the result is proved. □

Remark. The proof above may seem a little technical; it is not as hard as it appears. At its core, Parseval's theorem is just saying that $\widehat{G}$ is an orthogonal (note *not* orthonormal) basis, and then by generalised Parseval we have that

$$\|f\|_{L^2(\mu)}^2 = \sum_{\gamma\in\widehat{G}} \frac{|\langle f, \gamma\rangle_{L^2(\mu)}|^2}{\langle \gamma, \gamma\rangle_{L^2(\mu)}}.$$

The Spectral theorem and Proposition 3.3 tell us that $\widehat{G}$ generates $L^2(\mu)$; a calculation shows that the characters are orthogonal; and the fact that $|\gamma(x)| = 1$ shows that $\langle \gamma, \gamma\rangle_{L^2(\mu)} = \mu(G)$, so picking $\mu^*(\{\gamma\}) = 1/\mu(G)$ gives the result.

[12] This fact actually follows for all pairs of characters (γ_i, γ_j) with $i \neq j$ from the fact that the basis is an orthogonal basis. However, we do not know that the basis contains a multiple of *every* character until the end of the argument so we still need the calculation.

As usual an isometry between Hilbert spaces can be de-polarised to give a relationship between inner products.

Corollary 3.6 (Plancherel's theorem)**.** *We have the identity*

$$\langle f, g\rangle_{L^2(\mu)} = \langle \widehat{f}, \widehat{g}\rangle_{L^2(\mu^*)} \text{ for all } f, g \in L^2(\mu).$$

Proof. On the one hand for every $\lambda \in \mathbb{C}$ we have

$$\|f + \lambda g\|^2_{L^2(\mu)} = \|f\|^2_{L^2(\mu)} + 2\operatorname{Re}\overline{\lambda}\langle f, g\rangle_{L^2(\mu)} + \|g\|^2_{L^2(\mu)},$$

and on the other (for the same λ) we have

$$\|\widehat{f} + \lambda\widehat{g}\|^2_{L^2(\mu^*)} = \|\widehat{f}\|^2_{L^2(\mu^*)} + 2\operatorname{Re}\overline{\lambda}\langle \widehat{f}, \widehat{g}\rangle_{L^2(\mu^*)} + \|\widehat{g}\|^2_{L^2(\mu^*)}.$$

By Parseval's theorem (applied three times) and linearity of the Fourier transform we see that

$$2\operatorname{Re}\overline{\lambda}\langle f, g\rangle_{L^2(\mu)} = 2\operatorname{Re}\overline{\lambda}\langle \widehat{f}, \widehat{g}\rangle_{L^2(\mu^*)}.$$

Taking $\lambda = 1$ and $\lambda = i$ gives us that the real and imaginary parts are equal, and hence the result. □

Having understood what happens to functions $f \in L^2(\mu)$ under our change of basis we now turn to our operators. Again, motivated by Proposition 3.3 we make the following definition.

Definition 3.7 (The Fourier-Stieltjes transform)**.** Given $\nu \in M(G)$ we define the *Fourier-Stieltjes transform* of ν to be the map $\widehat{\nu} : \widehat{G} \to \mathbb{C}$ determined by

$$\widehat{\nu}(\gamma) = \int \overline{\gamma(y)} d\nu(y) \text{ for all } \gamma \in \widehat{G}.$$

Here the idea is that $\widehat{\nu}(\gamma)$ is the eigenvalue of the operator M_ν on the vector γ.

The Fourier-Stieltjes transform has a number of simple but important properties which we now collect together. Their content is to show that all the important information contained in the algebra $M(G)$ is preserved under the Fourier-Stieltjes transform.

Theorem 3.8. *The Fourier-Stieltjes transform*

$$\widehat{\cdot} : M(G) \to \ell^\infty(\widehat{G}); \nu \mapsto \widehat{\nu}$$

is an injective, norm 1 *algebra* $*$*-homomorphism*[13] *from* $M(G)$ *under convolution to* $\ell^\infty(\widehat{G})$ *under point-wise multiplication.*

[13] A $*$-homomorphism is a homomorphism which also preserves adjoints, so that in this case $\widehat{\tilde{\mu}} = \overline{\widehat{\mu}}$. Note that the adjoint depends on the underlying multiplication: on $M(G)$ our multiplication is convolution, not point-wise multiplication. Were it to be the latter then, like $\ell^\infty(\widehat{G})$, the adjoint of μ would be $\overline{\mu}$.

Proof. We see immediately that the Fourier-Stieltjes transform is a linear map from $M(G)$ to $\ell^\infty(\widehat{G})$, and it is easy to check that

$$\widehat{\tilde{\mu}}(\gamma) = \int \gamma(x) d\tilde{\mu}(x) = \int \gamma(x)\overline{d\mu(-x)} = \overline{\int \gamma(x) d\mu(x)} = \overline{\widehat{\mu}(\gamma)}$$

for all $\gamma \in \widehat{G}$ so that it preserves adjoints. To see that it is an algebra homomorphism note that by the converse part of Proposition 3.3 every character is an eigenvector of every convolution operator with eigenvalue equal to the Fourier-Stieltjes transform of the measure inducing the operator at that character, so

$$\widehat{\nu * \rho}(\gamma)\gamma = M_{\nu*\rho}\gamma = M_\nu M_\rho \gamma = M_\nu \widehat{\rho}(\gamma)\gamma = \widehat{\nu}(\gamma)\widehat{\rho}(\gamma)\gamma$$

for all $\gamma \in \widehat{G}$ and $\nu, \rho \in M(G)$. We conclude that

$$\widehat{\nu * \rho} = \widehat{\nu} \cdot \widehat{\rho} \text{ for all } \nu, \rho \in M(G),$$

as required.

The norm of this homomorphism can be easily seen to be at most 1: since $|\gamma(x)| \leqslant 1$ for all $x \in G$ we get that

$$\|\widehat{\nu}\|_{\ell^\infty(\widehat{G})} \leqslant \sup\{\int \overline{\gamma(x)} d\nu(x) : \gamma \in \widehat{G}\} = \|\nu\|$$

for all $\nu \in M(G)$, which is (a slight variant of) a particular instance of the Hausdorff-Young inequality. Moreover it is equal to 1 as can be seen by noting that for ν positive, $\widehat{\nu}$ at the constant 1 function (which is a character) is $\|\nu\|$, so for such ν we have $\|\widehat{\nu}\|_{\ell^\infty(\widehat{G})} = \|\nu\|$.

Finally, we need to check injectivity. Since the Fourier-Stieltjes transform is linear it suffices to check that it has trivial kernel, and this follows from our application of the Spectral Theorem and Proposition 3.3. Indeed, if $\widehat{\nu} \equiv 0$ then M_ν is identically 0 (by that pair of results), which means that $\nu * f = 0$ for all $f \in L^2(\mu)$. By testing this against the functions $(1_{\{x\}})_{x\in G}$ we see that $\nu \equiv 0$ □

Remark. In §2.8 we identified a way of embedding $L^1(\mu)$ into $M(G)$ and this naturally gives us a way to extend the Fourier-Stieltjes transform to $L^1(\mu)$: we should define the Fourier transform of f to be $\widehat{fd\mu}$. Fortunately this is consistent with Definition 3.4 in that if $f \in L^1(\mu) \cap L^2(\mu)$ then we have $\widehat{fd\mu} = \widehat{f}$.

As a result of this equivalence we inherit the algebra identity:

$$\widehat{f * g} = \widehat{f} \cdot \widehat{g} \text{ for all } f, g \in L^1(\mu) \cap L^2(\mu).$$

Of course, since G is finite this applies to *all* functions f and g. (If G were not finite we should just take limits.)

We also inherit a trivial instance of the Hausdorff-Young inequality via the same embedding. Namely that

$$\|\widehat{f}\|_{L^\infty(\mu^*)} \leqslant \|f\|_{L^1(\mu)} \text{ for all } f \in L^1(\mu). \tag{3.4}$$

This can, of course, just be proved directly from the triangle inequality.

One question left open by Parseval's theorem as we stated it is whether the Fourier transform is a surjection. We also know that both the Fourier transform and the Fourier-Stieltjes transform are injections from Parseval's theorem and Theorem 3.8 respectively, and so it is natural to ask if we can easily reconstruct functions from their Fourier transform. It turns out that we can and shall do so now.

Theorem 3.9 (Fourier inversion formula)**.** *The Fourier transform* $\widehat{\cdot} : L^2(\mu) \to L^2(\mu^*)$ *is a surjection. Moreover, if* $f \in L^1(\mu)$ *is such that then we have*

$$f(x) = \int \widehat{f}(\gamma)\gamma(x)d\mu^*(\gamma) \text{ for all } x \in G.$$

Proof. As usual, since G is finite the inversion formula applies to all functions on G. First suppose that $h \in L^1(\mu^*)$ and put

$$k(x) := \int h(\gamma)\gamma(x)d\mu^*(\gamma).$$

Then by linearity we have

$$\widehat{k}(\lambda) = \int h(\gamma) \int \gamma(x)\overline{\lambda(x)}d\mu(x)d\mu^*(\gamma) = \int h(\gamma)\widehat{\gamma}(\lambda)d\mu^*(\gamma) = h(\lambda)$$

since $\widehat{\gamma}(\lambda) = \mu(G)$ if $\gamma = \lambda$ and is 0 otherwise.[14] It follows that the Fourier transform is surjective.

Now, suppose that $f \in L^1(\mu)$ and put $h = \widehat{f}$ so that $\widehat{k} = \widehat{f}$. It then follows by uniqueness (Theorem 3.8) that $k = f$ and the result is proved. □

Given the Fourier inversion formula we make the following definition.

Definition 3.10 (Inverse Fourier transform)**.** Given $f \in L^1(\mu^*)$ we define

$$f^\vee(x) := \int f(\gamma)\gamma(x)d\mu^*(\gamma) \text{ for all } x \in G.$$

Shortly we shall see using Pontryagin duality that this is itself (almost) a Fourier transform.

From this point on we shall refer to both the Fourier transform and the Fourier-Stieltjes transform under the one banner of the Fourier transform.

The set $\widehat{G}$ is really just (equivalent to) the set of (equivalence classes of) irreducible representation of G. Moreover, since G is Abelian, they are all 1-dimensional, which means that the tensor product gives rise to a genuine group operation. This group structure is peculiar to the Abelian setting.

[14] See (3.3) in the proof of Parseval's theorem for this calculation.

3.11. **The dual group.** The product of two elements $\gamma, \lambda \in \widehat{G}$ is their point-wise product. It is easy to see that this is, again, a homomorphism:

$$(\gamma\lambda)(0_G) = \gamma(0_G)\lambda(0_G) = 1^2 = 1$$

and

$$\begin{aligned}(\gamma\lambda)(x+y) = \gamma(x+y)\lambda(x+y) &= \gamma(x)\gamma(y)\lambda(x)\lambda(y)\\ &= \gamma(x)\lambda(x)\gamma(y)\lambda(y) = (\gamma\lambda)(x)(\gamma\lambda)(y)\end{aligned}$$

for all $x, y \in G$. More than this, since multiplication of complex numbers is commutative, the multiplication on $\widehat{G}$ is commutative.

⚠**Warning** ⚠ *The group $\widehat{G}$ is an Abelian group so we shall denote its identity, that is the constant function equal to* 1*, as an additive identity, $0_{\widehat{G}}$. This means that $0_{\widehat{G}}(x) = 1$ for all $x \in G$.*

By construction the measure μ^* is a Haar measure on $\widehat{G}$ and there is a duality between (G, μ) and $(\widehat{G}, \mu^*)$ which it will be informative to draw out. It essentially identifies the dual of $(\widehat{G}, \mu^*)$ with (G, μ), and is very close to the duality between a Hilbert space H and its continuous double dual H^{**}.

Theorem 3.12 (Pontryagin duality)**.** *The map*

$$\phi_{G,\widehat{G}} : G \to \widehat{\widehat{G}}; x \mapsto (\gamma \mapsto \gamma(x))$$

is an isomorphism from G to $\widehat{\widehat{G}}$ such that $(\mu^*)^* \circ \phi_{G,\widehat{G}} = \mu$.

Proof. The fact that $\phi_{G,\widehat{G}}$ is a homomorphism is an easy check. To see that it is injective we import the injectivity of the Fourier-Stieltjes transform. Note that if $x \in \ker \phi_{G,\widehat{G}}$ then

$$\widehat{\delta_x}(\gamma) = \int \overline{\gamma(z)} d\delta_x(z) = \overline{\gamma(x)} = 1 \text{ for all } \gamma \in \widehat{G}.$$

Since $0_G \in \ker \phi_{G,\widehat{G}}$ it follows from the linearity of the Fourier transform that $(\delta_x - \delta_{0_G})^\wedge \equiv 0$ and hence, by the uniqueness in Theorem 3.8, that $\delta_x - \delta_{0_G} \equiv 0$ and so $x = 0_G$ as required.

The work now is to show that $\phi_{G,\widehat{G}}$ is a surjection. Suppose that $A \in \widehat{\widehat{G}}$ which we consider as a function on $\widehat{G}$. This time we use surjectivity of the Fourier transform and let f be such that $\widehat{f}(\gamma) = A(\gamma)$ for all $\gamma \in \widehat{G}$. Then by the inversion formula, the fact that A is a homomorphism and translation invariance of μ^* we have that

$$f(x) = \int A(\gamma)\gamma(x) d\mu^*(\gamma) = \int A(\gamma\lambda)(\gamma\lambda)(x) d\mu^*(\gamma) = A(\lambda)\lambda(x)f(x)$$

for all $x \in G$. Since f is not identically zero there is some $x \in G$ such that $f(x) \neq 0$, and dividing it follows that $A(\lambda) = \lambda(-x)$ and we see that $\phi_{G,\widehat{G}}$ is surjective as required.

Finally, we need to address the measure. The measure was design to make the inversion theorem (and Parseval's theorem) work so it is natural to check these using the Pontryagin isomorphism. Suppose that $g \in L^1(\mu)$ and write $f := \widehat{g}$. By the Fourier inversion formula on G we conclude that

$$g(x) = \int f(\gamma)\gamma(x)d\mu^*(\gamma).$$

On the other hand by definition of the Fourier transform on $\widehat{G}$ we have that

$$\widehat{f}(\phi_{G,\widehat{G}}(x)) = \int f(\gamma)\overline{\phi_{G,\widehat{G}}(x)(\gamma)}d\mu^*(\gamma) = \int f(\gamma)\overline{\gamma(x)}d\mu^*(\gamma),$$

and so $\widehat{f}(\phi_{G,\widehat{G}}(x)) = g(-x)$. Now, by the Fourier inversion formula on $\widehat{G}$, the fact that $\phi_{G,\widehat{G}}$ is an isomorphism and μ^{**} is a Haar measure (so $\mu^{**}(E) = \mu^{**}(-E)$) we have

$$\begin{aligned} f(\gamma) = \int \widehat{f}(A)A(\gamma)d\mu^{**}(A) &= \int \widehat{f}(\phi_{G,\widehat{G}}(x))\gamma(x)d\mu^{**}\circ\phi_{G,\widehat{G}}(x) \\ &= \int g(-x)\gamma(x)d\mu^{**}\circ\phi_{G,\widehat{G}}(x) \\ &= \int g(x)\overline{\gamma(x)}d\mu^{**}\circ\phi_{G,\widehat{G}}(x). \end{aligned}$$

On the other hand $f = \widehat{g}$, and so

$$\int g(x)\overline{\gamma(x)}d\mu(x) = \int g(x)\overline{\gamma(x)}d\mu^{**}\circ\phi_{G,\widehat{G}}(x) \text{ for all } \gamma \in \widehat{G}.$$

It follows that $gd\mu = gd\mu^{**}\circ\phi_{G,\widehat{G}}$ by the uniqueness of the Fourier-Stieltjes transform, but then $g \in L^1(\mu)$ was arbitrary and so $\mu = \mu^{**}\circ\phi_{G,\widehat{G}}$ as required. □

Note that in the proof of this duality we saw that

$$f^\vee(-x) = \widehat{\widehat{f}}(\phi_{G,\widehat{G}}(x)) \text{ for all } x \in G,$$

and so the inverse Fourier transform is (up to a minus sign) also a Fourier transform.

The duality expressed above extends to subgroups and quotient groups. To describe this we make a further definition which will be very useful in the sequel.

Definition 3.13 (Annihilators)**.** Given a set $A \subset G$ we write $A^\perp$ for its *annihilator* defined by

$$A^\perp := \{\gamma \in \widehat{G} : \gamma(x) = 1 \text{ for all } x \in A\}.$$

It is easy to check that annihilators are subgroups, and by Pontryagin duality we see that there is an isomorphism

$$\Gamma^\perp \xrightarrow{\phi_{G,\widehat{G}}} \{x \in G : \gamma(x) = 1 \text{ for all } \gamma \in \Gamma\},$$

and so if $V \leqslant G$ then there is an isomorphism $V^{\perp\perp} \xrightarrow{\phi_{G,\widehat{G}}} V$.

The aspect of duality which annihilators encode is that the annihilator of a subgroup is naturally isomorphic to the dual of the quotient group. Indeed, suppose that $V \leqslant G$ then

$$V^\perp \rightarrow \widehat{G/V}; \gamma \mapsto \quad \begin{aligned} \gamma' : \ & G/V \rightarrow S^1 \\ & x+V \mapsto \gamma(x) \end{aligned}$$

is a well-defined isomorphism.

Annihilators come up explicitly in the Fourier transform of measures as we shall now see.

Example 3.14 (The Fourier transform of subgroup measures)**.** Given $V \leqslant G$ we have that $\widehat{\mu_V}(\gamma) = 1_{V^\perp}$. To see this first note that $\gamma \in V^\perp$ if and only if $\widehat{\mu_V}(\gamma) = 1$. Indeed,

$$\operatorname{Re} \widehat{\mu_V}(\gamma) = \int \operatorname{Re} \overline{\gamma(x)} d\mu_V(x) \leqslant \int d\mu_V(x) = 1$$

with equality if and only if $\operatorname{Re}\overline{\gamma(x)} = 1$ for all $x \in V$, which in turn is true if and only if $\gamma(x) = 1$ for all $x \in V$. Now, from the calculation in Example 2.10 we see that $\widehat{\mu_V} = (\mu_V * \mu_V)^\wedge = \widehat{\mu_V}^2$ so that $\widehat{\mu_V}$ can only take the values 0 and 1. Thus, if $\widehat{\mu_V}(\gamma) \neq 0$ then $\widehat{\mu_V}(\gamma) = 1$ and $\gamma \in V^\perp$, and conversely. The claimed equality follows.

Using Parseval's theorem we can establish a little bit more: we have $(1_V)d\mu = \mu(V)d\mu_V$, whence $\widehat{1_V}(\gamma) = \mu(V)1_{V^\perp}$. But then by Parseval's theorem we have

$$\mu(V) = \|1_V\|_{L^2(\mu)}^2 = \|\widehat{1_V}\|_{L^2(\mu^*)}^2 = \mu(V)^2\mu^*(V^\perp),$$

and so

$$\mu(V)\mu^*(V^\perp) = 1.$$

The last equality above represents a critical case of the well-known uncertainty principle. We turn to this now.

Example 3.15 (Uncertainty principle)**.** This states that for all $f \not\equiv 0$ we have

$$\mu(\operatorname{supp} f)\mu^*(\operatorname{supp} \widehat{f}) \geqslant 1.$$

Roughly speaking, a (non-zero) function cannot be simultaneously concentrated on G (physical space) and $\widehat{G}$ (momentum space).

To prove the result we start by applying the triangle inequality and a special case of the Hausdorff-Young inequality (that is (3.4)) to get that

$$\|\widehat{f}\|_{L^2(\mu^*)}^2 \leqslant \|\widehat{f}\|_{L^1(\mu^*)}\|f\|_{L^1(\mu)}.$$

Now by the Cauchy-Schwarz inequality we have that

$$\|\widehat{f}\|_{L^1(\mu^*)} \leqslant \|\widehat{f}\|_{L^2(\mu^*)}\mu^*(\operatorname{supp} \widehat{f})^{1/2} \text{ and } \|f\|_{L^1(\mu)} \leqslant \|f\|_{L^2(\mu)}\mu(\operatorname{supp} f)^{1/2}.$$

But then by Parseval's theorem we conclude that

$$\begin{aligned} \|\widehat{f}\|_{L^2(\mu^*)}^2 &\leqslant \|\widehat{f}\|_{L^2(\mu^*)}\mu^*(\operatorname{supp} \widehat{f})^{1/2}.\|f\|_{L^2(\mu)}\mu(\operatorname{supp} f)^{1/2} \\ &= \|\widehat{f}\|_{L^2(\mu^*)}^2\mu^*(\operatorname{supp} \widehat{f})^{1/2}\mu(\operatorname{supp} f)^{1/2}. \end{aligned}$$

The claimed inequality follows on dividing.

Given $V \leqslant G$ our previous calculations (in Example 3.14) show that taking $f = 1_V$ the inequality is tight since $\operatorname{supp} \widehat{1_V} = V^{\perp}$. In fact more than this, if $\gamma \in \widehat{G}$ and $x \in G$ then $f := \gamma 1_{x+V}$ has

$$\widehat{f}(\gamma') = \gamma(x)\overline{\gamma'(x)}1_{V^{\perp}}(\gamma - \gamma'),$$

so

$$\operatorname{supp} f = x + V \text{ and } \operatorname{supp} \widehat{f} = \gamma + V^{\perp},$$

and hence equality holds in the uncertainty principle for f as well.

It turns out that the uncertainty principle is, in fact, robustly true in the sense that any function which is close to tight for this inequality is in some sense close to functions of this type. Indeed, suppose that $f \not\equiv 0$ is such that

$$\mu(\operatorname{supp} f)\mu^*(\operatorname{supp} \widehat{f}) \leqslant 1 + \eta$$

for some $\eta > 0$ sufficiently small. We shall show that there is some subgroup $H \leqslant \widehat{G}$ such that[15]

$$\mu^*((\operatorname{supp} \widehat{f})\triangle(\gamma + H)) = O(\sqrt{\eta}\mu^*(H)) \text{ for } \eta \text{ sufficiently small.}$$

We could equally well have looked at $\operatorname{supp} f$, but it will be slightly more notationally convenient for us to do it this way round. We shall follow an argument of Fournier [Fou77] for Young's inequality. Indeed, showing the above result is essentially equivalent to analysing the case when Young's inequality is close to critical.

We start by estimating the size of $S := \operatorname{supp} \widehat{f}$. From the Cauchy-Schwarz inequality as before we see that

$$\mu(\operatorname{supp} f) \geqslant \|f\|_{L^1(\mu)}^2/\|f\|_{L^2(\mu)}^2,$$

which inserted into our hypothesis tells us that

$$\mu^*(S) \leqslant (1 + \eta)\|f\|_{L^2(\mu)}^2/\|f\|_{L^1(\mu)}^2. \tag{3.5}$$

Of course, we expect this to be tight since it follows the inequalities we used to derive the uncertainty principle itself.[16] Now, by Plancherel's theorem and the Cauchy-Schwarz

[15] Recall that $\triangle$ denotes symmetric difference so that $A\triangle B = (A\backslash B) \cup (B\backslash A)$; in words $A\triangle B$ is the set of elements in exactly one of A and B.

[16] To see that it is tight in the sense that $\mu^*(S)$ is close to this upper bound, apply Parseval's theorem and the triangle inequality to get that

$$\begin{aligned} \|f\|_{L^2(\mu)}^2 = \|\widehat{f}\|_{L^2(\mu*)}^2 &= \int 1_S(\gamma)|\widehat{f}(\gamma)|^2 d\gamma \\ &= \int 1_S(\gamma)\left|\int f(x)\overline{\gamma(x)}dx\right|^2 d\gamma \leqslant \mu^*(S)\|f\|_{L^1(\mu)}^2. \end{aligned}$$

It follows that $\mu(S) \geqslant \|f\|_{L^2(\mu)}^2/\|f\|_{L^1(\mu)}^2$ which complements our inequality up to a factor of $1 + \eta$.

inequality (twice) we have

$$\begin{aligned}\|f\|_{L^2(\mu)}^2 = \int 1_S(\gamma)|\widehat{f}(\gamma)|^2 d\gamma &= \int f * \tilde{f}(x) 1_S^\vee(x) dx \\ &\leqslant \left(\int |f * \tilde{f}(x)| dx\right)^{1/2} \left(\int |f * \tilde{f}(x)||1_S^\vee(x)|^2 dx\right)^{1/2} \\ &\leqslant \left(\int |f * \tilde{f}(x)| dx\right)^{1/2} \\ &\quad \times \left(\int |f * \tilde{f}(x)|^2 dx\right)^{1/4} \left(\int |1_S^\vee(x)|^4 dx\right)^{1/4} \\ &= \|f * \tilde{f}\|_{L^1(\mu)}^{1/2} \|f * \tilde{f}\|_{L^2(\mu)}^{1/2} \|1_S^\vee\|_{L^4(\mu)}.\end{aligned}$$

Young's inequality provides estimates for these norms of $f * \tilde{f}$:

$$\|f * \tilde{f}\|_{L^2(\mu)} \leqslant \|f\|_{L^2(\mu)} \|\tilde{f}\|_{L^1(\mu)} = \|f\|_{L^2(\mu)} \|f\|_{L^1(\mu)}$$

and

$$\|f * \tilde{f}\|_{L^1(\mu)} \leqslant \|f\|_{L^1(\mu)} \|\tilde{f}\|_{L^1(\mu)} = \|f\|_{L^1(\mu)}^2.$$

Inserting these bounds, raising to the fourth power and using (3.5) we get that

$$\|1_S^\vee\|_{L^4(\mu)}^4 \geqslant \frac{\|f\|_{L^2(\mu)}^6}{\|f\|_{L^1(\mu)}^6} \geqslant \frac{1}{(1+\eta)^3} \mu^*(S)^3 \geqslant (1-3\eta)\mu^*(S)^3.$$

On the other hand by Parseval's theorem we have that

$$\|1_S * 1_{-S}\|_{L^2(\mu^*)}^2 = \|1_S^\vee\|_{L^4(\mu)}^4 \geqslant (1-3\eta)\mu^*(S)^3. \tag{3.6}$$

At this point we recall that

$$1_S * 1_{-S}(\gamma) = \mu^*(S \cap (\gamma + S)) \text{ for all } \gamma \in \widehat{G}.$$

This immediately tells us that (3.6) is tight.[17] More than this it turns out that the set of characters at which $1_S * 1_{-S}$ is large is a large subgroup, and it is proving this to which we turn our attention.[18] We write

$$H_\epsilon := \{\gamma \in \widehat{G} : 1_S * 1_{-S}(\gamma) \geqslant (1-\epsilon)\mu^*(S)\},$$

where ϵ (to be thought of as small) is a parameter to be optimised later. Translation by characters in γ does not vary $1_S * 1_{-S}$ very much. In particular we have the following claim.

[17] Indeed, note that

$$\|1_S * 1_{-S}\|_{L^2(\mu^*)}^2 \leqslant \|1_S * 1_{-S}\|_{L^\infty(\mu^*)} \|1_S * 1_{-S}\|_{L^1(\mu^*)} = \mu^*(S).\mu^*(S)^2 = \mu^*(S)^3.$$

[18] It should not come as a total surprise that this is the case: an easy calculation shows that if T is very close to being a (coset of a) subgroup then the set of characters where $1_T * 1_{-T}$ is large is *exactly* a subgroup.

Claim 1. *If $\gamma \in H_\epsilon$ then*

$$|1_S * 1_{-S}(\gamma+\gamma') - 1_S * 1_{-S}(\gamma')| \leqslant \epsilon\mu^*(S) \text{ for all } \gamma' \in \widehat{G}.$$

Proof. First note that

$$\begin{aligned} 1_S * 1_{-S}(\gamma+\gamma') - 1_S * 1_{-S}(\gamma') &= \mu^*(S \cap (\gamma+\gamma'+S)) - \mu^*(S \cap (\gamma'+S)) \\ &= \mu^*((S-\gamma) \cap (\gamma'+S)) - \mu^*(S \cap (\gamma'+S)) \\ &\leqslant \mu^*(((S-\gamma)\backslash S) \cap (\gamma'+S)) \\ &\quad -\mu^*((S\backslash(S-\gamma)) \cap (\gamma'+S)). \end{aligned}$$

It follows that

$$\mu^*((S-\gamma)\backslash S) \geqslant 1_S * 1_{-S}(\gamma+\gamma') - 1_S * 1_{-S}(\gamma') \geqslant -\mu^*(S\backslash(S-\gamma)),$$

and hence

$$\mu^*(S) - 1_S * 1_{-S}(\gamma) \geqslant 1_S * 1_{-S}(\gamma+\gamma') - 1_S * 1_{-S}(\gamma') \geqslant 1_S * 1_{-S}(-\gamma) - \mu^*(S).$$

Finally, by symmetry we get the claim. □

We now have the three main claims which help us show that the set of characters at which $1_S * 1_{-S}$ is large is a large subgroup.

Claim 2.

$$\mu^*(H_\epsilon) \geqslant (1-3\eta\epsilon^{-1})\mu^*(S).$$

Proof. Since $1_S * 1_{-S}(\gamma) \leqslant \mu^*(S)$, by the definition of H_ϵ we have

$$\begin{aligned} \mu^*(S)\int_{H_\epsilon} 1_S * 1_{-S}(\gamma)d\gamma + (1-\epsilon)\mu^*(S)\int_{G\backslash H_\epsilon} 1_S * 1_{-S}(\gamma)d\gamma &\geqslant \int 1_S * 1_{-S}(\gamma)^2 d\gamma \\ &\geqslant (1-3\eta)\mu^*(S)^3. \end{aligned}$$

However,

$$\int_{G\backslash H_\epsilon} 1_S * 1_{-S}(\gamma)d\gamma = \int_G 1_S * 1_{-S}(\gamma)d\gamma - \int_{H_\epsilon} 1_S * 1_{-S}(\gamma)d\gamma,$$

whence

$$\epsilon\mu^*(S)\int_{H_\epsilon} 1_S * 1_{-S}(\gamma)d\gamma \geqslant (\epsilon-3\eta)\mu^*(S)^3;$$

it follows that

$$\mu^*(H_\epsilon) \geqslant (1-3\eta\epsilon^{-1})\mu^*(S)$$

again using the fact that $1_S * 1_{-S}(\gamma) \leqslant \mu^*(S)$. □

We get the group structure in two steps.

Claim 3. *If $H_{4\epsilon}\backslash H_{2\epsilon}$ is empty then $H_{2\epsilon}$ is a subgroup of $\widehat{G}$.*

Proof. First note that $H_{2\epsilon}$ is non-empty since it contains $0_{\widehat{G}}$. Now, suppose $\gamma, \gamma' \in H_{2\epsilon}$. Then by symmetry $-\gamma' \in H_{2\epsilon}$ and by Claim 1 we have

$$\begin{aligned}|1_S * 1_{-S}(\gamma - \gamma') - \mu^*(S)| &\leqslant |1_S * 1_{-S}(\gamma - \gamma') - 1_S * 1_{-S}(\gamma)| \\ &\quad + |1_S * 1_{-S}(\gamma) - \mu^*(S)| \leqslant 4\epsilon\mu^*(S)\end{aligned}$$

and so $\gamma - \gamma' \in H_{4\epsilon}$. Since $H_{4\epsilon} \backslash H_{2\epsilon}$ is empty we conclude that $\gamma - \gamma' \in H_{2\epsilon}$ and it is a subgroup as required. □

Our final claim is aimed at satisfying the hypotheses of the previous one.

Claim 4. *The set $H_{1-\nu} \backslash H_{2\epsilon}$ is empty provided* $1 < (1 + \nu - 2\epsilon)(1 - 3\eta\epsilon^{-1})$.

Proof. Suppose, for a contradiction, that $\gamma' \in H_{1-\nu} \backslash H_{2\epsilon}$ so that

$$(1 - 2\epsilon)\mu^*(S) > 1_S * 1_{-S}(\gamma') \geqslant \nu\mu^*(S).$$

It follows from Claim 1 that if $\gamma \in H_\epsilon$ then $\gamma + \gamma' \in H_{1-\nu+\epsilon} \backslash H_\epsilon$, which is to say

$$(1 - \epsilon)\mu^*(S) > 1_S * 1_{-S}(\gamma + \gamma') \geqslant (\nu - \epsilon)\mu^*(S).$$

We conclude that

$$\int_{H_{1-\nu+\epsilon} \backslash H_\epsilon} 1_S * 1_{-S}(\gamma) d\gamma \geqslant (\nu - \epsilon)\mu^*(S)\mu^*(H_{1-\nu+\epsilon} \backslash H_\epsilon) \geqslant (\nu - \epsilon)\mu^*(S)\mu^*(H_\epsilon),$$

since $H_{1-\nu+\epsilon} \backslash H_\epsilon$ contains a translate of H_ϵ as a result of the hypothesis that $H_{1-\nu} \backslash H_{2\epsilon}$ is non-empty. On the other hand

$$\int_{H_{1-\nu+\epsilon} \backslash H_\epsilon} 1_S * 1_{-S}(\gamma) d\gamma + \int_{H_\epsilon} 1_S * 1_{-S}(\gamma) d\gamma \leqslant \int 1_S * 1_{-S}(\gamma) d\gamma = \mu^*(S)^2,$$

and so

$$\begin{aligned}\mu^*(S)^2 &\geqslant (\nu - \epsilon)\mu^*(S)\mu^*(H_\epsilon) + (1 - \epsilon)\mu^*(S)\mu^*(H_\epsilon) \\ &= (1 + \nu - 2\epsilon)\mu^*(S)\mu^*(H_\epsilon) \geqslant (1 + \nu - 2\epsilon)(1 - 3\eta\epsilon^{-1})\mu^*(S)^2\end{aligned}$$

by Claim 2, which contradicts the hypothesis proving this claim. □

We are now in a position to piece together our work. First we take $\nu := 1 - 4\epsilon$ and (provided η is sufficiently small) arrange ϵ such that[19]

$$1 < (1 + (1 - 4\epsilon) - 2\epsilon)(1 - 3\eta\epsilon^{-1}).$$

Given this we know from Claim 4 that $H_{4\epsilon} \backslash H_{2\epsilon}$ is empty and so by Claim 3 we have that $H := H_{2\epsilon}$ is a group. Moreover, $H \supset H_\epsilon$ and so by Claim 2 we conclude that

$$\int 1_S * 1_{-S}(\gamma) 1_H(\gamma) d\gamma \geqslant \int_{H_\epsilon} 1_S * 1_{-S}(\gamma) d\gamma \geqslant (1 - \epsilon)(1 - 3\eta\epsilon^{-1})\mu^*(S)^2.$$

[19] This is possible if, for example, we can take $\epsilon > 0$ such that $\epsilon + \eta\epsilon^{-1} < 1/6$.

On the other hand

$$\begin{aligned}\int 1_S * 1_{-S}(\gamma)1_H(\gamma)d\gamma &= \langle 1_S * 1_{-S}, 1_H\rangle_{L^2(\mu^*)}\\ &= \langle 1_S, 1_H * 1_S\rangle_{L^2(\mu^*)} \leqslant \mu^*(S)\|1_S * 1_H\|_{L^\infty(\mu^*)}.\end{aligned}$$

We conclude that there is some γ' such that

$$\mu^*(S \cap (\gamma' + H)) = 1_S * 1_H(\gamma') \geqslant (1-\epsilon)(1-3\eta\epsilon^{-1})\mu^*(S).$$

Additionally,

$$(1-2\epsilon)\mu^*(S)\mu^*(H) \leqslant \int_H 1_S * 1_{-S}(\gamma)d\gamma \leqslant \mu^*(S)^2,$$

so that (using Claim 2 for the left hand side) we have

$$(1-3\eta\epsilon^{-1})\mu^*(S) \leqslant \mu^*(H) \leqslant (1-2\epsilon)^{-1}\mu^*(S).$$

It follows from this that

$$\mu^*(S\triangle(\gamma' + H)) = \mu^*(S) + \mu^*(\gamma + H) - 2\mu^*(S \cap (\gamma' + H)) = O(\epsilon + \eta\epsilon^{-1})\mu^*(H),$$

and optimising the choice of ϵ gives the result.

Example 3.16 (The Fourier transform of random sets, Example 2.11 contd.)**.** Suppose again that each $x \in G$ is placed in the set A independently with probability $\alpha = \Omega(1)$. The phase of $\widehat{1_A}$ is hard to determine and in general we shall not try; the magnitude however is easier to estimate.

If $\gamma = 0_{\widehat{G}}$ then we have $\mathbb{E}|\widehat{1_A}(\gamma)| = \alpha$, but it turns out that for other characters we expect the magnitude to be small. Indeed, suppose that $\gamma \neq 0_{\widehat{G}}$. Then

$$\begin{aligned}\mathbb{E}|\widehat{1_A}(\gamma)| \leqslant \left(\mathbb{E}|\widehat{1_A}(\gamma)|^2\right)^{1/2} &= \left(\mathbb{E}\left|\mathbb{E}_{x\in G}1_A(x)\overline{\gamma(x)}\right|^2\right)^{1/2}\\ &= \left(\mathbb{E}_{x,y\in G}\mathbb{E}1_A(x)1_A(y)\gamma(x-y)\right)^{1/2}\\ &= \left(\mathbb{E}_{x,y\in G}\alpha^2\gamma(x-y) - \frac{\alpha^2}{|G|} + \frac{\alpha}{|G|}\right)^{1/2}\\ &= \left(\frac{\alpha-\alpha^2}{|G|}\right)^{1/2} = O(1/\sqrt{|G|}).\end{aligned}$$

In fact using the variance calculation from Example 2.11 one can show that most of the time $|\widehat{1_A}(\gamma)|$ is about this size, and in fact we shall see later that

$$\sup_{\gamma\neq 0_{\widehat{G}}} |\widehat{1_A}(\gamma)| = O\left(\frac{\log|G|}{|G|}\right)^{1/2} \text{ w.h.p.}$$

We now turn to the final example of this chapter which reflects another powerful way in which the Fourier transform may be used.

Example 3.17 (Spectral gap and convergence in distribution)**.** Suppose that $A \subset G$ is non-empty. We set up a homogenous random walk on G as follows. Suppose that at stage i we have a G-valued random variable X_i with law[20] μ_i – we think of X_i as being our position after i steps of the random walk. We assume the distribution of X_0 – the initial position – is given by some (arbitrary) probability measure $\mu_0 \in M(G)$:

$$\mathbb{P}(X_0 \in S) = \mu_0(S) \text{ for all } S \subset G.$$

Now, at stage i we move to $X_i + a$ with probability $|A|^{-1}$ for each $a \in A$. This translates to

$$\mathbb{P}(X_{i+1} \in S) = \int 1_S(x+a)d\mu_A(a)d\mu_i(x) = \mu_i * \mu_A(S).$$

We conclude that $\mu_{i+1} = \mu_i * \mu_A$, and ask whether the walk converges. If A is chosen poorly, for example if it is a non-zero coset of a subgroup of G, then the walk may oscillate between cosets and not, in general, converge. To avoid this we make the technical convenience of assuming that $0_G \in A$.

Given our assumption intuitively we expect the walk to converge: the set A generates some subgroup H and we expect after enough time that the random walk will have 'averaged' our initial distribution over cosets of H. This means that the real question of interest if how *fast* the walk converges.

To measure the difference between two distributions we use a quantity called total variation distance: suppose that ν and σ are two probability measures. Then their *total variation distance* is

$$\tau(\nu,\sigma) := \sup\{|\nu(S) - \sigma(S)| : S \subset G\}.$$

This quantity can be easily expressed in terms of the norm of the difference of the two measures. First, for any set $S \subset G$ we have

$$0 = \int (1_S + 1_{G\setminus S})d(\nu - \sigma),$$

so it follows that

$$|\nu(S) - \sigma(S)| = \frac{1}{2}\left|\int (1_S - 1_{G\setminus S})d(\nu-\sigma)\right| \leqslant \frac{1}{2}\|\nu - \sigma\|.$$

On the other hand taking

$$S := \{x : \nu(\{x\}) \geqslant \sigma(\{x\})\},$$

we see that this upper bound is achieved (since $\operatorname{sgn}(\nu - \sigma) = 1_S - 1_{G\setminus S}$) so that

$$\tau(\nu,\sigma) = \frac{1}{2}\|\nu - \sigma\|.$$

Returning to our random walk it may be instructive to consider a concrete example: take $G = \mathbb{Z}/N\mathbb{Z}$ and $A = \{0,1\}$ (so that A generates the whole of G). Suppose that X_0 is

[20] Recall that the *law* of a random variable X is the measure μ such that $\mu(S) = \mathbb{P}(X \in S)$.

the δ-distribution centred at 0_G. Then after k steps we have

$$\mathbb{P}(X_k = r) = 2^{-k} \sum_{Nl+r \leqslant k} \binom{k}{Nl+r}.$$

It follows that

$$\begin{aligned} \mathbb{P}(X_k = r) &= 2^{-k}\left(\sum_{Nl+r\leqslant k/2}\binom{k}{Nl+r} + \sum_{k/2<Nl+r\leqslant k}\binom{k}{Nl+r}\right)\\ &\leqslant 2^{-k}\left(\sum_{Nl\leqslant k/2}\binom{k}{N(l+1)} + \sum_{k/2<Nl\leqslant k}\binom{k}{Nl}\right)\\ &\leqslant \mathbb{P}(X_k=0) + 2^{-k}\binom{k}{k/2} = \mathbb{P}(X_k=0)+O(1/\sqrt{k}), \end{aligned}$$

where the last equality is by Stirling's formula. A similar argument shows that $\mathbb{P}(X_k = r) \geqslant \mathbb{P}(X_k = 0) - O(1/\sqrt{k})$ and so it follows that

$$\tau(\mu_k, \mu_G) = \sup\{|\mathbb{P}(X_k \in S) - \mu_G(S)| : S \subset G\} = O(|G|/\sqrt{k}) \tag{3.7}$$

so that after $O(\epsilon^{-2}|G|^2)$ steps we have $\tau(\mu_k, \mu_G) \leqslant \epsilon$. The $|G|$ dependence here is essentially tight as can be seen by some more careful analysis; the ϵ dependence can be improved and the whole approach can be set in a more general frame. This is the task to which we turn and for which we use the Fourier transform.

Returning to our general random walk we are interested in an upper bound on

$$\tau(\mu_k, \mu_0 * \mu_H) = \frac{1}{2}\|\mu_k - \mu_0 * \mu_H\|,$$

where, as mentioned, H is the subgroup of G generated by A. This is measuring the total variation distance between the distribution of our random walk after k steps and our initial distribution averaged over cosets of H. We write f_0 for the probability mass function associated with μ_0 then we see that

$$\tau(\mu_k, \mu_0 * \mu_H) = \frac{1}{2}\|f_0 * \mu_A^{(k)} - f_0 * \mu_H\|_{\ell^1(G)} \leqslant \frac{1}{2}\|f_0 * \mu_A^{(k)} - f_0 * \mu_H\|_{\ell^2(G)}|G|^{1/2}$$

by Cauchy-Schwarz, where $\mu_A^{(k)}$ is the k-fold convolution of μ_A with itself. Now we examine the ℓ^2-norm using Parseval's theorem (thinking of G as being endowed with counting measure):

$$\|f_0 * \mu_A^{(k)} - f_0 * \mu_H\|_{\ell^2(G)}^2 = \mathbb{E}_{\gamma\in\widehat{G}}|\widehat{f_0}(\gamma)|^2|\widehat{\mu_A}(\gamma)^k - \widehat{\mu_H}(\gamma)|^2.$$

By Example 3.14 we have that $\widehat{\mu_H} = 1_{H^\perp}$. On the other hand, $A \subset H$ and so if $\gamma \in H^\perp$ then $\widehat{\mu_A}(\gamma) = 1$. It follows that

$$\|f_0 * \mu_A^{(k)} - f_0 * \mu_H\|_{\ell^2(G)}^2 = \mathbb{E}_{\gamma\in\widehat{G}}1_{\widehat{G}\setminus H^\perp}(\gamma)|\widehat{f_0}(\gamma)|^2|\widehat{\mu_A}(\gamma)|^{2k}.$$

We say that A has an[21] *ϵ-spectral gap* if

$$\sup_{\gamma\notin A^{\perp}} |\widehat{\mu_A}(\gamma)| \leqslant 1-\epsilon, \tag{3.8}$$

that is to say there is a gap of size ϵ in the values of $\widehat{\mu_A}$ between when it is 1 and when it is less than 1. (Note that $A^{\perp}=H^{\perp}$ since A generates H.) It we talk about *the* spectral gap of A we mean the largest $\epsilon \in [0,1]$ such that (3.8) holds.

If A has ϵ-spectral gap then we see that

$$\|f_0 * \mu_A^{(k)} - f_0 * \mu_H\|_{\ell^2(G)}^2 \leqslant \mathbb{E}_{\gamma\in\widehat{G}} 1_{\widehat{G}\setminus H^{\perp}}(\gamma)|\widehat{f_0}(\gamma)|^2(1-\epsilon)^{2k},$$

and it follows that

$$\tau(\mu_k, \mu_0 * \mu_H) \leqslant \frac{1}{2}\|f_0\|_{\ell^2(G)}(1-\epsilon)^k|G|^{1/2} = O(|G|^{1/2}\exp(-O(\epsilon k))). \tag{3.9}$$

Roughly we expect to have achieved ‘good convergence’ after at most $O(\epsilon^{-1}\log|G|)$ steps of the walk.

It is trivial that A has some non-zero spectral gap: if A is a subgroup then we see that it has spectral gap[22] 1; if A is not a subgroup, suppose that $\gamma \notin A^{\perp}$. Then there is some $a_0 \in A$ such that $\gamma(a_0) \neq 1$. Also, $0_G \in A$ so that

$$1 - |\widehat{\mu_A}(\gamma)|^2 = \frac{1}{|A|^2}\sum_{a,a'\in A} 1 - \operatorname{Re}\gamma(a-a') \geqslant \frac{1}{|A|^2}2(1-\operatorname{Re}\gamma(a_0)).$$

Since γ maps from G we have that the order of $\gamma(a_0)$ divides $|G|$ and hence

$$2(1-\operatorname{Re}\gamma(a_0)) = |1-\gamma(a_0)|^2 \geqslant |1-\exp(2\pi i/|G|)|^2 = \Omega(|G|^{-2}).$$

It follows that A has $\Omega(1/|A|^2|G|^2)$-spectral gap; in fact it is somewhat better.[23]

Before proceeding it is worth noting that if $|A| = O(1)$ then the above tells us that A has spectral gap $\Omega(1/|G|^2)$ and so by (3.9) we have good convergence after at most $O(|G|^2\log|G|)$ steps. This should be compared with (3.7).

We now return to the general setting of G a finite Abelian group and $A \subset G$ a set containing the identity.

Claim. *Suppose that A has density α and $\lambda \in \widehat{G}$. Then at least one of the following is true:*

(i) (Fourier coefficient is small) $|\widehat{\mu_A}(\lambda)| \leqslant 1-\Omega(\alpha^2)$*;*

(ii) (Many pairs in kernel) *for at least* $49/64$ *of the pairs* $(a,a') \in A^2$ *have* $a-a' \in \ker\lambda$.

Proof. Our argument is, in a sense, dual to some of the work we did in Example 3.15, and our analogue of the sets H_ϵ from there are the sets

$$S_\epsilon := \{\gamma \in \widehat{G} : \|1-\gamma\|_{L^2(\mu_A*\mu_{-A})} \leqslant \epsilon\}.$$

[21] This is a slightly unorthodox definition of spectral gap. When the spectral gap is usually defined to be $\inf_{\gamma\neq 0_{\widehat{G}}} 1-\operatorname{Re}\widehat{\mu_A}(\gamma)$. The exercises have some more material on this.

[22] Waffle about infinity.

[23] We proved something a bit better in lectures; we shall prove something even better in a moment using the main claim from the lectures.

The reason for this slightly odd definition is because the triangle inequality for L^2-norms tells us immediately that if $\gamma \in S_\epsilon$ and $\gamma' \in S_\delta$ then

$$\begin{aligned}\|1-\gamma\gamma'\|_{L^2(\mu_A*\mu_{-A})} &= \|1-\gamma+\gamma(1-\gamma')\|_{L^2(\mu_A*\mu_{-A})}\\ &\leqslant \|1-\gamma\|_{L^2(\mu_A*\mu_{-A})}+\|\gamma(1-\gamma')\|_{L^2(\mu_A*\mu_{-A})}\\ &= \|1-\gamma\|_{L^2(\mu_A*\mu_{-A})}+\|1-\gamma'\|_{L^2(\mu_A*\mu_{-A})} \leqslant \epsilon+\delta,\end{aligned}$$

so that $\gamma+\gamma' \in S_{\epsilon+\delta}$. On the other hand the definition relates to spectral gap in the sense that

$$\|1-\gamma\|^2_{L^2(\mu_A*\mu_{-A})} = \int |1-\gamma(x)|^2 d\mu_A*\mu_{-A}(x) = 2-2|\widehat{\mu_A}(\gamma)|^2,$$

so that

$$|\widehat{\mu_A}(\gamma)| \geqslant 1-\epsilon^2/2 \text{ if and only if } \|1-\gamma\|_{L^2(\mu_A*\mu_{-A})} \leqslant \epsilon.$$

We shall let ϵ be a constant to be optimised later (it will turn out that $\epsilon = \Omega(\alpha)$). Our aim is to show that if $\lambda \in S_\epsilon$ and ϵ is sufficiently small then we have the second conclusion.

Suppose that $\lambda \in S_\epsilon$. If $r \leqslant \epsilon^{-1}/2$ then by the triangle inequality we have that

$$\gamma' \in S_{(r-1)\epsilon} \Rightarrow \lambda+\gamma' \in S_{r\epsilon}.$$

Now, suppose that $S_{r\epsilon}\backslash S_{(r-1)\epsilon}$ is non-empty for every $r \leqslant \epsilon^{-1}/2$. (This will turn out to lead to a contradiction provided ϵ is sufficiently small.) Since the sets $(S_{r\epsilon}\backslash S_{(r-1)\epsilon})_r$ are disjoint and contained in $S_{1/2}$ we have, by Parseval's theorem, that

$$\begin{aligned}(\lfloor \epsilon^{-1}/2\rfloor -1).(7\alpha/8)^2 &\leqslant \sum_{r=1}^{\epsilon^{-1}/2}\sum_{\gamma\in S_{r\epsilon}\backslash S_{(r-1)\epsilon}} |\widehat{1_A}(\gamma)|^2\\ &= \sum_{\gamma\in\bigcup_{r\leqslant\epsilon^{-1}/2} S_{r\epsilon}\backslash S_{(r-1)\epsilon}} |\widehat{1_A}(\gamma)|^2 \leqslant \sum_{\gamma\in\widehat{G}}|\widehat{1_A}(\gamma)|^2 = \alpha.\end{aligned}$$

It follows that $\epsilon = \Omega(\alpha)$; thus if ϵ is sufficiently small then we have a contradiction and we conclude that $S_{r\epsilon}\backslash S_{(r-1)\epsilon}$ is empty for some $r \leqslant \epsilon^{-1}/2$. Thus by the triangle inequality we have that $S_{(r-1)\epsilon}+\lambda \subset S_{(r-1)\epsilon}$, and by induction $S_{(r-1)\epsilon}+n\lambda \subset S_{(r-1)\epsilon}$. On the other hand $0_{\widehat{G}} \in S_0 \subset S_{(r-1)\epsilon}$ and so $\lambda^n \in S_{(r-1)\epsilon} \subset S_{1/2}$ for all $n \in \mathbb{N}$.

Writing k for the order of λ we know that $\mathbb{E}_{1\leqslant n\leqslant k}\lambda^n(x) = 1_{\ker\lambda}(x)$ for all $x \in G$. Thus, it follows from the above that

$$\begin{aligned}(7/8)^2 \leqslant \mathbb{E}_{1\leqslant n\leqslant k}|\widehat{\mu_A}(\lambda^n)|^2 &= \int \mathbb{E}_{1\leqslant n\leqslant k}\lambda^n(a-a')d\mu_A(a)d\mu_A(a')\\ &= \int 1_{\ker\lambda}(a-a')d\mu_A(a)d\mu_A(a').\end{aligned}$$

We conclude that we are in the second case of the claim and we are done. □

As an immediate consequence of this claim, suppose that $G = \mathbb{Z}/p\mathbb{Z}$ for some prime p and $A \subset \mathbb{Z}/p\mathbb{Z}$ contains the identity and has size at least 2. Every non-trivial character has trivial kernel and so it follows that if we are in the second case of the claim then

$(49/64)|A|^2 \leqslant |A|$. Since $|A| \geqslant 2$ this is a contradiction we see that A has $\Omega(\alpha^2)$-spectral gap.

To summarise, combining our spectral gap result with our earlier result (3.9) we see that for every initial distribution μ_0 on $\mathbb{Z}/p\mathbb{Z}$ the random walk associated to A will have achieved 'good convergence' to the uniform distribution on $\mathbb{Z}/p\mathbb{Z}$ in $O(\alpha^{-2}\log|G|)$ steps.

Returning to the general setting of G a finite Abelian group and $A \subset G$ containing the identity we can use the previous claim to bootstrap our earlier estimate for the spectral gap.

Claim. *The set A has $\Omega(|A|/|G|^2)$-spectral gap.*

Proof. We apply the previous claim for each character $\lambda \notin A^{\perp}$. If we are in the first case then we have the relevant upper bound on $|\widehat{\mu_A}(\lambda)|$; assume we are in the second.

In the first instance we note that $|\ker\lambda| \geqslant 49|A|/64$ by averaging. By the isomorphism theorem we conclude that $|\operatorname{Im}\lambda| \leqslant 64|G|/49|A|$ and so by Lagrange's theorem $\lambda(x)$ has order at most $64|G|/49|A|$ for all $x \in G$.

Suppose that $a_0 \in A$ is such that $\lambda(a_0) \neq 1$. (Which exists since $\lambda \notin A^{\perp}$.) Write m for the order of $\lambda(a_0)$ which has $m > 1$ and $m \leqslant 64|G|/49|A|$. We conclude that

$$|1-\lambda(a_0)| \geqslant |1-\exp(2\pi i/m)| = \Omega(1/m^2) = \Omega(|A|^2/|G|^2).$$

Now let $c_\lambda \in S^1$ be such that $|\widehat{1_A}(\lambda)| = c_\lambda\widehat{1_A}(\lambda)$. Since $0_G \in A$ we have that

$$\begin{aligned}
|A| - |\widehat{1_A}(\lambda)| &= \operatorname{Re}\sum_{a\in A}(1-c_\lambda\lambda(a))\\
&= \sum_{a\in A}(1-\operatorname{Re}(c_\lambda\lambda(a)))\\
&\geqslant (1-\operatorname{Re}c_\lambda) + (1-\operatorname{Re}(c_\lambda\lambda(a_0))\\
&= \frac{1}{2}(|1-c_\lambda|^2 + |1-c_\lambda\lambda(a_0)|^2) \geqslant \frac{1}{4}|1-\lambda(a_0)|^2 = \Omega(|A|^2/|G|^2).
\end{aligned}$$

The claim follows on dividing. □

In the general setting this estimate is tight as can be see by letting A be a subgroup adjoined by one other element of order close to $|G|/|A|$. In general we do not use sets A which are 'close' to subgroups – rather we think of A as a generating set, in which case better estimates are often available.

4. Roth's theorem and arithmetic progressions

Roth's theorem on arithmetic progressions [Rot53] is one of the central problems of additive combinatorics and was one of the routes by which the power of the Fourier transform became apparent (in this field), although Roth used the term exponential sums.

A *three-term arithmetic progression* is a triple of integers (x, y, z) such that $x + z = 2y$ or, equivalently, a triple $(x, x + d, x + 2d)$. We say that such a progression is *trivial* if $d = 0$ so that all of x, y and z are the same.

Theorem 4.1 (Roth's theorem on arithmetic progressions)**.** *Suppose that $A \subset \{1, \ldots, N\}$ contains no non-trivial arithmetic progressions. Then*

$$|A| = O(N/\log\log N).$$

This result should be compared with Proposition 1.5 where instead of looking for triples (x, y, z) with $x + z = 2y$ we looked for quadruples (x, y, z, w) such that $x + z = y + w$. While that proposition was rather easy, Roth's theorem is not, and one of the reasons that Roth's theorem is so hard is that there are examples of rather large sets of integers not containing any non-trivial arithmetic progressions.

Theorem 4.2 (Behrend's construction, [Beh46])**.** *There is a set $A \subset \{1, \ldots, N\}$ containing no non-trivial arithmetic progressions such that*

$$|A| = \Omega\left(\frac{N}{\exp(2\sqrt{2\log_2 N})\log^{1/2} N}\right).$$

Proof. The basic idea is that the surfaces of strictly convex bodies do not contain (non-trivial) arithmetic progressions and that the higher the dimension of the underlying space, the more of the mass of the body is near the surface. Given such a surface in high dimensional space we embed it into $\{1, \ldots, N\}$.

Concretely we shall look at a sphere. We let M and d be naturals to be optimised later and put

$$S_r := \{x \in \{1, \ldots, M\}^d : x_1^2 + \cdots + x_d^2 = r\}.$$

By averaging it follows that there is some r such that $|S_r| \geqslant M^d/(dM^2)$. Of course, S_r is a set of points on a sphere which is strictly convex and so it contains no non-trivial convex combinations of points and, in particular, no three-term progressions.

We now consider the embedding

$$\begin{aligned} \phi : \{1, \ldots, M\}^d &\rightarrow \{1, \ldots, N\} \\ (x_1, \ldots, x_d) &\mapsto x_1 + (2M + 1)x_2 + \cdots + (2M + 1)^{d-1}x_d. \end{aligned}$$

This is into provided

$$M + (2M + 1)M + \cdots + (2M + 1)^{d-1}M \leqslant N;$$

and by design if

$$\phi(x) + \phi(y) = \phi(z) + \phi(w) \tag{4.1}$$

then $x + y = z + w$. It follows that $A := \phi(S_r)$ does not contain any non-trivial arithmetic progressions since S_r does not provided $(2M + 1)^d \leqslant N$. Furthermore, since ϕ is injective (as a result of (4.1)) we conclude that

$$|A| \geqslant M^d/(dM^2).$$

To ensure that A does not contain any arithmetic progressions we take $M = \lfloor (N^{1/d} - 1)/2 \rfloor$, so that

$$|A| \geqslant \frac{4N}{d2^d N^{2/d}}(1 - O(dN^{-1/d})).$$

We optimise this by taking d as close to the solution to '$2^z = N^{2/z}$' as possible. In particular we take d natural such that $d = \sqrt{2\log_2 N} + O(1)$, and the result follows. □

As we shall see later in Example 5.4, with slightly more care one can save a $\log^{1/4} N$ from the bottom, but essentially nothing better than this is known. Indeed, it took some sixty years before Elkin [Elk10] improved this by a further $\log^{1/2} N$, and this is the best known bound. For a nice and short version of Elkin's result the reader can consult the paper [GW10] of Green and Wolf.

We shall not prove Roth's theorem in these notes, but we shall prove a model version in a different group which captures most of the ideas. Indeed, nothing about the definition of a three-term arithmetic progression really requires that it be in the integers.

Suppose that G is a finite Abelian group. Then $(x, y, z) \in G^3$ is a *three-term arithmetic progression* if $x + z = 2y$. If the group has 2-torsion then this results in some degeneracy. Indeed, if $G = (\mathbb{Z}/2\mathbb{Z})^n$ then *any* triple (x, y, x) is a three-term arithmetic progression so we avoid this case. On the other hand if $G = (\mathbb{Z}/3\mathbb{Z})^n$ then $2 = -1 \pmod 3$ so a three-term arithmetic progression is just three points in a line.

Theorem 4.3 (Roth-Meshulam Theorem). *Suppose that $G := (\mathbb{Z}/3\mathbb{Z})^n$ and $A \subset G$ contains no three points in a line. Then*

$$|A| = O(3^n/n).$$

This result is due to Meshulam [Mes95], although the argument follows Roth [Rot53] adapted to the group $G = (\mathbb{Z}/3\mathbb{Z})^n$. To compare this with Roth's theorem we think of 3^n as being the equivalent of N so that the upper bound on A in Roth-Meshulam is of the form $O(N/\log N)$. This is rather better than that in Theorem 4.1, and while the bound in Theorem 4.1 can be improved (see, for example, the paper [Bou99] of Bourgain or its exposition in [TV06, §10.4]) nothing of the shape $O(N/\log N)$ is known.

Remark (Qualitative vs quantitative). We have discussed bounds quite a bit in this chapter and will continue to do so in the notes. Improvement of bounds provides a way to measure progress on a problem and the extent to which we understand what is going on. For example, if one could prove a bound in Theorem 4.1 matching the lower bound of Behrend, one might hope that it would say something about spheres in some sense 'being extremal' for this question. This also means that some improvements on bounds are more interesting than others, depending on the understanding they afford.

We shall turn now to proving the Roth-Meshulam theorem. Our starting point for both this and Roth's theorem is the same: we try to prove that if A has large enough density then A contains so many arithmetic progressions that it necessarily contains a non-trivial three-term arithmetic progression.

Suppose that G is a finite Abelian group and $A \subset G$ has density α. We write

$$T_G(A) \quad := \quad \mathbb{E}_{x,u\in G} 1_A(x)1_A(x+u)1_A(x-u)$$

so that $T_G(A)|G|^2$ is the number of three-term progression in the set A. We aim to show that $T_G(A)|G|^2 > \alpha|G|$ since $\alpha|G|$ is the number of trivial progressions in A.

As we have done before we can rewrite $T_G(A)$ in a simple way using convolution, and then diagonalise using the Fourier transform (specifically, insert the inversion formula for $1_A * 1_A$):

$$\begin{aligned} T_G(A) &= \mathbb{E}_{x\in G}1_A(x)\mathbb{E}_{u\in G}1_A(x+u)1_A(x-u)\\ &= \mathbb{E}_{x\in G}1_A(x)1_A * 1_A(2x)\\ &= \mathbb{E}_{x\in G}1_A(x)\sum_{\gamma\in\widehat{G}}\widehat{1_A}(\gamma)^2\gamma(2x) = \sum_{\gamma\in\widehat{G}}\widehat{1_A}(\gamma)^2\overline{\widehat{1_A}(2\gamma)}. \end{aligned}$$

We now do a little calculation assuming that G has no 2-torsion. If A were chosen at random with probability α (as in Examples 2.11 and 3.16) then we should expect $T_G(A)$ to be about α^3, since $T_G(A)|G|^2$ is the number of three-term progressions in A, and there are $|G|^2$ possible progressions each of which (except the trivial ones) is present with probability α^3. On the other hand $\widehat{1_A}(0_{\widehat{G}}) = \alpha$, so

$$T_G(A) - \alpha^3 = \sum_{\gamma\neq 0_{\widehat{G}}}\widehat{1_A}(\gamma)^3. \tag{4.2}$$

In Example 3.16 we saw that $\sup_{\gamma\neq 0_{\widehat{G}}}|\widehat{1_A}(\gamma)|$ was very small when A was chosen randomly, which fits with our expectations.[24] On the other hand, what (4.2) suggests is that this is an equivalence as we now prove.

Lemma 4.4. *Suppose G has no 2-torsion and $A \subset G$ has density α. Then at least one of the following is true.*

(i) *(Many progression) We have the estimate $T_G(A) \geqslant \alpha^3/2$;*

(ii) *(Large Fourier coefficient) There is some $\gamma \neq 0_{\widehat{G}}$ such that $|\widehat{1_A}(\gamma)| \geqslant \alpha^2/2$.*

Proof. We return to (4.2) and apply the triangle inequality to see that

$$|T_G(A) - \alpha^3| \leqslant \sum_{\gamma\neq 0_{\widehat{G}}}|\widehat{1_A}(\gamma)|^3 \leqslant \sup_{\gamma\neq 0_{\widehat{G}}}|\widehat{1_A}(2\gamma)|\sum_{\gamma\neq 0_{\widehat{G}}}|\widehat{1_A}(\gamma)|^2.$$

On the other hand by Parseval's theorem we have

$$\sum_{\gamma\neq 0_{\widehat{G}}}|\widehat{1_A}(\gamma)|^2 \leqslant \sum_{\gamma\in\widehat{G}}|\widehat{1_A}(\gamma)|^2 = \mathbb{E}_{x\in G}1_A(x)^2 = \alpha,$$

[24] The alert reader may wonder here: if we pick the elements of A independently at random with probability α then the expected number of three-term progressions is $\alpha^3(|G|^2-|G|)+\alpha|G|$. This is because there are $|G|^2-|G|$ non-trivial progressions each of which has probability α^3 of being included and then $|G|$ trivial progressions each of which has probability α of being included. It follows that $T_G(A) - \alpha^3 = \Theta_\alpha(1/|G|)$.

On the other hand from Example 3.16 we expect $|\widehat{1_A}(\gamma)|$ to be $\Theta_\alpha(|G|^{-1/2})$ in modulus so that we might expect the right hand side of (4.2) to be about $|G|.\Theta_\alpha(|G|^{-1/2})^3 = \Theta_\alpha(|G|^{-1/2})$ in size. The reader may be concerned about these heuristics since $\Theta_\alpha(1/|G|) \neq \Theta_\alpha(|G|^{-1/2})$.

Of course, what we have forgotten is that the sign of $\widehat{1_A}$ also behaves randomly so we expect square-root cancellation in the right hand side of (4.2) and then our adjusted heuristic for the right hand side tells us it is about $|G|^{1/2}.\Theta_\alpha(|G|^{-1/2})^3 = \Theta_\alpha(|G|^{-1})$ as desired.

and it follows that

$$|T_G(A) - \alpha^3| \leqslant \alpha \sup_{\gamma \neq 0_{\widehat{G}}} |\widehat{1_A}(2\gamma)|.$$

This gives the result since if $T_G(A) < \alpha^2/2$ then the left hand side is at least $\alpha^3/2$, and if G has no 2-torsion then 2γ is a non-trivial character if and only if γ is non-trivial. □

The reason the above lemma is useful is that a large Fourier coefficient leads to a density increment on a coset of a subgroup. Indeed, since γ is constant on cosets of $\{\gamma\}^\perp$ we see that if $\widehat{1_A}(\gamma) = \langle 1_A, \gamma\rangle_{L^2(G)}$ is large in modulus then A cannot have the same density on all cosets of $\{\gamma\}^\perp$; this yields a density increment.

Lemma 4.5. *Suppose that $A \subset G$ has density α and γ is non-trivial (meaning $\gamma \neq 0_{\widehat{G}}$) and has $|\widehat{1_A}(\gamma)| \geqslant \epsilon\alpha$. Then writing $V := \{\gamma\}^\perp$ we have*

$$\|1_A * \mu_V\|_{L^\infty(G)} \geqslant (1 + \epsilon/2)\alpha.$$

Proof. This is simply an averaging argument: first

$$\begin{aligned} \|(1_A - \alpha) * \mu_V\|_{L^1(G)} &\geqslant \|(1_A - \alpha) * \mu_V\|_{\ell^\infty(\widehat{G})} \\ &\geqslant |(1_A - \alpha)^\wedge(\gamma)||\widehat{\mu_V}(\gamma)| = |\widehat{1_A}(\gamma)| \geqslant \epsilon\alpha. \end{aligned}$$

Note, crucially,

$$(1_A - \alpha)^\wedge(\gamma) := (1_A - \alpha 1_G)^\wedge(\gamma) = \widehat{1_A}(\gamma) - \alpha\widehat{1_G}(\gamma) = \widehat{1_A}(\gamma) - \alpha 1_{\{0_{\widehat{G}}\}}(\gamma) = \widehat{1_A}(\gamma)$$

since γ is non-trivial. On the other hand

$$\mathbb{E}_{x \in G}(1_A - \alpha) * \mu_V(x) = \alpha - \alpha = 0.$$

It follows that

$$2 \max_x (1_A - \alpha) * \mu_V(x) \geqslant \mathbb{E}_{x\in G}|(1_A - \alpha) * \mu_V(x)| + \mathbb{E}_{x\in G}(1_A - \alpha) * \mu_V(x) \geqslant \epsilon\alpha,$$

and we get the result. □

Our plan of action is to combine these two lemmas and iterate:

(i) either we have many three-term progressions;
(ii) or there is a coset of a subgroup on which we have increased density.

If we are in the second case we put the output back into the iteration (possible since three-term progressions are translation invariant, so a three-term progression in the coset $x + V$ is the same as one in V) and repeat. This process cannot go on forever since density is bounded above by 1 and so we terminate in the first case.

The main problem with this plan is that the subgroup we pass to may be very small and this is why the Roth-Meshulam theorem is easier to prove than Roth's theorem: if $G = (\mathbb{Z}/3\mathbb{Z})^n$ then every element has order 3 and so $\gamma(x)^3 = \gamma(3.x) = 1$ and so γ is a homomorphism with an image of size at most 3. It follows that $|\ker\gamma| \geqslant |G|/3$ which is large.

Proof of Theorem 4.3. We write α for the density of A in G and we proceed iteratively to define a sequence of spaces $G =: V_0 \geqslant V_1 \geqslant \ldots \geqslant V_k \geqslant \ldots$ and elements $0_G =: x_0, x_1, \ldots, x_k, \ldots$. We write

$$A_i := V_i \cap (x_i + A), \alpha_i := \mathbb{P}_{V_i}(A_i) \text{ and } K_i := |G/V_i|.$$

Suppose that we are at stage i of the iteration. Apply Lemma 4.4 to the set A_i considered as a subset of V_i. It follows that either

$$T_{V_i}(A_i) \geqslant \alpha_i^3/2 \text{ or } \sup_{\gamma \neq 0_{\widehat{V_i}}} |\widehat{1_{A_i}}(\gamma)| \geqslant \alpha_i^2/2. \tag{4.3}$$

In the first case this means that A_i contains at least $(\alpha_i^3/2).|V_i|^2$ arithmetic progressions. Since arithmetic progressions are translation invariant it follows that A contains at least $(\alpha_i^3/2)|V_i|^2 = (\alpha_i^3/2)K_i^{-2}|G|^2$ arithmetic progressions, and it will turn out that we shall be done.

On the other hand in the second case, by Lemma 4.5 there is a subgroup $V_{i+1} \leqslant V_i$ with $|V_i/V_{i+1}| = 3$ (since every (non-trivial) element of G is order 3 means that every (non-trivial) element of V_i is order 3, and hence so is every (non-trivial) element of $\widehat{V_i}$) and some $y_i \in V_i$ such that

$$1_{A_i} * \mu_{V_{i+1}}(y_i) = \|1_{A_i} * \mu_{V_{i+1}}\|_{L^\infty(V_i)} \geqslant (1 + \alpha_i/4)\alpha_i.$$

Putting $x_{i+1} = y_i + x_i$ we see that

$$\begin{aligned} \alpha_{i+1} = \mathbb{P}_{V_{i+1}}(A_{i+1}) &= \mathbb{P}_{V_{i+1}}(V_{i+1} \cap (x_{i+1} + A)) \\ &= \mathbb{P}_{V_{i+1}}(V_{i+1} \cap (y_i + x_i + A)) \\ &= \mathbb{P}_{V_{i+1}}(V_{i+1} \cap (y_i + A_i)) = 1_A * \mu_{V_{i+1}}(y_i) \geqslant (1 + \alpha_i/4)\alpha_i; \end{aligned}$$

additionally $K_{i+1} = |G/V_{i+1}| = 3|G/V_i| = 3K_i$.

Since $\alpha_{i+1} \geqslant (1 + \alpha_i/4)\alpha_i$ we see that after $k_i = O(\alpha_i^{-1})$ steps we have $\alpha_{i+k_i} \geqslant 2\alpha_i$. On the other hand $\alpha_0 = \alpha$ and $\alpha_i \leqslant 1$ and so the iteration terminates in the first case of (4.3) in

$$O(\alpha^{-1}) + O((2\alpha)^{-1}) + O((4\alpha)^{-1}) + \cdots + O((2^r\alpha)^{-1}) + \cdots = O(\alpha^{-1})$$

steps. When it terminates we conclude that $K_i \geqslant \exp(-O(\alpha^{-1}))$ (and $\alpha_i \geqslant \alpha$) whence A contains at least $\exp(-O(\alpha^{-1}))|G|^2$ three-term progressions.

Finally, since A contains only trivial progressions we have that

$$\alpha|G| \geqslant T_G(A)|G|^2 \geqslant \exp(-O(\alpha^{-1}))|G|^2$$

from which the bound follows. □

Interestingly Behrend's construction does *not* extend to the model setting of $G := (\mathbb{Z}/3\mathbb{Z})^n$. Indeed, the best know construction is due to Edel [Ede04] who showed that there is a set $A \subset (\mathbb{Z}/3\mathbb{Z})^n$ not containing any non-trivial three-term progressions such that $|A| \geqslant (2.2\ldots)^n$.

Despite the fact that for Roth's theorem one of the main constructions does not translate to the model setting, the general technique of translating problems for the integers into model settings such as $(\mathbb{Z}/3\mathbb{Z})^n$ is very useful. For more information it is recommended that the reader take a look at the wonderful survey paper [Gre05] of Green.

One can adapt Lemma 4.5 to work for 'approximate annihilators of characters'; such sets are called Bohr sets.

Definition 4.6 (Bohr sets)**.** Suppose that Γ is a set of characters on G and $\delta \in (0,2]$. Then the *Bohr set* with *frequency set* Γ and *width* δ is the set

$$\mathrm{Bohr}(\Gamma,\delta) := \{x \in G : |\gamma(x) - 1| \leqslant \delta \text{ for all } \gamma \in \Gamma\},$$

and it is said to have *rank* $|\Gamma|$. The rank is sometimes called the dimension, but this term is also used for another quantity so we avoid it here. Additionally, the term Bohr neighbourhood is also used for Bohr sets although we shall use this for any translate of a Bohr set.

It is immediate from the definition that

$$\Gamma^{\perp} \subset \mathrm{Bohr}(\Gamma,\delta) \subset G.$$

Since $|\gamma(x)| = 1$, the triangle inequality tells us that we have equality on the right when $\delta = 2$; on the other hand if $\delta < |1 - \exp(2\pi i/r)|$ where r is the maximum order of an element in Γ then we have equality on the left.

In light of this we think of the width parameter as measuring the degree to which the Bohr set $\mathrm{Bohr}(\Gamma,\delta)$ approximates the annihilator $\Gamma^{\perp}$.

Even with Bohr sets proving Roth's theorem is not as easy as simply plugging in the generalisation of Lemma 4.5 into the proof of Theorem 4.3. The problem is that instead of getting a density increment on a coset of a subgroup we get a density increment on a Bohr neighbourhood and so we then have to go back and adapt Lemmas 4.4 and 4.5 to subsets of Bohr sets, rather than just subsets of groups. This can be done provided the Bohr sets satisfy a certain technical condition called regularity. This condition does not always hold, but is ubiquitous and so holds enough of the time to make the argument work although the details are fierce.

Example 4.7 (Bohr sets in cyclic groups)**.** Bohr sets are at their most interesting in cyclic groups. Suppose that $G = \mathbb{Z}/N\mathbb{Z}$ and $\Gamma = \{\gamma\}$ where $\gamma(x) = \exp(2\pi i x/N)$. Then we have that

$$|1 - \gamma(x)| = \sqrt{2 - 2\cos(2\pi x/N)} = 2\pi|x|/N + O_{|x|/N\rightarrow 0}(|x|/N)^3,$$

and so

$$\begin{aligned}\mathrm{Bohr}(\Gamma,\delta) &= \{x \in G : |1 - \exp(2\pi i x/N)| \leqslant \delta\}\\ &= \{x + N\mathbb{Z} : |x| \leqslant (\delta + O_{\delta\rightarrow 0}(\delta^3))N/2\pi\}.\end{aligned}$$

We see that the Bohr set is a symmetric interval of width about $\delta N/\pi$. Note, in particular, that if δ is small enough then the Bohr set may only contain the identity despite being rank 1.

Taking other characters in the previous example leads to dilates of intervals – centred arithmetic progressions – and taking a Bohr set with multiple characters in the frequency set therefore leads to intersections of centred arithmetic progressions.

To understand Bohr sets better it will be useful to have an estimate for how big they are, and in light of the example in cyclic groups it should come as little surprise that this translates to a bound on the density (as opposed to size).

Lemma 4.8 (Bohr set size estimate)**.** *Suppose that* $\mathrm{Bohr}(\Gamma, \delta)$ *is a rank* k *Bohr set. Then*

$$\mathbb{P}_G(\mathrm{Bohr}(\Gamma, \delta)) \geqslant (\delta(1 - o_{\delta\to 0}(1))/2\pi)^k.$$

Proof. We write $\mathbb{T} := \mathbb{R}/\mathbb{Z}$ and consider the homomorphism

$$\phi : G \to \mathbb{T}^\Gamma; x \mapsto \left(\frac{\log \gamma(x)}{2\pi i}\right)_{\gamma\in\Gamma}.$$

Writing Q for the cube $[0, \eta]^k + \mathbb{Z}^\Gamma$ in $\mathbb{T}^\Gamma$, we have, interchanging the order of summation, that

$$\begin{aligned}\mathbb{E}_{z\in\mathbb{T}^\Gamma}\mathbb{P}_G(\{x \in G : \phi(x) \in z + Q\}) &= \mathbb{E}_{z\in\mathbb{T}^\Gamma}\mathbb{E}_{x\in G}1_{z+Q}(\phi(x))\\ &= \mathbb{E}_{x\in G}\mathbb{E}_{z\in\mathbb{T}^\Gamma}1_{z+Q}(\phi(x))\\ &= \mathbb{E}_{x\in G}\mathbb{E}_{z\in\mathbb{T}^\Gamma}1_{\phi(x)-Q}(z)\\ &= \mathbb{E}_{x\in G}\mu(\phi(x) - Q) = \mu(Q).\end{aligned}$$

Thus, by averaging that there is some $z \in \mathbb{T}^\Gamma$ such that

$$\mathbb{P}_G(\{x \in G : \phi(x) \in z + Q\}) \geqslant \mu(Q).$$

On the other hand if $x, y \in \{x \in G : \phi(x) \in z + Q\}$ then (since ϕ is a homomorphism) we have that

$$|1 - \gamma(x - y)| \leqslant |1 - \exp(2\pi i.\eta)|.$$

Choosing η such that $|1 - \exp(2\pi i.\eta)| = \delta$ yields the result since $\mu(Q) \geqslant \eta^{|\Gamma|} = \eta^k$. □

Up to the constant this is best possible as can be seen by considering, for example, a cube in $G = (\mathbb{Z}/M\mathbb{Z})^k$. The cube example also motivates our thinking for arithmetic progressions. Indeed, it turns out that every Bohr set contains a long arithmetic progression as we can show using the above.

Lemma 4.9 (Arithmetic progressions in Bohr sets)**.** *Suppose that* G *is cyclic and* Γ *is a set of characters on* G *of size* k*. Then the Bohr set* $\mathrm{Bohr}(\Gamma, \delta)$ *contains an arithmetic progression of length at least* $\delta|G|^{1/(k+1)}(1 - o_{\delta\to 0}(1))/\pi$ *centred at* 0_G*.*

Proof. Write $K := \bigcap_{\gamma\in\Gamma} \ker \gamma$ which we note is cyclic since it is a subgroup of a cyclic group and we write $|K| = M$. First note that since K is cyclic $\Gamma^\perp \subset \mathrm{Bohr}(\Gamma, \delta)$ contains an arithmetic progression of length M. Secondly, from the definition of K the characters in Γ induce a set of characters Γ' on $H := G/K$.

In light of Lemma 4.8 we can pick $\rho \sim 2\pi|H|^{-1/k}$ such that $\mathbb{P}_H(\mathrm{Bohr}(\Gamma',\rho)) > 1/|H|$. It follows that there is some $x_0 \in \mathrm{Bohr}(\Gamma',\rho)$ with $x_0 \neq 0_H$. But then if l is an integer with $|l| \leqslant \delta\rho^{-1}$ we have

$$|1-\gamma(lx_0)| \leqslant l|1-\gamma(x_0)| \leqslant \delta \text{ for all } \gamma \in \Gamma'.$$

It follows that $\{lx_0 : |l| \leqslant \delta\rho^{-1}\} \subset \mathrm{Bohr}(\Gamma',\delta)$; the problem is that these elements might not be distinct. However, since $x_0 \neq K = 0_H$ we conclude that there is some $\gamma \in \Gamma'$ such that $\gamma(x_0) \neq 1$. It follows that all the values $\gamma(lx_0)$ are distinct for $|l| \leqslant \delta\rho^{-1}$ and hence the arithmetic progression genuinely has length $2\lfloor\delta\rho^{-1}\rfloor + 1$.

It remains to optimise by minimising

$$\max\{M, 2\lfloor\delta\rho^{-1}\rfloor + 1\} \sim \max\{M, \delta(|G|/M)^{1/k}/\pi\}$$

which gives the result. □

The above result is easier and stronger if G is cyclic of prime order; we leave the proof of this to the exercises.

We now turn to the last result of the section which is closely related to Roth's theorem. The starting point is that if we have a set $A \subset \{1,\dots,N\}$ (with more than 1 element) then eventually if one keeps adding A to itself it will contain a translate of $\{1,\dots,N\}$. (The reader may wish to compare this with the work in Example 3.17.)

The following theorem, due to Freĭman, Halberstam and Ruzsa [FHR92], shows that even after adding the set A to itself three times we get a long arithmetic progression.

Theorem 4.10. *Suppose that $A \subset \{1,\dots,N\}$ has size αN. Then $A+A+A$ contains an arithmetic progression of length $\Omega(\alpha N^{\Omega(\alpha^3)})$.*

Proof. First we embed A in $G := \mathbb{Z}/6N\mathbb{Z}$ via the usual quotient map from Z and write A' for the image. This has density $\alpha' = \alpha/6$ and if $A'+A'+A'$ contains an arithmetic progression of length l then this lifts to an arithmetic progression of length l in $A+A+A$.

We now study $A'+A'+A'$ through the three-fold convolution of $1_{A'}$ with itself. Since we have a bound on the density of A' in G we think of the group as endowed with Haar probability measure. By Fourier inversion we have that

$$1_{A'} * 1_{A'} * 1_{A'}(x) = \sum_{\gamma\in\widehat{G}} (1_{A'} * 1_{A'} * 1_{A'})^\wedge(\gamma)\gamma(x) = \sum_{\gamma\in\widehat{G}} \widehat{1_{A'}}(\gamma)^3\gamma(x).$$

We shall split this sum into three parts which we shall deal with separately. First, recall that $\widehat{1_{A'}}(0_{\widehat{G}}) = \alpha'$. This term is going to be the main term and we shall want to show that the rest of the sum above is an error term. We write

$$\Gamma := \{\gamma \neq 0_{\widehat{G}} : |\widehat{1_{A'}}(\gamma)| \geqslant \epsilon\alpha'\} \text{ and } \Gamma' := \{\gamma : |\widehat{1_{A'}}(\gamma)| < \epsilon\alpha'\},$$

so that

$$1_{A'} * 1_{A'} * 1_{A'}(x) = \alpha'^3 + C(x) + E(x)$$

where

$$C(x) = \sum_{\gamma\in\Gamma} \widehat{1_{A'}}(\gamma)^3\gamma(x) \text{ and } E(x) = \sum_{\gamma\in\Gamma'} \widehat{1_{A'}}(\gamma)^3\gamma(x).$$

The term $E(x)$ will genuinely be an error term in L^∞; the term $C(x)$ will be roughly constant on the Bohr set with frequency set Γ. To see these facts we have

$$\left|\sum_{\gamma\in\Gamma'}\widehat{1_{A'}}(\gamma)^3\gamma(x)\right| \leqslant \sup_{\gamma\in\Gamma'}|\widehat{1_{A'}}(\gamma)|\sum_{\gamma\in\Gamma'}|\widehat{1_{A'}}(\gamma)^2| \leqslant \epsilon\alpha'\sum_{\gamma\in\widehat{G}}|\widehat{1_{A'}}(\gamma)|^2 = \epsilon\alpha'^2,$$

so that $|E(x)| \leqslant \epsilon\alpha'^2$. Secondly, if $y \in \mathrm{Bohr}(\Gamma,\delta)$ then we have

$$\begin{aligned}|C(x+y)-C(x)| &\leqslant \sum_{\gamma\in\Gamma}|\widehat{1_{A'}}(\gamma)|^3|\gamma(x+y)-\gamma(x)|\\ &= \sum_{\gamma\in\Gamma}|\widehat{1_{A'}}(\gamma)|^3|\gamma(y)-1| \leqslant \delta\sum_{\gamma\in\widehat{G}}|\widehat{1_{A'}}(\gamma)|^3 \leqslant \delta\alpha'^2.\end{aligned}$$

We are now in a position to find a Bohr neighbourhood on which $1_{A'} * 1_{A'} * 1_{A'}$ is large. If $\gamma \in \Gamma$ then $\mathbb{E}_x\gamma(x) = 0$ (since $\gamma \neq 0_{\widehat{G}}$) so that $\mathbb{E}_xC(x) = 0$ and there is some x_0 such that $C(x_0) \geqslant 0$. But then for every $x \in x_0 + \mathrm{Bohr}(\Gamma,\delta)$ we have

$$1_{A'} * 1_{A'} * 1_{A'}(x) \geqslant \alpha'^3 + 0 - \delta\alpha'^2 - \epsilon\alpha'^2$$

by the triangle inequality. If we take $\epsilon = \alpha'/3$ and $\delta = \alpha'/3$ then we see that $x \in A'+A'+A'$. Finally we use Parseval's theorem to bound the size of Γ:

$$|\Gamma|(\epsilon\alpha')^2 \leqslant \sum_{\gamma\in\Gamma}|\widehat{1_{A'}}(\gamma)|^2 \leqslant \alpha',$$

so $|\Gamma| \leqslant \alpha'^{-3}$. We conclude that $A'+A'+A'$ contains a translate of a Bohr set of width $\alpha'/3$ and rank $O(\alpha'^{-3})$. By Lemma 4.9 this means that $A'+A'+A'$ contains an arithmetic progression of length $\Omega(\alpha' N^{\Omega(\alpha'^3)})$ and we are done in light of the relationship between α and α'. □

It should be remarked that $A+A$ does not necessarily contain an arithmetic progression of length $N^{\Omega_\alpha(1)}$, so up to α-dependence the above result is best possible. The examples showing this are due to Ruzsa [Ruz91] and are technically formidable.

Remark 4.11. It may also be worth remarking that one often chooses the embeddings in proofs of the type given above to be into groups of prime order. This often has benefits, although typically not of an essential nature. For example Lemma 4.9 is stronger and easier to prove in this case, but not essentially more useful.

To get an appropriate prime we could use the prime number theorem, but a much weaker result works. Indeed, we use (a weak version of) Bertrand's postulate[25] that there is always a prime between n and $O(n)$. This is easy enough to prove: first we need an upper bound on the Chebychev function $\theta(n)$. On the one hand every prime between r and $2r$ divides

[25] Bertrand's postulate asserts that there is always a prime between n and $2n$ and in fact they are far more common.

$\binom{2r}{r}$ a least once and so

$$\prod_{p\leqslant n} p \leqslant \prod_{i=0}^{\lfloor \log_2 n\rfloor} \prod_{2^i<p\leqslant 2^{i+1}} p \leqslant \prod_{i=0}^{\lfloor \log_2 n\rfloor} \binom{2^{i+1}}{2^i} \leqslant 4^{\sum_{i=0}^{\lfloor \log_2 n\rfloor} 2^i} = \exp(O(n)). \tag{4.4}$$

On the other hand the number of times a prime p divides $n!$ is $\sum_{i\geqslant 1}\lfloor n/p^i\rfloor$, and

$$n/p - O(1) \leqslant \sum_{i\geqslant 1}\lfloor n/p^i\rfloor \leqslant n/p + O(n/p^2).$$

Hence

$$\begin{aligned} n\log n + O(n) = \log n! &= \sum_{p\leqslant n}\log p \sum_{i\geqslant 1}\left\lfloor \frac{n}{p^i}\right\rfloor \\ &= n\sum_{p\leqslant n}\frac{\log p}{p} + \sum_{p\leqslant n} O(1)\log p + \sum_{p\leqslant n} O(n/p^2)\log p. \end{aligned}$$

Inserting the log of the bound (4.4) and the fact that $\sum_{p\leqslant n} p^{-2}\log p = O(\sum_{x\leqslant n} x^{-3/2}) = O(1)$ into this and dividing by n gives

$$\sum_{p\leqslant n}\frac{\log p}{p} = \log n + O(1).$$

This result is called Merten's theorem and the weak version of Bertand's postulate follows immediately since the sum of $(\log p)/p$ for primes between n and Dn is then at least $\log D - O(1)$ which is positive provided D is a sufficiently large absolute constant. That is to say there is a prime between n and $Dn = O(n)$.

5. Sums of independent random variables

We have spent considerable time developing the Fourier transform and examining basic examples of its uses. At this point we shall introduce a new tool called measure concentration and some variants which are very powerful. In particular they will help us revise and improve much of our earlier work with some much stronger estimates.

Our starting point is sums of independent random variables which, in the spirit of the rest of the notes, will be assumed based on a finite probability space. Suppose that $X_1, \dots, X_n$ are independent with mean 0 and variance 1. We are interested in $\sum_i X_i$ which has

$$\mathbb{E}\sum_i X_i = 0 \text{ and } \operatorname{Var}(\sum_i X_i) = n$$

where the variance calculation is as a result of the X_is being (pair-wise) independent. By Chebychev's inequality we see that most of the mass of $\sum_i X_i$ is concentrated in $[-C\sqrt{n}, C\sqrt{n}]$; indeed it gives

$$\mathbb{P}(|\sum_i X_i|/\sqrt{n} \notin [-C, C]) \leqslant 1/C^2.$$

There are now two questions about $\sum_i X_i/\sqrt{n}$:

(i) What does $\sum_i X_i/\sqrt{n}$ look like inside the interval $[-C,C]$? Answers to this are called local limit theorems.

(ii) How rapidly does $\sum_i X_i/\sqrt{n}$ concentrate on the interval $[-C,C]$? Answers to this go by various names such as concentration of measure.

Our interest is more in the second of these two questions.

At this point it is instructive to consider the prototypical example. Suppose that the X_is above are identically distributed with $\mathbb{P}(X_i=1)=\mathbb{P}(X_i=-1)=1/2$. Then for $r\equiv n \pmod 2$ we have

$$\mathbb{P}(\sum_i X_i = r) = \frac{1}{2^n}\binom{n}{(n+r)/2}.$$

Provided $n-|r|\rightarrow\infty$ as $n\rightarrow\infty$ we can use Stirling's formula to get that

$$\begin{aligned}\binom{n}{(n+r)/2} &\sim \frac{n^{n+1/2}e^{-n}\sqrt{2\pi}}{((n+r)/2)^{(n+r+1)/2}e^{-(n+r)/2}\sqrt{2\pi}((n-r)/2)^{(n-r+1)/2}e^{-(n-r)/2}\sqrt{2\pi}}\\ &= \frac{2^{n+1}}{\sqrt{2\pi n}}\left(\frac{n^2}{n^2-r^2}\right)^{(n+1)/2}\left(\frac{n-r}{n+r}\right)^{r/2},\end{aligned}$$

so that

$$\mathbb{P}(\sum_i X_i = r) \sim \frac{2}{\sqrt{2\pi n}}\left(\frac{n^2}{n^2-r^2}\right)^{(n+1)/2}\left(\frac{n-r}{n+r}\right)^{r/2}. \tag{5.1}$$

Initially we are looking for an upper bound on this in terms of r, and we begin by considering the range $0\leqslant r\leqslant n/4$. When this is the case we have

$$\begin{aligned}\mathbb{P}(\sum_i X_i = r) &= O\left(\left(\frac{1}{1-(r^2/n^2)}\right)^{n/2}\left(1-\frac{2r}{n+r}\right)^{r/2}\right)\\ &= O(\exp(3r^2/4n)\exp(-r^2/(n+r))\\ &= O(\exp(-r^2/20n))\end{aligned}$$

since $(1-x)^{-1}\leqslant\exp(3x/2)$ whenever $x\in[0,1/3]$ and $1-x\leqslant\exp(-x)$ for all $x\in[0,\infty)$. Of course, binomial coefficients are decreasing away from the centre and are symmetric so we conclude that

$$\mathbb{P}(\sum_i X_i = r) = O(\exp(-r^2/320n)) = O(\exp(-\Omega(r^2/n))),$$

and hence summing gives an upper bound of

$$\mathbb{P}(|\sum_i X_i| > r) = O(\exp(-\Omega(r^2/n))).$$

This bound is called a Chernoff-type bound and more information can be found in the classic text [Shi96] of Shiryaev. For comparison it is exponentially stronger than Chebychev's estimate, giving

$$\mathbb{P}(|\sum_i X_i|/\sqrt{n}\notin[-C,C]) = O(\exp(-\Omega(C^2))). \tag{5.2}$$

Of course Chebychev's bound holds for any pairwise independent set of random variables, whereas the estimate just proved is for the specific case of the distributions defined above. It turns out, however, that results of this shape are true in much greater generality and the key properties are three-fold: the X_is have mean 0; bounded L^∞; and are independent.

To prove this we start with a clever lemma. It is unlikely that the discovery of this lemma was made with this proof, but it is a natural refinement of *a* proof.

Lemma 5.1. *Suppose that $p \in [2,\infty)$ and $X_1,\dots,X_n$ are independent random variables with mean 0. Then*

$$\mathbb{E}|\exp(\sum_i X_i)| \leqslant \exp(\frac{1}{2}\sum_i \|X_i\|_{L^\infty(\mathbb{P})}^2).$$

Proof. Begin by noting that

$$\mathbb{E}|\exp(\sum_i X_i)| = \mathbb{E}\exp(\mathrm{Re}\sum_i X_i) = \mathbb{E}\prod_i \exp(\mathrm{Re}\, X_i).$$

Now we have the elementary inequality $\exp(ty) \leqslant \cosh t + y\sinh t$ whenever $t \in \mathbb{R}$ and $-1 \leqslant y \leqslant 1$, so

$$\mathbb{E}|\exp(\sum_i X_i)| \leqslant \mathbb{E}\prod_i \left(\cosh \|X_i\|_{L^\infty(\mathbb{P})} + \frac{\mathrm{Re}(X_i)}{\|X_i\|_{L^\infty(\mathbb{P})}} \sinh \|X_i\|_{L^\infty(\mathbb{P})}\right),$$

with the usual convention that $t^{-1}\sinh t$ is 1 if $t = 0$. Since $\mathbb{E}X_i = 0$ we conclude that $\mathbb{E}\mathrm{Re}\, X_i = 0$ and so, by independence we pass $\mathbb{E}$ through the product, and get

$$\begin{aligned}\mathbb{E}|\exp(\sum_i X_i)| &\leqslant \prod_i \left(\cosh \|X_i\|_{L^\infty(\mathbb{P})} + \frac{\mathbb{E}\,\mathrm{Re}(X_i)}{\|X_i\|_{L^\infty(\mathbb{P})}} \sinh \|X_i\|_{L^\infty(\mathbb{P})}\right)\\ &= \prod_i \cosh \|X_i\|_{L^\infty(\mathbb{P})}.\end{aligned}$$

The result follows since $\cosh x \leqslant \exp(x^2/2)$. □

A rather general Chernoff-type bound follows immediately from this by a quadratic optimisation.

Proposition 5.2. *Suppose that $p \in [2,\infty)$ and $X_1,\dots,X_n$ are independent random variables with mean 0. Then if the X_is are real we have*

$$\mathbb{P}\left(|\sum_i X_i| \geqslant C\sqrt{\sum_i \|X_i\|_{L^\infty(\mathbb{P})}^2}\right) \leqslant 2\exp(-C^2/2) \textit{ for all } C \geqslant 0,$$

and if the X_is are complex we have

$$\mathbb{P}\left(|\sum_i X_i| \geqslant C\sqrt{\sum_i \|X_i\|_{L^\infty(\mathbb{P})}^2}\right) \leqslant 4\exp(-C^2/4) \textit{ for all } C \geqslant 0,$$

Proof. Write $S := \sqrt{\sum_i \|X_i\|_{L^\infty(\mathbb{P})}^2}$ and let λ be a non-negative real to be optimised later. Apply the previous lemma to the variables $\lambda X_1, \dots, \lambda X_n$ to get that

$$\mathbb{E}|\exp(\lambda \sum_i X_i)| \leqslant \exp(\lambda^2 S^2/2).$$

It now follows that

$$\mathbb{P}(\operatorname{Re}\sum_i X_i \geqslant CS)\exp(\lambda CS) = \mathbb{P}(\lambda \operatorname{Re}\sum_i X_i \geqslant \lambda CS)\exp(\lambda CS) \leqslant \exp(\lambda^2 S^2/2).$$

Thus solving the quadratic we put $\lambda = C/S$ and get that

$$\mathbb{P}(\operatorname{Re}\sum_i X_i \geqslant CS) \leqslant \exp(-C^2/2).$$

Similarly applying the above argument to the variables $i^r X_1, \dots, i^r X_n$ for $r \in \{1,2,3\}$ tells us that

$$\mathbb{P}(\pm \operatorname{Re}\sum_i X_i \geqslant CS), \mathbb{P}(\pm \operatorname{Im}\sum_i X_i \geqslant CS) \leqslant \exp(-C^2/2),$$

and the result follows by the triangle inequality. □

Note that the above argument significantly improves the constant in the exponent we obtained in the bound (5.2). On the other hand, by returning to (5.1) we recall that for X_i independent with $\mathbb{P}(X_i = 1) = \mathbb{P}(X_i = -1) = 1/2$ and $r \sim C\sqrt{n}$ we get

$$\mathbb{P}(\sum_i X_i = r) \sim \frac{2}{\sqrt{2\pi n}}\left(\frac{n^2}{n^2 - r^2}\right)^{(n+1)/2}\left(\frac{n-r}{n+r}\right)^{r/2} \sim \frac{2}{\sqrt{2\pi n}}\exp(-C^2/2)$$

provided $C = o(\sqrt{n})$. By summing over the range of $r \in [C\sqrt{n}, C\sqrt{n} + \sqrt{n}/C]$ we conclude that

$$\mathbb{P}(|\sum_i X_i| > r) = \Omega(C^{-1}\exp(-C^2/2))$$

so that the bound above is close to tight.[26]

Our first application of this bound is in improving the estimates in Example 3.16.

Example 5.3 (The Fourier transform of random sets, Example 3.16 contd.)**.** As in Examples 2.11 and 3.16 suppose again that each $x \in G$ is placed in the set A independently with probability α. Our aim is to show the previously promised statement that

$$\sup_{\gamma \neq 0_{\widehat{G}}} |\widehat{1_A}(\gamma)| = O\left(\frac{\log|G|}{|G|}\right)^{1/2} \text{ w.h.p.} \tag{5.3}$$

[26] The tightness we have in mind here is in the constant of C^2 in the exponent, and must disappear as C approaches $\sqrt{n}$ since $\mathbb{P}(|\sum_i X_i| > n-1) = 2^{n-1}$ whereas $\exp(-C^2/2) = \exp(-n/2)$ in that case and $\exp(-1/2) > 1/2$.

We start by fixing $\gamma \neq 0_{\widehat{G}}$ and for each $x \in G$ let $X_x := (1_A(x) - \alpha)\overline{\gamma(x)}$. The random variables $(X_x)_{x\in G}$ are independent, have mean 0, and $\|X_x\|_{L^\infty(\mathbb{P})} = \max\{1-\alpha, \alpha\}$. It follows from Proposition 5.2 that

$$\mathbb{P}(|\sum_{x\in G} X_x| \geqslant C\sqrt{|G|\max\{1-\alpha,\alpha\}}) \leqslant 2\exp(-C^2/2).$$

But since $\gamma \neq 0_{\widehat{G}}$ we have

$$|\widehat{1_A}(\gamma)| = \frac{1}{|G|}|\sum_{x\in G}(1_A(x) - \alpha)\overline{\gamma(x)}| = \frac{1}{|G|}|\sum_{x\in G} X_x|,$$

and moreover $\max\{1-\alpha,\alpha\} \leqslant 1$, so

$$\mathbb{P}(|\widehat{1_A}(\gamma)| \geqslant C/\sqrt{|G|}) \leqslant 2\exp(-C^2/2).$$

But then by the triangle inequality we get

$$\mathbb{P}(\sup_{\gamma\neq 0_{\widehat{G}}} |\widehat{1_A}(\gamma)| \geqslant C/\sqrt{|G|}) \leqslant 2(|G|-1)\exp(-C^2/2),$$

and we can take $C = O(\sqrt{\log|G|})$ so that (5.3) holds. It may be worth saying that even this estimate is not best possible, and there are sets A with $\sup_{\gamma\neq 0_{\widehat{G}}} |\widehat{1_A}(\gamma)| = O(|G|^{-1/2})$. These sets are *not* typical in the sense that the upper bound above is tight up to a constant for randomly chosen sets, but they can be constructed by more sophisticated random arguments of Beck and Spencer (see [AS08, §12.2]) or explicitly Rudin-Shapiro polynomials.

Example 5.4 (Improving the Behrend construction)**.** As mentioned at the end of the Behrend construction (Theorem 4.2) with slightly more care one can construct a subset $A \subset \{1,\dots,N\}$ containing no three-term arithmetic progressions with

$$|A| = \Omega\left(\frac{N}{\exp(2\sqrt{2\log_2 N})\log^{1/4} N}\right).$$

With our Chernoff-type estimates we are in a position to make this improvement. Recall that we were considering the sets

$$S_r := \{x \in \{1,\dots,M\}^d : x_1^2 + \cdots + x_d^2 = r\},$$

and noted by that there is some r such that $|S_r| \geqslant M^d/(dM^2)$. We shall now show that there is some r with that $|S_r| = \Omega(M^d/\sqrt{d}M^2)$ from which the claimed improvement follows. To see this we consider the independent random variables $(X_i)_{i=1}^d$ defined by

$$\mathbb{P}(X_i = j^2 - \frac{1}{M}\sum_{k=1}^{M} k^2) = 1/M \text{ for all } j \in \{1,\dots,M\}.$$

These variables have mean 0 and $\|X_i\|_{L^\infty(\mathbb{P})} = \max\{\frac{1}{M}\sum_{k=1}^M k^2, M^2 - \frac{1}{M}\sum_{k=1}^M k^2\} \leqslant M^2$. It follows from Proposition 5.2 that

$$\mathbb{P}(|\sum_i X_i| > C\sqrt{dM^4}) \leqslant \mathbb{P}(|\sum_i X_i| > C\sqrt{\sum_{i=1}^d \|X_i\|_{L^\infty(\mathbb{P})}^2}) \leqslant 2\exp(-C^2/2).$$

It follows that we can pick $C = O(1)$ such that for at least $1/2$ of $x \in \{1, \dots, M\}^d$ we have

$$|\sum_i (x_i^2 - \frac{1}{M}\sum_{k=1}^M k^2)| = O(\sqrt{dM^4});$$

equivalently for at least $1/2$ of the points $x \in \{1, \dots, M\}^d$ the sum $\sum_i x_i^2$ comes from a range of length $O(\sqrt{dM^4}) = O(\sqrt{d}M^2)$. By averaging we conclude that there is some r with $|S_r| = \Omega(M^d/\sqrt{d}M^2)$ as required.

It may be worth saying that on a careful examination the reader will see that we did not really use the full power of the Chernoff bound; we could have made do with Chebychev's inequality, but conceptually we are thinking about the result as being a statement about concentration.

In Proposition 5.2 we looked at probabilities of the form

$$\mathbb{P}(|\sum_i X_i| \geqslant C\sqrt{\sum_i \|X_i\|_{L^\infty(\mathbb{P})}^2}),$$

whereas Chebychev's bound looks at probabilities of the form

$$\mathbb{P}(|\sum_i X_i| \geqslant C\sqrt{\sum_i \|X_i\|_{L^2(\mathbb{P})}^2}),$$

since $\mathbb{E}X_i = 0$ implies that $\operatorname{Var} X_i = \|X_i\|_{L^2(\mathbb{P})}^2$. We should like to recover the situation and to some extent we can in the Marcinkiewicz-Zygmund inequality. To understand this we first need a lemma which relates Chernoff-type bounds to inequalities about L^p-norms.

Lemma 5.5. *Suppose that X is a random variable. Then the following are equivalent:*

(i) (Chernoff-type bound) *For all $t \in [0, \infty)$ we have*

$$\mathbb{P}(|X| \geqslant t\|X\|_{L^2(\mathbb{P})}) = O(\exp(-\Omega(t^2))).$$

(ii) (Bounded L^p-norm growth) *For all $p \in [2, \infty)$ we have*

$$\|X\|_{L^p(\mathbb{P})} = O(\sqrt{p}\|X\|_{L^2(\mathbb{P})}).$$

Proof. There is no loss of generality in proving this for random variables with $\|X\|_{L^2(\mathbb{P})} = 1$. We start by showing that (ii) implies (i): let $C > 0$ be the constant in the big-O of the hypothesis and let $c > 0$ be a constant to be optimised later. Note that

$$\mathbb{P}(|X| \geqslant t)\exp(ct^2) \leqslant \mathbb{E}\exp(c|X|^2) = \sum_{k=0}^\infty \frac{c^k}{k!}\mathbb{E}|X|^{2k} \leqslant \sum_{k=0}^\infty \frac{c^k}{k!}.(2C^2k)^k = \sum_{k=0}^\infty O(cC^2)^k.$$

Pick $c = \Omega(1/C^2)$ such that the right hand side is $O(1)$ so that $\mathbb{P}(|X| \geqslant t)\exp(\Omega(t^2)) = O(1)$, and the result follows.

Secondly we show that (i) implies (ii). By nesting of norms it suffices to show (ii) for even integers, and then we note that there is some $c > 0$ such that

$$\|X\|_{L^{2k}(\mathbb{P})}^{2k} = \int_0^\infty 2kt^{2k-1}\mathbb{P}(|X| \geqslant t)dt = O\left(\int_0^\infty 2kt^{2k-1}\exp(-ct^2)dt\right).$$

We now proceed by parts to see that for $r > 1$ we have

$$\begin{aligned}\int_0^\infty t^r\exp(-ct^2)dt &= \left[-t^{r-1}c^{-1}\exp(-ct^2)\right]_0^\infty + c^{-1}\int_0^\infty (r-1)t^{r-2}\exp(-ct^2)dt\\ &= c^{-1}\int_0^\infty (r-1)t^{r-2}\exp(-ct^2)dt,\end{aligned}$$

hence by induction we have $\|X\|_{L^{2k}(\mathbb{P})}^{2k} = O(c^{-1}k)^k$. □

The proof actually gives us that the constant in the Ω in the first case is roughly the reciprocal of the square of the constant in the big-O of the second case. This is not quite true, but it is asymptotically for large p. For small p the big-O in the first hypothesis has an effect.

As an aside, behind much of the above material is the fact that one can define a Banach space of *sub-gaussian* random variables following Kahane [Kah60] to be the real (we made no such assumption above) random variables such that there is some $c > 0$ for which

$$\mathbb{E}\exp(\lambda X) \leqslant \exp(c^2\lambda^2/2) \text{ for all } \lambda \in \mathbb{R};$$

the norm of X is the smallest c such that this holds. The fact that this is a norm is not completely trivial. Homogeneity is easy, but the triangle inequality requires a little more work.

An examination of Lemma 5.1 shows that we actually proved that $\sum_i X_i$ is sub-gaussian with norm at most $\sqrt{\sum_i \|X_i\|_{L^\infty(\mathbb{P})}^2}$.

As a corollary of Proposition 5.2 and Lemma 5.5 we have Khintchine's inequality.

Proposition 5.6 (Khintchine's inequality). *Suppose that $p \in [2,\infty)$ and $X_1,\dots,X_n$ are random variables with $\mathbb{P}(X_i = a_i) = \mathbb{P}(X_i = -a_i) = 1/2$. Then*

$$\|\sum_i X_i\|_{L^p(\mathbb{P})} = O\left(\sqrt{p}\left(\sum_i \|X_i\|_{L^2(\mathbb{P})}^2\right)^{1/2}\right).$$

This can be bootstrapped to the following.

Theorem 5.7 (Marcinkiewicz-Zygmund inequality). *Suppose that $p \in [2,\infty)$ and we are given independent random variables $X_1,\dots,X_n \in L^p(\mathbb{P})$ with $\mathbb{E}\sum_i X_i = 0$. Then*

$$\|\sum_i X_i\|_{L^p(\mathbb{P})} = O\left(\sqrt{p}\|\sum_i |X_i|^2\|_{L^{p/2}(\mathbb{P})}^{1/2}\right).$$

Proof. For complex random variables the result follows from the real case by taking real and imaginary parts and applying the triangle inequality.

We now proceed in two parts. First we prove the inequality with the X_is assumed symmetric[27] (whence the mean of each X_i is automatically 0). Partition the probability space Ω into sets $\Omega_1, \dots, \Omega_M$ (and write $\mathbb{P}_j$ for the induced measure) such that all X_is are symmetric and take at most 2 values on each Ω_j. Then by Khintchine's inequality we have that

$$\|\sum_i X_i\|_{L^p(\mathbb{P}_j)}^p = O(p)^{p/2}\|\sum_i |X_i|^2\|_{L^{p/2}(\mathbb{P}_j)}^{p/2}.$$

Summing over j and taking roots gives the result in the symmetric case.

Now we suppose that the variables $X_1, \dots, X_n$ are given and $Y_1, \dots, Y_n$ are such that $X_i \sim Y_i$ and $X_1, \dots, X_n, Y_1, \dots, Y_n$ are independent. We now apply the symmetric result to the variables $X_i - Y_i$ to get that

$$\begin{aligned}\|\sum_i (X_i - Y_i)\|_{L^p(\mathbb{P}\times\mathbb{P})} &= O\left(\sqrt{p}\|\sum_i |X_i - Y_i|^2\|_{L^{p/2}(\mathbb{P}\times\mathbb{P})}^{1/2}\right)\\ &= O\left(\sqrt{p}\|\sum_i |X_i|^2\|_{L^{p/2}(\mathbb{P}\times\mathbb{P})}^{1/2}\right).\end{aligned}$$

But then it follows from nesting of norms and the fact that $\mathbb{E}\sum_i Y_i = 0$ that

$$\|\sum_i X_i\|_{L^p(\mathbb{P})} = \|\sum_i X_i - \mathbb{E}\sum_i Y_i\|_{L^p(\mathbb{P})} \leqslant \|\sum_i (X_i - Y_i)\|_{L^p(\mathbb{P}\times\mathbb{P})},$$

and the result is proved. □

For random variables satisfying the hypotheses of Khintchine's inequality the $L^{p/2}$-norm on the right is an L^1-norm, and there is something close to this true for variables in the generality considered above called Rosenthal's inequality. Indeed, suppose that $X_1, \dots, X_n \in L^p(\mathbb{P})$ are independent and $\mathbb{E}\sum_i X_i = 0$. Then

$$\|\sum_i X_i\|_{L^p(\mathbb{P})} = O\left(\frac{p}{\log p}\max\left\{\left(\sum_i \|X_i\|_{L^p(\mathbb{P})}^p\right)^{1/p}, \|\sum_i X_i\|_{L^2(\mathbb{P})}\right\}\right). \tag{5.4}$$

For p large the second term in the max takes over and we recover a strengthening of Khintchine's inequality. Of course, precisely when this takes over depends on the specific variables X_i and how large their L^p mass is compared to their L^2 – that is how often they take *very* large values.

The p dependence in (5.4) is best possible (up to the precise constant; see [JSZ85] for details), and it is weaker than that for the Marcinkiewicz-Zygmund inequality. This fits with the fact that the critical distributions for Rosenthal's inequality are Poisson whereas for the Marcinkiewicz-Zygmund inequality they are Gaussians.

[27] That is when $X_i \sim -X_i$; equivalently $\mathbb{P}(X_i = a) = \mathbb{P}(X_i = -a)$ for all $a \in \mathbb{R}$.

5.8. **Constants in Khintchine's inequality.** The example we considered after Proposition 5.2 to show that nothing much stronger is true also applies to show that Khintchine's inequality is best possible up to the constants. Of course we are not, for the most part, actually interested in the specific constants although they are known (see [Haa81] for the original proof, or [NP00] for a very accessible and more recent proof).

It is should be said that the price we pay for the a_js being arbitrary in Khintchine's inequality is that all bar one of them could be zero (or just very small) so that the term we are looking to bound is dominated by one random variable. This shows us that while for $p \in [2, \infty)$ we have

$$\|\sum_i X_i\|_{L^p(\mathbb{P})} \geqslant \|\sum_i X_i\|_{L^2(\mathbb{P})} = \left(\sum_i \|X_i\|_{L^2(\mathbb{P})}^2\right)^{1/2}$$

by nesting of norms, this cannot be improved and so in particular there is no matching lower bound to Khitchine's inequality.

We have concentrated on comparing the L^p-norm with the L^2-norm for $p \geqslant 2$ so far, but things can also be said for $p \in [1, 2]$. On the one hand

$$\|\sum_i X_i\|_{L^1(\mathbb{P})} \leqslant \|\sum_i X_i\|_{L^2(\mathbb{P})}$$

by nesting of norms; on the other it turns out that we have

$$\|\sum_i X_i\|_{L^2(\mathbb{P})} = O(\|\sum_i X_i\|_{L^1(\mathbb{P})}). \tag{5.5}$$

To see this note by log-convexity of L^p-norms that

$$\|\sum_i X_i\|_{L^2(\mathbb{P})}^2 \leqslant \|\sum_i X_i\|_{L^4(\mathbb{P})}^{4/3} \|\sum_i X_i\|_{L^1(\mathbb{P})}^{2/3},$$

and then apply Khintchine's inequality for $p = 4$ and rearrange. As a result of (5.5) and the remarks immediately before it we conclude that

$$\|\sum_i X_i\|_{L^1(\mathbb{P})} \leqslant \|\sum_i X_i\|_{L^p(\mathbb{P})} = O(\sqrt{p}\|\sum_i X_i\|_{L^1(\mathbb{P})}) \text{ for all } p \in [1, \infty)$$

which is often the result that is actually called Khintchine's inequality in the literature; the result we call Khintchine's inequality is seen as a special case.

There is a rather nice argument due to Latała and Oleszkiewicz [LO94] using the Fourier transform which actually gives the correct constant in (5.5). To get a lower bound on the constant consider X_1 and X_2 independent random variables with mean 0 taking the values 1 and -1. Then

$$\|X_1 + X_2\|_{L^1(\mathbb{P})} = 1 \text{ and } \|X_1 + X_2\|_{L^2(\mathbb{P})} = \sqrt{2},$$

so the constant is at least $\sqrt{2}$. This turns out to be the worst case.

Theorem 5.9. *Suppose that $X_1, \ldots, X_n$ are independent random variables with $\mathbb{P}(X_i = a_i) = \mathbb{P}(X_i = -a_i) = 1/2$. Then*

$$\|\sum_i X_i\|_{L^2(\mathbb{P})} \leqslant \sqrt{2}\|\sum_i X_i\|_{L^1(\mathbb{P})}.$$

Proof. We put $G := (\mathbb{Z}/2\mathbb{Z})^n$ which we think of as endowed with Haar probability measure, $\gamma_i(x) := (-1)^{x_i}$ and

$$f(x) := \left|\sum_{i=1}^n a_i\gamma_i(x)\right|.$$

Now by symmetry the value of $f(x)$ and $f(x + \sum_j x_j)$ are the same and so for all i we have $\widehat{f}(\gamma_i) = 0$.

The Laplacian of f is then defined via the measure $\nu := \frac{1}{2}\sum_i (\delta_{0_G} - \delta_{e_i})$ where the e_is are the canonical basis vectors. This means the Laplacian of f is $f * \nu$.

Now, by the triangle inequality we see that

$$\begin{aligned} f * \nu(x) &= \frac{1}{2}\left(nf(x) - \sum_i \left|\sum_j a_j\gamma_j(x) - 2a_i\gamma_i(x)\right|\right) \\ &\leqslant \frac{1}{2}\left(nf(x) - \left|n\sum_j a_j\gamma(x) - 2\sum_i a_i\gamma_i(x)\right|\right) = f(x). \end{aligned}$$

On the other hand the Fourier transform of ν is easy to compute:

$$\widehat{\nu}(\gamma) = \sum_i \frac{1}{2}(1 - \gamma(e_i)) = |\gamma|,$$

that is to say it is the number of 1s in γ when it is written with respect to the standard basis.

Combining this information with Parseval's theorem we have

$$\|f\|_{L^2(G)}^2 \geqslant \langle f * \nu, f\rangle_{L^2(G)} = \sum_{\gamma\in\widehat{G}} |\gamma||\widehat{f}(\gamma)|^2 \geqslant 2\sum_{|\gamma|\geqslant 2} |\widehat{f}(\gamma)|^2 = 2(\|f\|_{L^2(G)}^2 - |\widehat{f}(0_{\widehat{G}})|^2),$$

and the result follows on rearranging since $f \geqslant 0$ and so $\widehat{f}(0_{\widehat{G}}) = \|f\|_{L^1(G)}$. □

It should be remarked that this proof extends to random variables taking values in an arbitrary normed space; for complex values the result was already known but without such an elegant proof.

5.10. **Harmonic analysis of thin sets: Rudin's inequality.** Rudin's inequality is an extension of Khintchine's inequality to the case when the random variables are characters which do not necessarily need to be totally independent. Before starting it is useful to consider the example of $G := (\mathbb{Z}/2\mathbb{Z})^n$ endowed with Haar probability measure. The dual group $\widehat{G}$ can be thought of in two ways:

(i) *(Probabilistically)* The characters in $\widehat{G}$ are random variables on G and there is a notion of statistical independence for a sequence of characters $\gamma_1, \dots, \gamma_n \in \widehat{G}$;

(ii) *(Algebraically)* The group $\widehat{G}$ is naturally a vector space over $\mathbb{F}_2$ and there is a notion of algebraic independence for a sequence of characters $\gamma_1, \dots, \gamma_n \in \widehat{G}$.

It turns out that in this setting these two notions are equivalent.[28]

In a general finite Abelian group G the algebraic notion of independence has a very useful extension called dissociativity which we now define.

Definition 5.11 (Dissociativity)**.** Suppose that G is a finite Abelian group. We say that $S \subset G$ is *dissociated* if

$$\sum_{s\in S} \epsilon_s s = 0_G \text{ and } \epsilon \in \{-1,0,1\}^S \Rightarrow \epsilon \equiv 0.$$

Moreover, we write

$$\mathrm{Span}(S) := \{\sum_{s\in S} \epsilon_s s : \epsilon \in \{-1,0,1\}^S\}.$$

It may be useful to note that a set in $G = (\mathbb{Z}/2\mathbb{Z})^n$ is dissociated if and only if it is independent; in $G = \mathbb{Z}/p\mathbb{Z}$ this is not the case and an example of a dissociated set of size greater than 1 (that being the largest size of a genuinely independent set) is given by $\{1, 2, 4, \dots, 2^r\}$ for $r \leqslant \log_3 |G|$.

Algebraically a set is contained in the space generated by any maximal independent subset. With dissociativity and span we have a similar relationship.

Lemma 5.12. *Suppose that S is a maximal dissociated subset of T. Then $T \subset \mathrm{Span}(S)$.*

Proof. The proof is as for the algebraic version of this: suppose that $t \in T\backslash S$. Then adjoining t to S violates the dissociativity condition since otherwise we would contradict the maximality of S. Thus there is some $\epsilon \in \{-1,0,1\}^{S\cup\{t\}}$ with $\epsilon_t \neq 0$ such that

$$\epsilon_t t + \sum_{s\in S} \epsilon_s s = 0_G.$$

It follows that $t \in \mathrm{Span}(S)$. On the other hand $S \subset \mathrm{Span}(S)$ trivially and so the lemma is proved. □

[28] Since the characters γ are homomorphisms we see that any $y, z \in \{x : \gamma(x) = w_\gamma \text{ for all } \gamma \in \Lambda'\}$ have $\gamma(y - z) = 1$ for all $\gamma \in \Lambda'$ and so the set is just a translate of the annihilator of Λ'. Thus Λ is statistically independent iff

$$\mathbb{P}_G(\Lambda'^{\perp}) = \prod_{\gamma\in\Lambda'} \mathbb{P}_G(\{\gamma\}^{\perp}) \text{ for all } \Lambda' \subset \Lambda.$$

Now, if Λ is algebraically independent then none of the γs in Λ are identically 1 and so $\mathbb{P}_G(\{\gamma\}^{\perp}) = 1/2$ for all $\gamma \in \Lambda$. On the other hand $(\Lambda'^{\perp})^{\perp}$ is the subspace generated by Λ' which has size 2^d since Λ' is independent. Hence $\mathbb{P}_G(\Lambda'^{\perp}) = \mathbb{P}_G(((\Lambda'^{\perp})^{\perp})^{\perp}) = 2^{-d}$ and we see that Λ is statistically independent.

On the other hand if Λ is statistically independent then by a similar argument the subspace generated by Λ has size $2^{|\Lambda|}$ and hence is $|\Lambda|$-dimensional. It follows that Λ is algebraically independent.

Dissociativity is really a property of sets in groups endowed with Haar counting measure. In the example at the start of the subsection where we considered $G = (\mathbb{Z}/2\mathbb{Z})^n$ endowed with Haar probability measure, but that meant that $\widehat{G}$ was endowed with Haar counting measure and so was the natural setting for dissociativity.

It turns out that when a set of characters is dissociated we have an analogue of Khintchine's inequality (and Chernoff's bound) which can be established in the same way as the Chernoff bounds.[29]

Theorem 5.13 (Rudin's inequality)**.** *Suppose that Γ is a dissociated set of characters and $p \in [2,\infty)$. Then*

$$\|f^\vee\|_{L^p(G)} = O(\sqrt{p}\|f\|_{\ell^2(\Gamma)}) \textit{ for all } f \in \ell^2(\Gamma).$$

Proof. We proceed as in the proof of Lemma 5.1. Suppose that $f \in \ell^2(\Gamma)$ and begin by noting that

$$\mathbb{E}\exp(\operatorname{Re} f^\vee) = \mathbb{E}\prod_{\gamma\in\Gamma}\exp(\operatorname{Re}(f(\gamma)\gamma)).$$

As before we see that

$$\begin{aligned}\mathbb{E}\exp(\operatorname{Re} f^\vee) &\leqslant \mathbb{E}\prod_{\gamma\in\Gamma}\left(\cosh|f(\gamma)| + \frac{\operatorname{Re}(f(\gamma)\gamma)}{|f(\gamma)|}\sinh|f(\gamma)|\right)\\ &= \left(\prod_{\gamma\in\Gamma}\cosh|f(\gamma)|\right)\mathbb{E}\prod_{\gamma\in\Gamma}\left(1+\frac{\sinh|f(\gamma)|}{\cosh|f(\gamma)|}\frac{(f(\gamma)\gamma+\overline{f(\gamma)}\overline{\gamma})}{2|f(\gamma)|}\right),\end{aligned}$$

with the obvious convention that the factor is 1 if $f(\gamma) = 0$. Now, when we multiply out the second product on the right we get terms of the form

$$\prod_{\gamma\in S}\frac{f(\gamma)\sinh|f(\gamma)|}{2|f(\gamma)|\cosh|f(\gamma)|}\cdot\prod_{\gamma\in T}\frac{\overline{f(\gamma)}\sinh|f(\gamma)|}{2|f(\gamma)|\cosh|f(\gamma)|}\cdot\prod_{\gamma\in S}\gamma\cdot\prod_{\gamma\in T}\overline{\gamma} \tag{5.6}$$

where S and T are disjoint subsets of Γ. By dissociativity we have

$$\prod_{\gamma\in S}\gamma\cdot\prod_{\gamma\in T}\overline{\gamma} = 0_{\widehat{G}} \text{ if and only if } S = T = \emptyset,$$

since S and T are disjoint subsets of Γ. On the other hand if $\gamma \neq 0_{\widehat{G}}$ then $\mathbb{E}\gamma = 0$ and hence the expectation of (5.6) is 0 unless $S = T = \emptyset$ which which case we see that it is 1. It follows that

$$\mathbb{E}\exp(\operatorname{Re} f^\vee) \leqslant \left(\prod_{\gamma\in\Gamma}\cosh|f(\gamma)|\right) \leqslant \exp(\frac{1}{2}\sum_{\gamma\in\Gamma}|f(\gamma)|^2).$$

We now have a conclusion of the shape of Lemma 5.1 and so we could follow the arguments of Proposition 5.2 and then Lemma 5.5 to get the desired conclusion. We shall proceed

[29] The proof below is different from the one lectured, but seems easier to understand given the difficulties we had with dissociativity.

directly instead: suppose that $\lambda \in \mathbb{R}$ is arbitrary. Then

$$\begin{aligned} \mathbb{E}\sum_{n=0}^{\infty} \frac{\lambda^{2n}(\operatorname{Re} f^{\vee})^{2n}}{(2n)!} &= \frac{1}{2}\mathbb{E}(\exp(\operatorname{Re}(\lambda f)^{\vee}) + \exp(\operatorname{Re}(-\lambda f)^{\vee})) \\ &\leqslant \exp(\frac{1}{2}\lambda^2\|f\|_{\ell^2(\Gamma)}^2) = \sum_{n=0}^{\infty} \frac{\lambda^{2n}\|f\|_{\ell^2(\Gamma)}^{2n}}{2^n n!}. \end{aligned}$$

By equating coefficients of λ and then taking $2n$-th roots we conclude that

$$\|\operatorname{Re} f^{\vee}\|_{L^{2n}(G)} \leqslant \left(\frac{(2n)!}{2^n n!}\right)^{1/2n} \|f\|_{\ell^2(\Gamma)}.$$

The result follows for even integers by the triangle inequality and applying this for f and if. Nesting of norms then completes the proof. □

It is worth noting that Rudin's inequality actually *is* Khintchine's inequality when $G = (\mathbb{Z}/2\mathbb{Z})^n$ since then being dissociated is the same as being algebraically independent which, as we noted earlier, is the same as being statistically independent.

Dissociated sets are examples of Sidon sets. The first basic result about Sidon sets is Rudin's inequality, and that along with some more basic results may be found in [Rud90, §5.7]. There is also the dedicated book [LR75] of López and Ross and some material in the book [GM79] of Graham and McGehee, although both of these are harder to find.

There is also one further remark relating to §5.8 on the constants in Khintchine's inequality which may be of interest. Contained in the preceding proof of Rudin's inequality was a version of Khintchine's inequality for real random variables with a particularly good constant: we showed that

$$\|\sum_i X_i\|_{L^{2n}(\mathbb{P})} \leqslant \left(\frac{(2n)!}{2^n n!}\right)^{1/2n} \left(\sum_i \|X_i\|_{L^2(\mathbb{P})}^2\right)^{1/2}.$$

It turns out that this constant is tight. Of course, we use nesting of norms to deduce Khintchine's inequality for L^p-norms where p is not an even integer, and so it will not be surprising that for these other values of p the constants are not tight. In fact for other values of $p \in [2, \infty)$ the tight inequality is

$$\|\sum_i X_i\|_{L^p(\mathbb{P})} \leqslant \left(2^{(p-2)/2}\frac{\Gamma((p+1)/2)}{\Gamma(3/2)}\right)^{1/p} \left(\sum_i \|X_i\|_{L^2(\mathbb{P})}^2\right)^{1/2},$$

which coincides with the above for p an even integer. The critical example arises by taking the X_is to be identically distributed.

⚠**Warning** ⚠ *In number theory the term Sidon set is used for something also called a $B_2[1]$-set. It refers to any set of positive integers A such that ay $n \in \mathbb{N}$ has at most one representation $n = a + b$ with $a, b \in A$ and $a \leqslant b$. The reader should compare this with the condition of having no proper additive quadruples considered in Proposition 1.5. In general these number theoretic Sidon sets are much bigger than analytic Sidon sets.*

5.14. **Application: structure of the large spectrum.** We have often found ourselves with some $A \subset G$ of density α, examining the set of characters at which its Fourier transform is large[30]:

$$\Gamma := \{\gamma \in \widehat{G} : |\widehat{1_A}(\gamma)| \geqslant \epsilon\alpha\}.$$

There are two trivial pieces of structural information about Γ:

(i) Γ is symmetric since 1_A is real;
(ii) $0_{\widehat{G}} \in \Gamma$ since 1_A is non-negative.

Moreover, we have the so-called 'Parseval bound' which says that[31]

$$|\Gamma| \leqslant \epsilon^{-2}\alpha^{-1}. \tag{5.7}$$

The question we now consider is whether we can say anything more about the structure of the set Γ. It turns out when A has density tending to 0 we can and we have a celebrated result of Chang [Cha02].

Theorem 5.15 (Chang's theorem)**.** *Suppose that A is a subset of G of density $\alpha > 0$ and Γ is a dissociated subset of $\{\gamma \in \widehat{G} : |\widehat{1_A}(\gamma)| \geqslant \epsilon\alpha\}$. Then*

$$|\Gamma| = O(\epsilon^{-2}\log\alpha^{-1}).$$

Proof. Given such a Γ, by Rudin's inequality the operator

$$T : \ell^2(\Gamma) \to L^p(G); f \mapsto f^\vee$$

has norm $O(\sqrt{p})$. It follows that its adjoint T^* also has norm $O(\sqrt{p})$, but a short calculation[32] shows us that the adjoint is just the restriction of the Fourier transform:

$$T^* : L^{p'}(G) \to \ell^2(\Gamma); f \mapsto \widehat{f}|_\Gamma$$

where $1/p + 1/p' = 1$. It follows that

$$|\Gamma|(\epsilon\alpha)^2 \leqslant \sum_{\gamma\in\Gamma}|\widehat{1_A}(\gamma)|^2 = \|T^*1_A\|_{\ell^2(\Gamma)}^2 = O(p\|1_A\|_{L^{p'}(G)}^2) = O(p\alpha^{2/p'}) = \alpha^2 O(p\alpha^{-2/p}).$$

We optimise this by setting $p = 2 + \log\alpha^{-1}$ and arrive at the result on rearranging. □

Chang's theorem has a huge number of applications and a number of proofs. When comparing it with the Parseval bound (5.7) it should be clear that we gain when $\alpha \to 0$.

We now turn to a sample application where we improve the dependencies in Theorem 4.10.

Theorem 5.16 (Theorem 4.10, improved)**.** *Suppose that $A \subset \{1,\dots,N\}$ has size αN. Then $A + A + A$ contains an arithmetic progression of length $\Omega(\alpha^{3+o(1)}N^{\Omega(\alpha^{2+o(1)})})$.*

[30] As usual, since we are considering density we regard G as endowed with Haar probability measure.

[31] The proof is simply the usual application of Parseval's theorem to show us that

$$(\epsilon\alpha)^2|\Gamma| \leqslant \sum_{\gamma\in\Gamma}|\widehat{1_A}(\gamma)|^2 \leqslant \sum_{\gamma\in\widehat{G}}|\widehat{1_A}(\gamma)|^2 = \alpha.$$

[32] Simply check that $\langle f, Tg\rangle_{L^2(G)} = \langle T^*f, g\rangle_{\ell^2(\Gamma)}$ for all $f \in L^{p'}(G)$ and $g \in \ell^2(\Gamma)$.

Sketch proof. We proceed as in the proof before to get a set A' (which is morally the same as the set A), a set

$$\Gamma := \{\gamma \neq 0_{\widehat{G}} : |\widehat{1_{A'}}(\gamma)| \geqslant \alpha'^2/3\},$$

and an element x_0 such that for all $y \in \mathrm{Bohr}(\Gamma, \alpha'/3)$ we have

$$1_{A'} * 1_{A'} * 1_{A'}(x_0 + y) > 0.$$

Let $\Lambda \subset \Gamma$ be maximal dissociated and suppose that $y \in \mathrm{Bohr}(\Lambda, \alpha'/3|\Lambda|)$. Then by Lemma 5.12 we have that $\Gamma \subset \mathrm{Span}(\Lambda)$ and so given $\gamma \in \Gamma$ there is some $\epsilon \in \{-1, 0, 1\}^\Lambda$ such that $\gamma = \prod_{\lambda \in \Lambda} \lambda^{\epsilon_\lambda}$. Thus, by the triangle inequality

$$|\gamma(y) - 1| = |\prod_{\lambda \in \Lambda} \lambda^{\epsilon_\lambda}(y) - 1| \leqslant \sum_{\lambda \in \Lambda} |\lambda^{\epsilon_\lambda}(y) - 1| \leqslant |\Lambda|.(\alpha'/3|\Lambda|) = \alpha'/3,$$

and hence $\mathrm{Bohr}(\Lambda, \alpha'/3|\Lambda|) \subset \mathrm{Bohr}(\Gamma, \alpha'/3)$. Now by Chang's theorem we have the bound $|\Lambda| = O(\alpha'^{-2} \log \alpha'^{-1})$ and the result now follows from Lemma 4.9 applied to $\mathrm{Bohr}(\Lambda, \alpha'/3|\Lambda|)$. □

In fact it is possible to use the Marcinkiewicz-Zygmund inequality in a rather clever way to improve this further and get an arithmetic progression of length $\Omega(\alpha^{O(1)} N^{\Omega(\alpha^{1+o(1)})})$. The details of this may be found in [Hen12, Theorem 1.5] and make use of the rather powerful techniques of Croot and Sisask as developed in [CS10] and [CŁS11].

5.17. **Application: limitations on the structure of the large spectrum.** In the same setting as above we can ask about what happens if $\alpha = \Omega(1)$; in this case Chang's bound is no better than Parseval's bound since $\alpha^{-1} = O(1)$ and $\log \alpha^{-1} = O(1)$. One might be hopeful since $1_A^2 = 1_A$ and so

$$\widehat{1_A} * \widehat{1_A}(\gamma) = \widehat{1_A}(\gamma) \text{ for all } \gamma \in \widehat{G},$$

which looks like it places restrictions on what $\widehat{1_A}$ can look like. It turns out somewhat surprisingly that it does not significantly impact the modulus, and we have the following theorem.

Theorem 5.18. *Suppose that $\Gamma \subset \widehat{G}$ is a symmetric neighbourhood of $0_{\widehat{G}}$ of size k. Then there is a set $A \subset G$ of density $\Omega(1)$ such that*

$$|\widehat{1_A}(\gamma)| = \Omega(k^{-1/2}) \text{ for all } \gamma \in \Gamma$$

provided k is sufficiently small.[33]

We prove this in two parts. First we need a lemma which shows that L^∞ functions are much the same as sets.

Lemma 5.19. *Suppose that $f \in C(G)$ is a real-valued function with $\|h\|_{L^\infty(G)} \leqslant M$. Then there is some set $A \subset G$ with $\mathbb{P}_G(A) = \Omega(1/M)$ and*

$$\left|\widehat{1_A}(\gamma) - \frac{\widehat{h}(\gamma)}{3M}\right| = O\left(\frac{\log |G|}{|G|}\right)^{1/2} \text{ for all } \gamma \neq 0_{\widehat{G}}.$$

[33] Taking $k = o(|G|/\log |G|)$ works and this can hardly be said to be a major restriction.

Sketch proof. The argument is essentially the same as that for Example 5.3. We pick $x \in G$ with independently with probability $(h(x) + 2M)/3M$ and let A be the resulting set. □

With this information at hand we have the following theorem of de Leeuw, Kahane and Katznelson [dLKK77].

Theorem 5.20. *Suppose that $f \in \ell^2(\widehat{G})$. Then there is some real-valued $h \in C(G)$ such that $|\widehat{h}(\gamma)| \geqslant |f(\gamma)|$ for all $\gamma \in \widehat{G}$ and $\|h\|_{L^\infty(G)} = O(\|f\|_{\ell^2(\widehat{G})})$.*

Proof. If f is symmetric then it is easy to check that the argument below produces a real function, and if it is not symmetric then we replace f by $\gamma \mapsto |f(\gamma)| + |f(-\gamma)|$ (which is symmetric) and apply the argument to that from which the result follows (with slightly worse constants).

Our starting point, then, is a '99%' version of the conclusion which we distill into the following claim which illustrates the power of Khintchine's inequality already for $p = 4$.

Claim. *Given $k \in \ell^2(\widehat{G})$ and $\eta \in (0, 1/2]$ there is some choice of signs ϵ on the support of k (meaning $\epsilon : \operatorname{supp} k \to \{-1, 1\}$) and a function g with*

$$\|\widehat{g} - \epsilon k\|_{\ell^2(\widehat{G})} \leqslant \eta \|k\|_{\ell^2(\widehat{G})} \text{ and } \|g\|_{L^\infty(G)} = O(\eta^{-1}\|k\|_{\ell^2(\widehat{G})}).$$

Proof. We suppose that $(\epsilon_\gamma)_{\gamma \in \operatorname{supp} k}$ are independent and $\mathbb{P}(\epsilon_\gamma = -1) = \mathbb{P}(\epsilon_\gamma = 1) = 1/2$, and examine

$$g_\epsilon(x) = \sum_{\gamma \in \operatorname{supp} k} \epsilon_\gamma k(\gamma)\gamma(x).$$

By linearity, uniqueness and Fourier inversion we have $\widehat{g_\epsilon}(\gamma) = \epsilon_\gamma k(\gamma)$ for all $\gamma \in \widehat{G}$. Now $|\gamma(x)| = 1$ for all $x \in G$ and so

$$\|k\|_{\ell^2(\widehat{G})}^2 = \sum_{\gamma \in \widehat{G}} |k(\gamma)|^2 = \sum_{\gamma \in \operatorname{supp} k} |k(\gamma)\gamma(x)|^2$$

for all $x \in G$. It follows by Khintchine's inequality (for fixed $x \in G$ with random variables $X_\gamma := \epsilon_\gamma k(\gamma)\gamma(x)$) that we have

$$\mathbb{E}\|g_\epsilon\|_{L^4(G)}^4 = \mathbb{E}_{x \in G}\mathbb{E}|g_\epsilon(x)|^4 = \mathbb{E}_{x \in G} O(\|k\|_{\ell^2(\widehat{G})}^4).$$

It follows that we can make a choice of signs ϵ (supported on $\operatorname{supp} k$) such that

$$\|g_\epsilon\|_{L^4(G)}^4 = O(\|k\|_{\ell^2(\widehat{G})}^4).$$

Finally we let

$$g(x) := \begin{cases} g_\epsilon(x) & \text{if } |g_\epsilon(x)| \leqslant C\|k\|_{\ell^2(\widehat{G})} \\ 0 & \text{otherwise,} \end{cases}$$

and note that

$$\|g - g_\epsilon\|_{L^2(G)}^2 (C\|k\|_{\ell^2(\widehat{G})})^2 = O(\|k\|_{\ell^2(\widehat{G})}^4).$$

By Parseval's theorem and our earlier calculation of $\widehat{g_\epsilon}$ we conclude that

$$\|g - \epsilon k\|_{\ell^2(\widehat{G})} = O(C^{-1}\|k\|_{\ell^2(\widehat{G})}).$$

Optimising we can take $C = O(\eta^{-1})$ such that the right hand side is at most $\eta\|k\|_{\ell^2(\widehat{G})}$ and the claim is proved. □

We define functions $h_1, \dots, h_r$ iteratively as better and better approximations to our eventual function h. We think of them as 'almost-majorants'[34] and to measure their quality we introduce a sequence of positive real parameters $(\delta_i)_i$ and sets

$$\Gamma_r := \{\gamma \in \widehat{G} : |\widehat{h_r}(\gamma)| < (1+\delta_r)|f(\gamma)|\};$$

this set is the set where $\widehat{h_r}$ fails to 'robustly majorise' f. A sensible choice for the δ_is emerges from the proof below.

We write $f_r := f|_{\Gamma_r}$ and the plan is to use the above claim to produce a function g_r such that $|\widehat{g_r}|$ is a good approximation to $|f_r|$ and then add this to h_r to get h_{r+1}. The hope is that this new function is 'better' in the sense that Γ_{r+1} is smaller than Γ_r.

We let $(\eta_i)_i$ be another sequence of parameters which will be optimised later. Applying the claim to $4f_r$ with parameter η_r we get g_r such that

$$\|\widehat{g_r} - 4\epsilon f_r\|_{\ell^2(\widehat{G})} \leqslant \eta_r\|f_r\|_{\ell^2(\widehat{G})} \text{ and } \|g_r\|_{L^\infty(G)} = O(\eta_r^{-1}\|f_r\|_{\ell^2(\widehat{G})}),$$

and then put $h_{r+1} := h_r + g_r$. Suppose that $\gamma \in \Gamma_{r+1}$. Then we have two possibilities:

(i) $\gamma \in \Gamma_r$: which is to say that γ was 'bad' for h_r and we have not succeeded in dealing with it in h_{r+1}. In this case we have

$$|\widehat{g_r}(\gamma)| \leqslant |\widehat{h_{r+1}}(\gamma)| + |\widehat{h_r}(\gamma)| < (2+\delta_r+\delta_{r+1})|f(\gamma)|,$$

and so

$$\begin{aligned} |\widehat{g_r}(\gamma) - 4\epsilon_\gamma f_r(\gamma)| &\geqslant 4|f(\gamma)| - |\widehat{g_r}(\gamma)| \\ &\geqslant (2-\delta_r-\delta_{r+1})|f(\gamma)| \geqslant (\delta_r - \delta_{r+1})|f(\gamma)|. \end{aligned}$$

This last inequality holds if we take $\delta_r \leqslant 1$ for all r which we can certainly do and is, in any case, very weak. It can easily be replaced with a lower bound of $|f(\gamma)|$ for example, but we make the estimate in light of the next case.

(ii) $\gamma \notin \Gamma_r$: which is to say that γ was 'good' but adding in our new approximation has made things worse and it is now 'bad'. In this case we have

$$\begin{aligned} |\widehat{g_r}(\gamma)| &\geqslant |\widehat{h_r}(\gamma)| - |\widehat{h_{r+1}}(\gamma)| \\ &\geqslant (1+\delta_r)|f(\gamma)| - (1+\delta_{r+1})|f(\gamma)| \geqslant (\delta_r - \delta_{r+1})|f(\gamma)|. \end{aligned}$$

Of course $f_r(\gamma) = 0$ and so it follows that

$$|\widehat{g_r}(\gamma) - 4\epsilon_\gamma f_r(\gamma)| \geqslant (\delta_r - \delta_{r+1})|f(\gamma)|.$$

In both cases we have shown the same inequality and so we conclude that

$$\begin{aligned} \|f\|^2_{\ell^2(\Gamma_{r+1})} = \sum_{\gamma\in\Gamma_{r+1}} |f(\gamma)|^2 &\leqslant (\delta_r - \delta_{r+1})^{-2}\|\widehat{g_r} - 4\epsilon f_r\|^2_{\ell^2(\widehat{G})} \\ &\leqslant (\delta_r - \delta_{r+1})^{-2}\eta_r^2\|f_r\|^2_{\ell^2(\widehat{G})} = (\delta_r - \delta_{r+1})^{-2}\eta_r^2\|f\|^2_{\ell^2(\Gamma_r)}. \end{aligned}$$

[34] A function F majorises a function G if $F(x) \geqslant G(x)$ for all x.

Now, since G is finite, f has a minimum non-zero modulus and so taking $\eta_r := (\delta_r - \delta_{r+1})/2$ we see that after a finite number, say r, iterations $\Gamma_r = \emptyset$ at which point we terminate and put $h := h_r$. Since Γ_r is empty we certainly have that $\widehat{h}$ majorises f; it remains to note that

$$\|h\|_{L^\infty(G)} = O\left(\sum_{i=1}^{\infty} 2^{-i}(\delta_i - \delta_{i+1})^{-1}\right)\|f\|_{\ell^2(\widehat{G})}.$$

For the right hand side to converge we can take, for example, $\delta_i = 1/i^2$ and the result is proved. □

It should be noted that the proof only really needed Khintchine's inequality for some $p > 2$, although since we deduced Khintchine for $p \in (2, 4)$ from Khintchine for $p = 4$ this is not immediately useful. It turns out, however, that there is a wonderful generalisation due to Nazarov [Naz97] which takes an even weaker input.

Suppose that $(\phi_j)_j$ of unit vectors in $L^1(\nu)$ satisfying an $L^1(\nu) - \ell^2$ Khintchine inequality, that is to say such that

$$\|\sum_j a_j\phi_j\|_{L^1(\nu)} \leqslant M\left(\sum_j a_j^2\right)^{1/2}$$

for all real sequences $(a_j)_j$. Then Nazarov proved that for any sequence $(f_j)_j$ of positive numbers there is a function $g \in L^\infty(\nu)$ with

$$\|g\|_{L^\infty(\nu)} = O(M^2\|f\|_{\ell^2}) \text{ and } |\langle \phi_j, g\rangle_{L^2(\nu)}| \geqslant f_j \text{ for all } j.$$

In our case the ϕ_js were characters and the f_js were the values of f. Of course when the ϕ_js are characters, (5.17) follows from Khintchine's inequality for any $p > 2$ by the same argument we used to deduce (5.5), but it is a strictly weaker assumption.

Proof of Theorem 5.18. We simply take $f = 1_\Gamma$ which is symmetric and apply Theorem 5.20 to get a function h. Lemma 5.19 then completes the argument. □

⚠**Warning** ⚠ *Note that the function produced in Theorem 5.20 is a majorant not an approximation, so that its Fourier transform may be large at more characters than just those at which f is large. In particular this result can be used to find counter-examples of the form 'there are continuous functions whose large spectrum contains the following bad structure', but is* not *useful for showing that 'there are continuous functions whose large spectrum does not contain the following good structure'.*

References

[AS08] N. Alon and J. H. Spencer. *The probabilistic method.* Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons Inc., Hoboken, NJ, third edition, 2008. With an appendix on the life and work of Paul Erdős.

[Bec75] W. Beckner. Inequalities in Fourier analysis. *Ann. of Math. (2)*, 102(1):159–182, 1975.

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

[Bou99] J. Bourgain. On triples in arithmetic progression. *Geom. Funct. Anal.*, 9(5):968–984, 1999.

[Cha02] M.-C. Chang. A polynomial bound in Freĭman's theorem. *Duke Math. J.*, 113(3):399–419, 2002.

[CŁS11] E. S. Croot, I. Łaba, and O. Sisask. Arithmetic progressions in sumsets and L^p-almost-periodicity. 2011, arXiv:1103.6000.

[CS10] E. S. Croot and O. Sisask. A probabilistic technique for finding almost-periods of convolutions. *Geom. Funct. Anal.*, 20(6):1367–1396, 2010.

[dLKK77] K. de Leeuw, J.-P. Kahane, and Y. Katznelson. Sur les coefficients de Fourier des fonctions continues. *C. R. Acad. Sci. Paris Sér. A-B*, 285(16):A1001–A1003, 1977.

[Ede04] Y. Edel. Extensions of generalized product caps. *Des. Codes Cryptogr.*, 31(1):5–14, 2004.

[Elk10] M. Elkin. An improved construction of progression-free sets. In *Symposium on Discrete Algorithms*, pages 886–905, 2010, arXiv:0801.4310.

[FHR92] G. A. Freiman, H. Halberstam, and I. Z. Ruzsa. Integer sum sets containing long arithmetic progressions. *J. London Math. Soc. (2)*, 46(2):193–201, 1992.

[Fou77] J. J. F. Fournier. Sharpness in Young's inequality for convolution. *Pacific J. Math.*, 72(2):383–397, 1977.

[GM79] C. C. Graham and O. C. McGehee. *Essays in commutative harmonic analysis*, volume 238 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Science]*. Springer-Verlag, New York, 1979.

[Gre05] B. J. Green. Finite field models in additive combinatorics. In *Surveys in combinatorics 2005*, volume 327 of *London Math. Soc. Lecture Note Ser.*, pages 1–27. Cambridge Univ. Press, Cambridge, 2005.

[GW10] B. J. Green and J. Wolf. A note on Elkin's improvement of Behrend's construction. In *Additive number theory: Festschrift in honor of the sixtieth birthday of Melvyn B. Nathanson*, pages 141–144. Springer-Verlag, 1st edition, 2010.

[Haa81] U. Haagerup. The best constants in the Khintchine inequality. *Studia Math.*, 70(3):231–283 (1982), 1981.

[Hen12] K. Henriot. On arithmetic progressions in $A + B + C$. 2012, arXiv:1211.4917.

[JL93] G. James and M. Liebeck. *Representations and characters of groups*. Cambridge Mathematical Textbooks. Cambridge University Press, Cambridge, 1993.

[JSZ85] W. B. Johnson, G. Schechtman, and J. Zinn. Best constants in moment inequalities for linear combinations of independent and exchangeable random variables. *Ann. Probab.*, 13(1):234–253, 1985.

[Kah60] J.-P. Kahane. Propriétés locales des fonctions à séries de Fourier aléatoires. *Studia Math.*, 19:1–25, 1960.

[Kat04] Y. Katznelson. *An introduction to harmonic analysis*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, third edition, 2004.

[LO94] R. Latała and K. Oleszkiewicz. On the best constant in the Khinchin-Kahane inequality. *Studia Math.*, 109(1):101–104, 1994.

[LR75] J. M. López and K. A. Ross. *Sidon sets*. Marcel Dekker Inc., New York, 1975. Lecture Notes in Pure and Applied Mathematics, Vol. 13.

[Mes95] R. Meshulam. On subsets of finite abelian groups with no 3-term arithmetic progressions. *J. Combin. Theory Ser. A*, 71(1):168–172, 1995.

[Naz97] F. L. Nazarov. The Bang solution of the coefficient problem. *Algebra i Analiz*, 9(2):272–287, 1997.

[NP00] F. L. Nazarov and A. N. Podkorytov. Ball, Haagerup, and distribution functions. In *Complex analysis, operators, and related topics*, volume 113 of *Oper. Theory Adv. Appl.*, pages 247–267. Birkhäuser, Basel, 2000.

[Rot53] K. F. Roth. On certain sets of integers. *J. London Math. Soc.*, 28:104–109, 1953.

[Rud90] W. Rudin. *Fourier analysis on groups*. Wiley Classics Library. John Wiley & Sons Inc., New York, 1990. Reprint of the 1962 original, A Wiley-Interscience Publication.

[Ruz91] I. Z. Ruzsa. Arithmetic progressions in sumsets. *Acta Arith.*, 60(2):191–202, 1991.

[Shi96] A. N. Shiryaev. *Probability*, volume 95 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1996. Translated from the first (1980) Russian edition by R. P. Boas.

[Ten95] G. Tenenbaum. *Introduction to analytic and probabilistic number theory*, volume 46 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1995. Translated from the second French edition (1995) by C. B. Thomas.

[Ter99] A. Terras. *Fourier analysis on finite groups and applications*, volume 43 of *London Mathematical Society Student Texts*. Cambridge University Press, Cambridge, 1999.

[TV06] T. C. Tao and H. V. Vu. *Additive combinatorics*, volume 105 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2006.

Mathematical Institute, University of Oxford, 24-29 St. Giles', Oxford OX1 3LB, England

E-mail address: tom.sanders@maths.ox.ac.uk