ANALYSIS OF BOOLEAN FUNCTIONS

TOM SANDERS

1. Introduction

There are two sources of problems which motivate the ideas in this course: the first is additive number theory, and the second is computer science. In number theory it is the many and varied questions on the representation of integers which we shall want to keep at the back of our minds.

It will be helpful to have a little notation. For sets of integers A and B we write

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

and then if $k \in \mathbb{N}$ we put

$$kA := A + \cdots + A,$$

where the sum is k-fold, and

$$k.A := \{ka : a \in A\}.$$

Note that kA and $k.A$ are very different beasts, for example $2\mathbb{N}$ is the set of natural numbers bigger than 1, whereas $2.\mathbb{N}$ is the set of even natural numbers.

We turn to some examples.

Theorem (Lagrange's theorem)**.** *Every non-negative integer can be written as the sum of four squares, that is* $\mathbb{N}_0 = 4S$ *where* $S := \{0, 1, 4, 9, \dots\}$.

Conjecture (Goldbach's conjecture)**.** *Every even integer bigger than* 2 *can be written as the sum of two primes, that is* $2.\mathbb{N} \setminus \{2\} = 2P$ *where* $P := \{2, 3, 5, 7, 11, \dots\}$.

Theorem (Roth's theorem)**.** *Every subset of the integers of positive relative density contains three distinct elements in arithmetic progression.*

The problems above involve showing the existence of something and often when we try to do this it is helpful to show that there are many of that thing by counting. To this end we introduce the notion of convolution: given sets of integers A and B we write

$$1_A * 1_B(x) := \sum_{z+y=x} 1_A(z)1_B(y)$$

and call it the convolution of 1_A with 1_B. What is important about convolution is that

$$A+B = \operatorname{supp} 1_A * 1_B := \{x : 1_A * 1_B(x) \neq 0\},$$

the support of $1_A * 1_B$.

Last updated: 28th April, 2012.

Focussing on the example of Lagrange's theorem, we should like to show that

$$1_S * 1_S * 1_S * 1_S(x) > 0 \text{ whenever } x \in \mathbb{N}_0.$$

In fact, it is easier to show that it is really quite large if x is large, that is to say one develops an asymptotic for this four-fold convolution. The other two examples also have expressions involving convolutions (and the inner product) and these can be analysed with varying degrees of success.

Convolution requires nothing more than a group structure and, indeed, many of the questions of additive number theory have formulations in general abelian groups. More than this it turns out that there are much better behaved groups which are very good models for $\mathbb{Z}$: the dyadic groups. These groups are the focus of this course.

1.1. **The dyadic groups.** Throughout the course we shall write G for a finite dyadic group, that is a group in which every element has order 2. It is an exercise to check that any such G is isomorphic to (the additive group of) $\mathbb{F}_2^n$ for some n, where $\mathbb{F}_2$ is the field with two elements. In view of this we shall often put $G := \mathbb{F}_2^n$, and tend to use the languages of vector spaces for discussing these groups, so that subgroup are (vector) subspaces and cosets of subgroups are affine subspaces

One of the key ideas of harmonic analysis is to analyse a group G through the space of functions on G and to this end we introduce the Lebesgue spaces.

1.2. **Lebesgue spaces.** Given a finite set X there are two classes of Lebesgue spaces which we shall be interested in corresponding to two natural measures on X. The first is counting measure δ_X defined by

$$\delta_X(A) := |A| \text{ for all } A \subset X;$$

the second is normalised[1] counting measure μ_X defined by

$$\mu_X(A) = \frac{|A|}{|X|} \text{ for all } A \subset X.$$

For obvious reasons we call $\delta_X(A)$ the size of the set A, and $\mu_X(A)$ the density of the set A. Note that for a set $A \subset X$ the measure μ_A can be decomposed as

$$\mu_A = \frac{1}{|A|}\sum_{a\in A}\delta_{\{a\}}$$

and as the measure induced by the map

$$f \mapsto \frac{1}{\mu_X(A)}\int f 1_A d\mu_X, \text{ or equivalently } \mu_A = \frac{1_A}{\mu_X(A)}\mu_X.$$

[1]Normalised here refers to the fact that the integral is 1; it is normalised to have norm 1 with respect to the natural norm on measures $\mu \in M(X)$ defined by $\|\mu\| := \int d|\mu| := \sup\{\int f d\mu : \|f\|_{L^\infty(X)} \leqslant 1\}$. In particular it does *not* refer to the L^2-norm, which, to the extent that it makes sense, has $\|\mu_X\|_{L^2(X)}^2 = |X|^{-1}$.

The Lebesgue spaces of interest are $L^p(X)$, the space of real valued functions on X with norm defined by

$$\|f\|_{L^p(X)} := \left(\int |f|^p d\mu_X\right)^{1/p} = \left(\frac{1}{|X|}\sum_{x\in X}|f(x)|^p\right)^{1/p},$$

and $\ell^p(X)$, the same space of functions with the norm

$$\|f\|_{\ell^p(X)} := \left(\int |f|^p d\delta_X\right)^{1/p} = \left(\sum_{x\in G}|f(x)|^p\right)^{1/p}.$$

The Lebesgue spaces satisfy a useful nesting of norms property:

$$\|f\|_{L^p(X)} \leqslant \|f\|_{L^q(X)} \text{ whenever } p \leqslant q$$

and

$$\|f\|_{\ell^p(X)} \leqslant \|f\|_{\ell^q(X)} \text{ whenever } p \geqslant q.$$

We take the usual convention for $p = \infty$ that

$$\|f\|_{L^\infty(X)} = \|f\|_{\ell^\infty(X)} = \sup_{x\in X}|f(x)| = \max_{x\in X}|f(x)|,$$

and so $\|f\|_{L^p(X)}$ tends to $\|f\|_{L^\infty(X)}$ from below as $p \to \infty$ and $\|f\|_{\ell^p(X)}$ tends to $\|f\|_{\ell^\infty(X)}$ from above in the same limit.

When $p = 2$ the spaces are, of course, Hilbert spaces so that they have an inner product and we write these

$$\langle f, g\rangle_{L^2(X)} := \int fg d\mu_X = \frac{1}{|X|}\sum_{x\in X} f(x)g(x) \text{ for all } f, g \in L^2(X),$$

and

$$\langle f, g\rangle_{\ell^2(X)} := \int fg d\delta_X = \sum_{x\in X} f(x)g(x) \text{ for all } f, g \in \ell^2(X).$$

Finally we have Hölder's inequality that

$$\langle f, g\rangle_{L^2(X)} \leqslant \|f\|_{L^p(X)}\|g\|_{L^q(X)} \text{ for all } f \in L^p(X), g \in L^q(X)$$

and

$$\langle f, g\rangle_{\ell^2(X)} \leqslant \|f\|_{\ell^p(X)}\|g\|_{\ell^q(X)} \text{ for all } f \in \ell^p(X), g \in \ell^q(X)$$

whenever $p^{-1} + q^{-1} = 1$. It is easy to check that these inequalities are sharp by considering δ-functions.

The pair (p, q) is called a pair of conjugate indices, and the case $p = q = 2$ is the Cauchy-Schwarz inequality.

1.3. **Convolution.** Suppose that $G := \mathbb{F}_2^n$ and $f, g \in L^1(G)$. Then we define the convolution of f and g with a slightly different normalisation to before:

$$f * g(x) := \int f(x-y)g(y)d\mu_G(y) = \frac{1}{|G|}\sum_{z+y=x} f(z)g(y) \text{ for all } x \in G,$$

and so if μ and ν are measures on G then

$$\mu * \nu(E) = \int 1_E(x+y)d\mu(x)d\nu(y) \text{ for all } E \subset G.$$

Since our groups are always finite we often make the abuse of writing $\mu(x)$ for $\mu(\{x\})$.

We shall frequently find ourselves changing the order of integration (really summation), and here we get that

$$\int f * g d\mu_G = \int f d\mu_G \int g d\mu_G.$$

It turns out that with absolute value signs this becomes a special case of Young's inequality. In general by Young's inequality we shall mean the statement

$$\|f * g\|_{L^r(G)} \leqslant \|f\|_{L^p(G)}\|g\|_{L^q(G)} \text{ for all } f \in L^p(G) \text{ and } g \in L^q(G)$$

for a triple p, q, r provided $1 + r^{-1} = p^{-1} + q^{-1}$. Of particular interest are the cases $p, q = 2$ and $r = \infty$ which encodes the idea that the convolution of two functions in L^2 is 'continuous', and $p, q, r = 1$ which tells us that L^1 is 'closed under convolution'.

As a check of understanding it may be helpful to note that

$$f * f(0_G) = \int f(-x)f(x)d\mu_G(x) = \|f\|^2_{L^2(G)}$$

since $-x = x$ in G. It follows that Young's inequality certainly can't be any better than this for $r = \infty$ and $p, q = 2$ and, in fact, it is relatively easy to see that it is tight by considering δ-functions.

As before the crucial identity for us is that

$$A + B = \operatorname{supp} 1_A * 1_B.$$

Indeed, it is far easier to analyse the function $1_A * 1_B$ than 1_{A+B} because the former is far smoother: it is literally an average of 1_A over translates of B. Indeed, there is a useful maxim here that the more times you convolve the smoother a function becomes.

It will be instructive to bear some examples in mind.

Example (Convolution of subspaces)**.** Suppose that $W = x + V$ is an affine subspace with vector subspace V. We see immediately that $\operatorname{supp} 1_W * 1_W = W + W = V$ (since we are in characteristic 2) and if $y \in V$ then

$$1_W * 1_W(y) = \int 1_W(y-z)1_W(z)d\mu_G(z) = \int 1_W(-z)1_W(z)d\mu_G(z) = \mu_G(W),$$

so

$$1_W * 1_W = \mu_G(W)1_V.$$

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

$$\mathbb{E}1_A * 1_A(y) = \int \mathbb{E}1_A(y-x)1_A(x)d\mu_G(x) = \begin{cases} \alpha^2 & \text{if } y \neq 0_G \\ \alpha & \text{otherwise,} \end{cases}$$

so we expect it to be very likely that $A + A$ is essentially the whole of G provided α is not too small. In particular,

$$\mathrm{Var}(1_A * 1_A(x)) = \alpha^2(1-\alpha^2)/|G| \text{ if } x \neq 0_G,$$

and so by the central limit theorem we expect

$$\begin{aligned} \mathbb{P}(1_A * 1_A(x) - \alpha^2 < -\alpha^2/2) &\sim \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-\sqrt{\frac{\alpha^2|G|}{4(1-\alpha^2)}}} \exp(-u^2/2)du \\ &= O_\alpha(\sqrt{|G|}\exp(-|G|\alpha^2/8(1-\alpha^2))) \end{aligned}$$

for each $x \neq 0_G$. If $x \notin A+A$ then $1_A*1_A(x) = 0$ and so (certainly) $1_A*1_A(x)-\alpha^2 < -\alpha^2/2$. Thus the expected number of such x is $O_\alpha(|G|^{3/2}\exp(-\Omega_\alpha(|G|)))$. This is much less than 1 if $|G|$ is large (and α is not too small) which leads to the conclusion.

Example (Convolution as a sum of random variables)**.** Suppose that $A \subset G$ and X and Y are two independent A-valued uniform random variables, so that

$$\mathbb{P}(X = x) = \begin{cases} \frac{1}{|A|} & \text{if } x \in A \\ 0 & \text{otherwise,} \end{cases}$$

i.e. $\mathbb{P}(X = x) = \mu_A(x)$ and similarly for Y. Then $Z = X + Y$ has

$$\mathbb{P}(Z = a) = \mu_A * \mu_A(x);$$

the law of the sum of two independent random variables is the convolution of their laws. In general if we sample uniformly and independently at random k times from the set A then the probability that the sum of the samples is x is $\mu_A * \cdots * \mu_A(x)$ where the convolution is k-fold.

Example (Convolution as a measure of relative density)**.** Suppose that $X, A \subset G$. Then $1_A * \mu_X(x)$ is the relative density of A on the set $x + X$, that is the number of points in $A \cap (x + X)$ divided by the number of points in $x + X$. To see this note that

$$\begin{aligned} 1_A * \mu_X(x) &= \int 1_A(y)d\mu_X(x+y) \\ &= \int 1_A(y)\frac{1_X(x+y)}{\mu_G(X)}d\mu_G(x+y) \\ &= \frac{1}{\mu_G(X)}\int 1_A(y)1_{x+X}(y)d\mu_G(x+y). \end{aligned}$$

But then by translation invariance of μ_G we have

$$\begin{aligned}\frac{1}{\mu_G(X)}\int 1_A(y)1_{x+X}(y)d\mu_G(x+y) &= \frac{1}{\mu_G(X)}\int 1_A(y)1_{x+X}(y)d\mu_G(y)\\ &= \frac{\mu_G(A\cap(x+X))}{\mu_G(X)} = \frac{|A\cap(x+X)|}{|x+X|}.\end{aligned}$$

We shall frequently use this in the case when $X = V$ for some subspace $V \leqslant G$. In this case $1_A * \mu_V(x)$ is the relative density of A on the coset $x + V$.

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 case of highly structured sets such as subspaces, and random-like sets (such as random sets!) will form a dichotomy which will pervade our work; to quantify the notion of being random-like we introduce the Fourier transform.

1.4. **The Fourier transform.** For $G := \mathbb{F}_2^n$ we write $\widehat{G}$ for the collection of characters on G, that is maps of the form

$$x \mapsto (-1)^{r.x} \text{ where } r.x := r_1x_1 + \cdots + r_nx_n$$

and $r \in \mathbb{F}_2^n$. Characters can be added via the slightly confusing formula

$$(\gamma + \gamma')(x) := \gamma(x)\gamma'(x) \text{ for all } x \in G$$

and form a group which is (non-canonically) isomorphic to G. They are easily seen to be homomorphisms from G to $\{-1, 1\}$ under multiplication and remarkably they turn out to be an orthonormal basis of $L^2(G)$. To see that they are pair-wise orthogonal we note that

$$\begin{aligned}\langle\gamma,\gamma'\rangle_{L^2(G)} &= \int \gamma(x)\gamma'(x)d\mu_G(x)\\ &= \gamma(y)\gamma'(y)\int \gamma(x)\gamma'(x)d\mu_G(x) = \gamma(y)\gamma'(y)\langle\gamma,\gamma'\rangle_{L^2(G)}.\end{aligned}$$

Hence, either $\gamma(y)\gamma'(y) = 1$ for all $y \in G$, whereupon $\gamma = \gamma'$ and we have that $\|\gamma\|_{L^2(G)} = 1$; or $\gamma(y)\gamma'(y) = -1$ for some y and we conclude that the inner product is 0. We write this formally as

$$\langle\gamma,\gamma'\rangle_{L^2(G)} = \begin{cases}1 & \text{if } \gamma = \gamma'\\ 0 & \text{otherwise.}\end{cases}$$

The characters then form a basis of $L^2(G)$ because they are orthogonal (and so linearly independent) and there are $|G|$ of them which is the dimension of $L^2(G)$. Since they form an orthonormal basis we define the Fourier transform to be the map taking $f \in L^1(G)$ to $\widehat{f} \in \ell^\infty(\widehat{G})$ determined by

$$\widehat{f}(\gamma) := \langle f,\gamma\rangle_{L^2(G)} = \int f(x)\gamma(x)d\mu_G(x),$$

and so that it is completely clear if μ is a measure on G then

$$\widehat{\mu}(\gamma) := \int \gamma(x) d\mu(x).$$

It is easy to see by the triangle inequality that we have the Hausdorff-Young inequality:

$$\|\widehat{f}\|_{\ell^\infty(\widehat{G})} \leqslant \|f\|_{L^1(G)} \text{ for all } f \in L^1(G).$$

Since $\widehat{G}$ is an orthonormal basis we have the Fourier inversion formula:

$$f(x) = \sum_{\gamma \in \widehat{G}} \widehat{f}(\gamma)\gamma(x) \text{ for all } x \in G.$$

More than this the change of basis is unitary and so we have Plancherel's theorem:

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

and the special case when $f = g$, called Parseval's theorem:

$$\|f\|_{L^2(G)} = \|\widehat{f}\|_{\ell^2(\widehat{G})} \text{ for all } f \in L^2(G).$$

The Fourier transform is so useful because it is an (essentially unique) change of basis which simultaneously diagonalises all convolution operators. Specifically, given $f \in L^1(G)$, we get a linear operator

$$L^2(G) \to L^2(G); g \mapsto f * g.$$

This operator is diagonalised by the Fourier basis:

$$f * \gamma(y) = \int f(y - x)\gamma(x) d\mu_G(x) = \gamma(y) \int f(z)\gamma(z) d\mu_G(z) = \widehat{f}(\gamma)\gamma(y)$$

by the change of variables $z = y - x$. Thus,

$$f * g = \sum_{\gamma \in \widehat{G}} \widehat{f}(\gamma)\widehat{g}(\gamma)\gamma,$$

so that $\widehat{f * g} = \widehat{f} \cdot \widehat{g}$.

Example (Annihilators and the Fourier transform of subspaces)**.** Given $A \subset G$ we write $A^\perp$ for the annihilator of A, that is the set $\{\gamma \in \widehat{G} : \gamma(x) = 1 \text{ for all } x \in A\}$, and similarly if $\Gamma \subset \widehat{G}$ we write $\Gamma^\perp$ for the annihilator of Γ, the set $\{x \in G : \gamma(x) = 1 \text{ for all } \gamma \in \Gamma\}$.

It is immediate that annihilators are subspaces and that if $V \leqslant G$ then $V \subset (V^\perp)^\perp$; in fact we have equality as we shall see shortly.

From our calculation on the convolution of indicator functions of subspaces we see that $\widehat{1_V}^2 = \mu_G(V)\widehat{1_V}$, whence $\widehat{1_V}(\gamma)$ takes only the values 0 and $\mu_G(V)$. On the other hand, if $\widehat{1_V}(\gamma) = \mu_G(V)$ then $\gamma \in V^\perp$ and conversely so we have

$$\widehat{1_V}(\gamma) = \begin{cases} \mu_G(V) & \text{if } \gamma \in V^\perp \\ 0 & \text{otherwise.} \end{cases}$$

It follows by Parseval's theorem that

$$\mu_G(V)|V^\perp| = \frac{1}{\mu_G(V)}\sum_{\gamma\in V^\perp}|\widehat{1_V}(\gamma)|^2 = \frac{1}{\mu_G(V)}.\mu_G(V) = 1.$$

It follows from this that $\mu_G((V^\perp)^\perp) = \mu_G(V)$ and hence that $V = (V^\perp)^\perp$. Finally, the co-dimension of V is the dimension of G/V, that is $n - \dim V$ and so from the previous

$$\dim V^\perp = \operatorname{cod} V.$$

Example (An uncertainty principle)**.** By Hölder's inequality, the Hausdorff-Young inequality and Parseval's theorem we have that any function with unit L^2-norm has

$$\|f\|_{L^1(G)}\|\widehat{f}\|_{\ell^1(\widehat{G})} \geqslant \|\widehat{f}\|_{\ell^\infty(\widehat{G})}\|\widehat{f}\|_{\ell^1(\widehat{G})} \geqslant \|\widehat{f}\|^2_{\ell^2(\widehat{G})} = \|f\|^2_{L^2(G)} = 1.$$

It follows that a function cannot both have concentrated support on G (physical space) and $\widehat{G}$ (momentum space). In particular, by Cauchy-Schwarz we have

$$\|f\|_{L^1(G)} \leqslant \mu_G(\operatorname{supp} f)^{1/2}\|f\|_{L^2(G)} = \mu_G(\operatorname{supp} f)^{1/2}$$

and similarly $\|\widehat{f}\|_{\ell^1(\widehat{G})} \leqslant |\operatorname{supp}\widehat{f}|^{1/2}$. Thus,

$$\mu_G(\operatorname{supp} f)|\operatorname{supp}\widehat{f}| \geqslant 1$$

and we see from the preceding example that equality can be achieved when f is a scalar multiple of an affine subspace.

2. Subspaces, sumsets and counting solutions to equations

Our first result along the theme of the course shows that by adding a set to itself a few times we can ensure that the resulting set contains a large algebraic structure.

To get a grip on the problem notice that if A is a subspace of density α then $4A$ is also a subspace of density α (and so co-dimension $\log_2\alpha^{-1}$); whereas if A is a random set of density about α then in all likelihood $4A$ is the whole of G (that is a subspace of co-dimension 0).

Theorem 2.1 (Bogolyubov's lemma)**.** *Suppose that $G := \mathbb{F}_2^n$ and $A, B \subset G$ have density $\alpha, \beta > 0$. Then $2(A-B)$ contains a subspace of co-dimension $O(\alpha^{-1}\beta^{-1})$.*

Proof. We shall examine the convolution convolution of $1_A * 1_B$ with itself:

$$1_A * 1_B * 1_A * 1_B(x) = \sum_{\gamma\in\widehat{G}}|\widehat{1_A}(\gamma)|^2|\widehat{1_B}(\gamma)|^2\gamma(x).$$

We separate into those characters supporting large and small values of the Fourier transform:

$$L := \{\gamma\in\widehat{G} : |\widehat{1_B}(\gamma)| \geqslant \epsilon\beta\}.$$

Then Parseval's theorem gives an upper bound on L:

$$|L|(\epsilon\beta)^2 \leqslant \sum_{\gamma\in L}|\widehat{1_B}(\gamma)|^2 \leqslant \|\widehat{1_B}\|^2_{\ell^2(\widehat{G})} = \|1_B\|^2_{L^2(G)} = \beta,$$

so $|L| \leqslant \epsilon^{-2}\beta^{-1}$. We put $V := L^{\perp}$ and note that the bound on L implies that the co-dimension of V is at most $\epsilon^{-2}\beta^{-1}$. On the other hand if $x \in V$ then $\gamma(x) = 1$ for all $\gamma \in L$ and so we have that

$$\sum_{\gamma\in L} |\widehat{1_A}(\gamma)|^2|\widehat{1_B}(\gamma)|^2\gamma(x) \geqslant |\widehat{1_A}(0_{\widehat{G}})|^2|\widehat{1_B}(0_{\widehat{G}})|^2 = \alpha^2\beta^2$$

since $\widehat{1_A}(0_{\widehat{G}}) = \alpha$, and $\widehat{1_B}(0_{\widehat{G}}) = \beta$ and hence $0_{\widehat{G}} \in L$. On the other hand by the triangle inequality and Parseval's theorem we have

$$\begin{aligned}\left|\sum_{\gamma\notin L} |\widehat{1_A}(\gamma)|^2|\widehat{1_B}(\gamma)|^2\gamma(x)\right| &\leqslant \sup_{\gamma\notin L}|\widehat{1_B}(\gamma)|^2\sum_{\gamma\in\widehat{G}}|\widehat{1_A}(\gamma)|^2\\ &\leqslant (\epsilon\beta)^2\|1_A\|^2_{L^2(G)} = \epsilon^2\beta^2\alpha.\end{aligned}$$

Thus, if $\epsilon := \sqrt{\alpha/2}$ then we see that for all $x \in V^{\perp}$ we have $1_A * 1_B * 1_A * 1_B(x) \geqslant \alpha^2\beta^2/2$. The result is proved. □

This lemma is typical of the sort of results in the course. It inspires the definition of the spectrum of a function: suppose that $f \in L^2(G)$ and $\epsilon \in (0, 1]$. Then we write

$$\mathrm{Spec}_{\epsilon}(f) := \{\gamma \in \widehat{G} : |\widehat{f}(\gamma)| \geqslant \epsilon\|f\|_{L^1(G)}\},$$

for the ϵ-spectrum of f. Note that if f is non-negative then $0_{\widehat{G}}$ is in $\mathrm{Spec}_{\epsilon}(f)$ for all $\epsilon \in (0, 1]$, and it makes no sense to consider $\epsilon > 1$ since the set is invariable empty by the Hausdorff-Young inequality.

An essential ingredient of Bogolyubov's lemma was the so called Parseval bound on the size of the spectrum which lets us project out large Fourier coefficients. In particular we implicitly proved the following lemma.

Lemma 2.2. *Suppose that $B \subset G$ has density β and $\epsilon \in (0, 1]$. Then* $\operatorname{cod}\mathrm{Spec}_{\epsilon}(1_B)^{\perp} = O(\epsilon^{-2}\beta^{-1})$ *and*

$$\sup_{\gamma\notin(\mathrm{Spec}_{\epsilon}(1_B)^{\perp})^{\perp}} |\widehat{1_B}(\gamma)| \leqslant \epsilon\beta.$$

This may be seen as giving us a low complexity approximation to 1_B. In particular, we get a subspace V of controlled co-dimension such that $1_B \approx 1_B * \mu_V$ in a certain norm. It is not, however, clear why this norm should be useful.

Suppose that we have a set $A \subset G$ and wish to count the number of sums in A, that is triples $(x, y, z) \in A^3$ such that $x + y = z$. We write the density of such as

$$T(A) := \langle 1_A * 1_A, 1_A\rangle_{L^2(G)} = \sum_{\gamma\in\widehat{G}} \widehat{1_A}(\gamma)^3.$$

As usual $\widehat{1_A}(0_{\widehat{G}}) = \alpha$, and if $\sup_{\gamma\neq 0_{\widehat{G}}} |\widehat{1_A}(\gamma)| \leqslant \epsilon\alpha$ then

$$T(A) \geqslant \alpha^3 - \epsilon\alpha\sum_{\gamma\in\widehat{G}} |\widehat{1_A}(\gamma)|^2.$$

It follows that $T(A) \geqslant \alpha^3/2$ if $\epsilon \leqslant \alpha/2$. We think of α^3 as being the 'expected' number of solutions to $x+y=z$ in a random set of density α; $\sup_{\gamma \neq 0_{\widehat{G}}} |\widehat{1_A}(\gamma)|$ measures how far from being random we are.

At the other end of the spectrum from appearing random we have subspaces. Of course, if A is a vector subspace then $T(A) = \alpha^2$. However, if A is an affine subspace that is not a vector subspace then $T(A) = 0$ and so we see that $T(A)$ is not always large. Despite this we shall prove the following theorem.

Theorem 2.3 (Arithmetic removal lemma)**.** *Suppose that $A \subset G$, and that if $A' \subset A$ has $T(A') = 0$ then $\mu_G(A \setminus A') \geqslant \epsilon$. Then $T(A) = \Omega_\epsilon(1)$.*

This result and the approach we take to it was first developed by Green in [Gre05]. It will be useful to begin by introducing a more general tri-linear form based on T: for functions f_0, f_1, f_2 on G we put

$$T(f_0, f_1, f_2) := \langle f_0 * f_1, f_2 \rangle_{L^2(G)},$$

so that $T(A) = T(1_A, 1_A, 1_A)$. Importantly we have the following lemma for governing the behaviour of T which captures the content of our earlier argument for sets behaving 'randomly'.

Lemma 2.4. *Suppose that f_0, f_1, f_2 are functions on G. Then*

$$|T(f_0, f_1, f_2)| \leqslant \|\widehat{f_i}\|_{\ell^\infty(\widehat{G})} \|f_j\|_{L^2(G)} \|f_k\|_{L^2(G)} \leqslant \|f_i\|_{L^1(G)} \|f_j\|_{L^2(G)} \|f_k\|_{L^2(G)}.$$

for any permutation $\{i,j,k\}$ of $\{0,1,2\}$.

Proof. Notice that the second inequality is a consequence of the first and the Hausdorff-Young inequality. Otherwise, by Fourier inversion we have

$$T(f_0, f_1, f_2) = \sum_{\gamma \in \widehat{G}} \widehat{f_0}(\gamma)\widehat{f_1}(\gamma)\widehat{f_2}(\gamma) = \sum_{\gamma \in \widehat{G}} \widehat{f_i}(\gamma)\widehat{f_j}(\gamma)\widehat{f_k}(\gamma)$$

for any permutation $\{i,j,k\}$ of $\{0,1,2\}$. We apply Hölder's inequality and Cauchy-Schwarz to this to see that

$$|T(f_0, f_1, f_2)| \leqslant \sup_{\gamma \in \widehat{G}} |\widehat{f_i}(\gamma)| \left(\sum_{\gamma \in \widehat{G}} |\widehat{f_j}(\gamma)|^2 \right)^{1/2} \left(\sum_{\gamma \in \widehat{G}} |\widehat{f_k}(\gamma)|^2 \right)^{1/2}.$$

The result now follows by Parseval's theorem. □

For this to be useful we require a scale on which we can control the uniformity of a function and this is provided to us by the following arithmetic regularity lemma.

Proposition 2.5 (Arithmetic regularity lemma)**.** *Suppose that $A \subset G$ has density α, $B \subset G$ and $\delta, \eta \in (0,1]$. Then there are subspaces $V' \leqslant V \leqslant G$ with $\operatorname{cod} V' = O_{\delta,\eta}(1)$ such that*

$$\|1_A * \mu_V - 1_A * \mu_{V'}\|_{L^2(G)}^2 \leqslant \delta\alpha \text{ and } \sup_{\gamma \notin V'^\perp} |\widehat{1_B}(\gamma)| \leqslant \eta \mu_G(V).$$

Proof. We define a sequence of subspaces iteratively letting $V_0 = G$, and assuming we have defined V_i we let U be the subspace of co-dimension $O(\eta^{-2}\mu_G(V_i)^{-2})$ provided by Lemma 2.2 (with parameter $\min\{1, \eta\mu_G(V_i)\beta^{-1}\}$), and $V_{i+1} := V_i \cap U$ so that

$$\sup_{\gamma \notin V_{i+1}^\perp} |\widehat{1_B}(\gamma)| \leqslant \sup_{\gamma \notin U^\perp} |\widehat{1_B}(\gamma)| \leqslant \eta\mu_G(V_i).$$

By Parseval's theorem and the fact that convolution goes to multiplication we have

$$\|1_A * \mu_{V_i} - 1_A * \mu_{V_{i+1}}\|_{L^2(G)}^2 = \sum_{\gamma \in V_{i+1}^\perp \setminus V_i^\perp} |\widehat{1_A}(\gamma)|^2.$$

However, the sets $(V_{i+1}^\perp \setminus V_i^\perp)_i$ are disjoint so it follows by averaging that there is some $i = O(\eta^{-1})$ such that

$$\|1_A * \mu_{V_i} - 1_A * \mu_{V_{i+1}}\|_{L^2(G)}^2 \leqslant \eta \sum_{\gamma \in \widehat{G}} |\widehat{1_A}(\gamma)|^2 = \eta\alpha.$$

The result follows on setting $V := V_i$ and $V' := V_{i+1}$. □

Notice that the proof leads to a very poor bound on $\operatorname{cod} V'$. Indeed, it is a tower of $\eta^{-O(1)}$s of height δ^{-1}.

We can now prove the arithmetic removal lemma.

Proof of Theorem 2.3. We apply the regularity lemma with the sets A and $B := A$ to get subspaces $V' \leqslant V \leqslant G$ with $\operatorname{cod} V' = O_{\delta,\eta}(1)$ and

$$\|1_A * \mu_V - 1_A * \mu_{V'}\|_{L^2(G)}^2 \leqslant \delta\alpha \text{ and } \sup_{\gamma \notin V'^\perp} |\widehat{1_A}(\gamma)| \leqslant \eta\mu_G(V).$$

Let

$$A_1 := \{x \in A : 1_A * \mu_V(x) \leqslant \epsilon/4\}$$

and

$$A_2 := \{x \in A : |1_A * \mu_{V'} - 1_A * \mu_V(x)|^2 * \mu_V(x) \geqslant 4\delta\epsilon^{-1}\}.$$

We want to get upper bounds for the sizes of the sets A_1 and A_2. For A_1, first note that

$$A_1' := \{x \in G : 1_A * \mu_V(x) \leqslant \epsilon/4\}$$

is invariant under translation by elements of V, so $1_{A_1'} * \mu_V = 1_{A_1'}$. Now

$$\begin{aligned}
\mu_G(A_1) = \int 1_{A_1'} 1_A d\mu_G &= \int 1_{A_1'} * \mu_V 1_A d\mu_G \\
&= \int\int 1_{A_1'}(y) \frac{1_V(x-y)}{\mu_G(V)} d\mu_G(y) 1_A(x) d\mu_G(x) \\
&= \int\int 1_A(x) \frac{1_V(y-x)}{\mu_G(V)} d\mu_G(x) 1_{A_1'}(y) d\mu_G(y) \\
&= \int 1_A * \mu_V 1_{A_1'} d\mu_G \leqslant \epsilon/4
\end{aligned}$$

where the last inequality is since $1_A * \mu_V(x) \leqslant \epsilon/4$ on A_1'.

For A_2 note that

$$\begin{aligned}\mu_G(A_2)4\delta\epsilon^{-1} &\leqslant \int |1_A*\mu_{V'}-1_A*\mu_V|^2*\mu_V(x)d\mu_G(x)\\ &= \int\int |1_A*\mu_{V'}-1_A*\mu_V|^2(y)\frac{1_V(x-y)}{\mu_G(V)}d\mu_G(y)d\mu_G(x)\\ &= \int |1_A*\mu_{V'}-1_A*\mu_V|^2(y)\int \frac{1_V(x-y)}{\mu_G(V)}d\mu_G(x)d\mu_G(y)\\ &= \int |1_A*\mu_{V'}-1_A*\mu_V|^2(y)d\mu_G(y)\\ &= \|1_A*\mu_{V'}-1_A*\mu_V\|_{L^2(G)}^2\leqslant\delta\alpha,\end{aligned}$$

whence $\mu_G(A_2)\leqslant\epsilon/4$. It follows that $A':=A\setminus(A_1\cup A_2)$ has $T(A')\neq 0$, and so there is a triple (x_0,x_1,x_2) with $x_0+x_1=x_2$ and

$$1_A*\mu_V(x_i)\geqslant\epsilon/4 \text{ and } |1_A*\mu'_V-1_A*\mu_V(x_i)|^2*\mu_V(x_i)\leqslant 4\delta\epsilon^{-1}$$

for all $i\in\{0,1,2\}$. We put $S_i:=A\cap(x_i+V)$, so that

$$\mu_G(S_i)=1_A*\mu_V(x_i)\mu_G(V),$$

and put $f_i:=(1_A-1_A*\mu_V)|_{x_i+V}$ and $g_i:=(1_A-1_A*\mu_{V'})|_{x_i+V}$. By Cauchy-Schwarz we have

$$\|f_i-g_i\|_{L^1(G)}^2\leqslant\mu_G(V)\|f_i-g_i\|_{L^2(G)}^2 \tag{2.1}$$

since $\mathrm{supp}(f_i-g_i)\subset x_i+V$. However,

$$\begin{aligned}\|f_i-g_i\|_{L^2(G)}^2 &= \int |1_A*\mu_{V'}(y)-1_A*\mu_V(y)|^2 1_{x_i+V}(y)d\mu_G(y)\\ &= \mu_G(V)\int |1_A*\mu_{V'}(y)-1_A*\mu_V(y)|^2\frac{1_V(x_i-y)}{\mu_G(V)}d\mu_G(y).\end{aligned}$$

Now $1_A*\mu_V(y)=1_A*\mu_V(x_i)$ if $y\in x_i+V$, whence

$$\begin{aligned}\|f_i-g_i\|_{L^2(G)}^2 &= \mu_G(V)\int |1_A*\mu_{V'}(y)-1_A*\mu_V(x_i)|^2\frac{1_V(x_i-y)}{\mu_G(V)}d\mu_G(y)\\ &= \mu_G(V)|1_A*\mu_{V'}-1_A*\mu_V(x_i)|^2*\mu_V(x_i).\end{aligned}$$

Combining this with (2.1) we get that

$$\|f_i-g_i\|_{L^1(G)}^2\leqslant\mu_G(V)^2|1_A*\mu_{V'}-1_A*\mu_V(x_i)|^2*\mu_V(x_i)\leqslant 4\delta\epsilon^{-1}\mu_G(V)^2.$$

Finally,

$$\begin{aligned}\|\widehat{g_i}\|_{\ell^\infty(\widehat{G})} &= \sup_{\gamma\in\widehat{G}}\left|\sum_{\gamma'\notin V'^\perp}\widehat{1_A}(\gamma')\int\gamma'(x)1_{x_i+V}(x)\gamma(x)d\mu_G(x)\right| \\ &= \sup_{\gamma\in\widehat{G}}\left|\sum_{\gamma'\notin V'^\perp}\widehat{1_A}(\gamma')\gamma(x_i)\gamma'(x_i)\mu_G(x_i+V)1_{V^\perp}(\gamma+\gamma')\right| \leqslant \eta\mu_G(V)\end{aligned}$$

by Fourier inversion.

Now we examine

$$T(1_{S_0},1_{S_1},1_{S_2}) = \mu_G(V)^2\prod_{i=0}^{2}1_A*\mu_V(x_i) + 1_A*\mu_V(x_1)T(f_0,1_V,1_{S_2}) + T(1_{S_0},f_1,1_{S_2}).$$

The first of these terms is at least $(\epsilon/4)^3\mu_G(V)^2$; the second has

$$\begin{aligned}|T(f_0,1_V,1_{S_2})| &\leqslant |T(f_0-g_0,1_V,1_{S_2})| + |T(g_0,1_V,1_{S_2})| \\ &\leqslant (\|f_0-g_0\|_{L^1(G)} + \eta\mu_G(V))\|1_V\|_{L^2(G)}\|1_{S_2}\|_{L^2(G)} \\ &\leqslant (2\delta^{1/2}\epsilon^{-1/2}+\eta)\mu_G(V)^2\end{aligned}$$

by Lemma 2.4; and similarly for the third. We conclude that

$$T(A) \geqslant ((\epsilon/4)^3 - 4\delta^{1/2}\epsilon^{-1/2} - 2\eta)\mu_G(V)^2$$

and the result follows on suitable choice of δ and η. □

The bound resulting from this proof shows that $T(A)$ is at least the reciprocal of a tower of 2s of height $\epsilon^{-O(1)}$. It is possible (see the work of Fox [Fox11]) to make this a tower of height $O(\log\epsilon^{-1})$, but nothing better is known.

3. Sums of independent random variables

The setting $G := \mathbb{F}_2^n$ affords a remarkable unification of the algebraic and statistical notions of independence. Thinking of $\widehat{G}$ as another vector space over $\mathbb{F}_2$, a set $\Lambda\subset\widehat{G}$ is algebraically independent if

$$\sum_{\gamma\in\Lambda}\sigma_\gamma.\gamma = 0_{\widehat{G}} \text{ and } \sigma:\Lambda\to\mathbb{F}_2 \text{ iff } \sigma\equiv 0.$$

However, the elements of $\widehat{G}$ can also be thought of as random variables on G with underlying probability measure μ_G. A set $\Lambda\subset\widehat{G}$ is statistically independent if

$$\mu_G(\{x:\gamma(x)=z_\gamma \text{ for all } \gamma\in\Lambda'\}) = \prod_{\gamma\in\Lambda'}\mu_G(\{x:\gamma(x)=z_\gamma\})$$

for all $\Lambda'\subset\Lambda$ and $z:\Lambda'\to\{-1,1\}$.

Theorem 3.1. *Suppose that $\Lambda\subset\widehat{G}$. Then Λ is algebraically independent iff it is statistically independent.*

Proof. Since the characters γ are homomorphisms we see that any $y, z \in \{x : \gamma(x) = z_\gamma \text{ for all } \gamma \in \Lambda'\}$ 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

$$\mu_G(\Lambda'^{\perp}) = \prod_{\gamma\in\Lambda'} \mu_G(\{\gamma\}^{\perp}) \text{ for all } \Lambda' \subset \Lambda.$$

Now, if Λ is algebraically independent then none of the γs in Λ are identically 1 and so $\mu_G(\{\gamma\}^{\perp}) = 1/2$ 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 $\mu_G(\Lambda'^{\perp}) = \mu_G(((\Lambda'^{\perp})^{\perp})^{\perp}) = 2^{-d}$ 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. □

In view of the above theorem we shall treat sums of algebraically independent characters as sums of independent random variables, and hence the central limit theorem will provide a useful heuristic. This asserts that if Λ is an independent set of characters then

$$\mu_G(x : \sum_{\lambda\in\Lambda} \lambda(x) \leqslant \eta\sqrt{n}) \sim \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\eta} \exp(-x^2/2)dx$$

as $n \to \infty$. We should like to use this to estimate the probability that the sum when η is a large negative and show that the probability is very small. However, the rate of convergence in the central limit theorem is not very rapid, and necessarily so when η is close to zero. Since this is not our range of interest we shall formulate some other, rather simpler tools, for dealing with this range.

If one is interested in estimating the probability that a function (on G) takes large values then one naturally turns to the higher moments of that function. To this end we have the following simple inequality which may also be found in [Rud90].

Proposition 3.2 (Rudin's inequality)**.** *Suppose that $\Lambda \subset \widehat{G}$ is independent and $p \in [2, \infty)$. Then*

$$\|\sum_{\gamma\in\Lambda} f(\gamma)\gamma\|_{L^p(G)} = O(\sqrt{p}\|f\|_{\ell^2(\Lambda)}) \text{ for all } f \in \ell^2(\Lambda).$$

Proof. By nesting of norms the result follows if we can show it for p an even integer, say $2k$. In this case we can multiply out the left hand side:

$$\begin{aligned}\|\sum_{\gamma\in\Lambda} f(\gamma)\gamma\|^{2k}_{L^{2k}(G)} &= \sum_{\gamma_1,\dots,\gamma_{2k}\in\Lambda} \prod_{i=1}^{2k} f(\gamma_i) \int \prod_{i=1}^{2k} \gamma_i d\mu_G \\ &= \sum_{\substack{\gamma_1,\dots,\gamma_{2k}\in\Lambda \\ \gamma_1+\dots+\gamma_{2k}=0_{\widehat{G}}}} \prod_{i=1}^{2k} f(\gamma_i).\end{aligned}$$

Since Λ is independent it follows that for each summand there is a set $I \subset \{1,\dots,2k\}$ of size k and bijection $\phi : I \to \{1,\dots,2k\} \setminus I$ such that $\gamma_i = \gamma_{\phi(i)}$ for all $i \in I$. In this case the summand is just $\prod_{i\in I} |f(\gamma_i)|^2$, and so

$$\begin{aligned}\sum_{\substack{\gamma_1,\dots,\gamma_{2k}\in\Lambda\\ \gamma_1+\dots+\gamma_{2k}=0_{\widehat{G}}}} \prod_{i=1}^{2k} f(\gamma_i) &\leqslant \sum_{I,\phi} \sum_{\gamma_i\in\Lambda \text{ for all } i\in I} \prod_{i\in I} |f(\gamma_i)|^2 \\ &= \sum_{I,\phi} \left(\sum_{\gamma\in\Lambda} |\widehat{f}(\gamma)|^2\right)^k \leqslant \binom{2k}{k} k! \|f\|_{\ell^2(\Lambda)}^{2k}.\end{aligned}$$

The result follows. □

Although we call this Rudin's inequality, it is properly called Kintchine's inequality as established by Kintchine and Littlewood in the early 1920s. We use the name Rudin's inequality because in the more general case of arbitrary finite abelian groups that is the result one uses and it is the result to which modern literature points.

We shall find the dual formulation of this particularly useful.

Proposition 3.3. *Suppose that $\Lambda \subset \widehat{G}$ is independent and $p \in (1,2]$. Then*

$$\|\widehat{f}\|_{\ell^2(\Lambda)} = O\left(\sqrt{\frac{p}{p-1}}\|f\|_{L^p(G)}\right) \text{ for all } f \in L^p(G).$$

Proof. The argument is by duality. Suppose that $f \in L^p(G)$ and put

$$g := \sum_{\lambda\in\Lambda} \widehat{f}(\lambda)\lambda.$$

Then by Plancherel's theorem and Hölder's inequality we have

$$\|\widehat{f}\|_{\ell^2(\Lambda)}^2 = \langle \widehat{f}, \widehat{f}\rangle_{\ell^2(\Lambda)} = \langle \widehat{f}, \widehat{g}\rangle_{\ell^2(\widehat{G})} = \langle f, g\rangle_{L^2(G)} \leqslant \|f\|_{L^p(G)}\|g\|_{L^{p'}(G)}$$

where p' is conjugate to p. We then apply Rudin's inequality to see that

$$\|g\|_{L^{p'}(G)} = O(\sqrt{p'}\|g\|_{\ell^2(\Lambda)}) = O\left(\sqrt{\frac{p}{p-1}}\|f\|_{\ell^2(\Lambda)}\right)$$

and the result follows on dividing out by $\|f\|_{\ell^2(\Lambda)}$. □

As a simple corollary of this we can refine Lemma 2.2, improving the bound on the co-dimension of the subspace there from $O(\epsilon^{-2}\beta^{-1})$ to $O(\epsilon^{-2}\log\beta^{-1})$.

Corollary 3.4 (Chang's theorem, [Cha02])**.** *Suppose that $B \subset G$ has density β and $\epsilon \in (0,1]$. Then*

$$\operatorname{cod}\operatorname{Spec}_\epsilon(1_B)^\perp = O(\epsilon^{-2}\log\beta^{-1}).$$

Proof. Let $\Lambda \subset \mathrm{Spec}_\epsilon(1_B)$ be a maximal algebraically independent subset. Then we have that $\mathrm{Spec}_\epsilon(1_B)^\perp = \Lambda^\perp$, and $\operatorname{cod}\mathrm{Spec}_\epsilon(1_B)^\perp = |\Lambda|$. On the other hand

$$\epsilon^2\beta^2|\Lambda| \leqslant \sum_{\lambda\in\Lambda}|\widehat{1_B}(\lambda)|^2 = O\left(\frac{p}{p-1}\|1_B\|^2_{L^p(G)}\right) = O\left(\frac{p}{p-1}\beta^{2/p}\right)$$

by the dual of Rudin's inequality. Setting $p = 1 + 1/\log\beta^{-1}$ then gives the result. □

Chang's theorem can be easily used to refine Bogolyubov's lemma (Theorem 2.1), but we shall use it to look at a harder problem: two-fold sumsets.

3.5. **Application: Subspaces in sumsets.** We have seen that if A has density α then $4A$ contains a subspace of co-dimension $O_\alpha(1)$. On the other hand a random set shows that A itself need not contain a large subspace. What happens in between?

If A is highly structured or highly random then $A+A$ contains a large subspace. However there is an example due to Ruzsa [Ruz91] adapted to the model setting by Green [Gre02b] which indicates some limitations.

Example (Niveau set construction)**.** Let $G := \mathbb{F}_2^n$ and

$$A := \{x \in G : x \text{ has at least } n/2 + \eta\sqrt{n}/2 \text{ ones in it.}\}.$$

Then

$$\mu_G(A) \sim \frac{1}{\sqrt{2\pi}}\int_\eta^\infty \exp(-x^2/2)dx \geqslant \frac{1}{2} - O(\eta),$$

by the central limit theorem so we think of A as having density close to $1/2$. On the other hand, if $x, y \in A$ then $x + y$ has at most $n - 2\eta\sqrt{n}$ ones in it.

Now suppose that W is an affine subspace, say a coset of some linear subspace V, and $\operatorname{cod} V \leqslant d$. Then it is an exercise in linear algebra to show that W contains a vector with at least $n-d$ ones in it, and it follows that $A+A$ cannot contain a subspace of co-dimension less than $2\eta\sqrt{n} + 1$.

Example (Sumsets of very large sets)**.** Suppose that $A \subset G$ has $\mu_G(A) > 1/2$. Then

$$\begin{aligned} 1_A * 1_A(x) &= \mu_G(A \cap (x + A)) \\ &= 2\mu_G(A) - \mu_G(A \cup (x + A)) \geqslant 2\mu_G(A) - 1 > 0 \end{aligned}$$

for all $x \in G$. It follows that $A + A = G$ and so the sumset contains a subspace of co-dimension 0.

Complementing this we have the following result due to Green. It is formally proved in [Gre02b], although the method is that of [Gre02a].

Theorem 3.6. *Suppose that $A \subset G$ has density α. Then $A+A$ contains an affine subspace of dimension $\Omega(\alpha^2 n)$.*

The argument is iterative and based around the following lemma. The proof of this lemma uses Chang's theorem which is not altogether surprising given the Niveau set example.

Lemma 3.7. *Suppose that $G := \mathbb{F}_2^n$, $A \subset G$ has density α and $k \leqslant n$ is a natural. Then either $A + A$ contains an affine subspace of dimension k or else there is a subspace V of co-dimension $O(\alpha^{-2}k)$ such that $\|1_A * \mu_V\|_{L^\infty(G)} \geqslant \alpha(1 + 1/2)$.*

Proof. Write $S := (A + A)^c$ and suppose that $\sigma := \mu_G(S) < 2^{-k}$. Now, suppose that $H \leqslant G$ is any subspace of dimension k, then

$$\int \mu_G(S \cap (x + H))d\mu_G(x) = \int 1_S * 1_H(x)d\mu_G(x) = \sigma\mu_G(H) < 1/|G|,$$

and so there is some $x \in G$ such that $S \cap (x + H) = \emptyset$, whence $A + A \supset x + H$. Thus we assume that this is not the case so that $\sigma \geqslant 2^{-k}$.

Now we apply Plancherel's theorem to the obvious inner product:

$$0 = \langle 1_A * 1_A, 1_S\rangle_{L^2(G)} = \sum_{\gamma\in\widehat{G}} |\widehat{1_A}(\gamma)|^2\widehat{1_S}(\gamma).$$

We have $\widehat{1_A}(0_{\widehat{G}}) = \alpha$ and $\widehat{1_S}(0_{\widehat{G}}) = \sigma$, so

$$\alpha^2\sigma \leqslant \sum_{\gamma\neq 0_{\widehat{G}}} |\widehat{1_A}(\gamma)|^2|\widehat{1_S}(\gamma)|$$

by the triangle inequality. As before we apply Parseval to get that

$$\sum_{\gamma\in \mathrm{Spec}_{\alpha/2}(1_S)\setminus\{0_{\widehat{G}}\}} |\widehat{1_A}(\gamma)|^2|\widehat{1_S}(\gamma)| \geqslant \alpha^2\sigma - (\alpha\sigma/2).\sum_{\gamma\in\widehat{G}} |\widehat{1_A}(\gamma)|^2 \geqslant \alpha^2\sigma/2.$$

Dividing out by $|\widehat{1_S}|$ (which is at most σ by Hausdorff-Young) we get that

$$\sum_{\gamma\in \mathrm{Spec}_{\alpha/2}(1_S)} |\widehat{1_A}(\gamma)|^2 \geqslant \alpha^2 + \sum_{\gamma\in \mathrm{Spec}_{\alpha/2}(1_S)\setminus\{0_{\widehat{G}}\}} |\widehat{1_A}(\gamma)|^2 \geqslant \alpha^2(1 + 1/2).$$

Let $V = \mathrm{Spec}_{\alpha/2}(1_S)^\perp$ and apply Chang's theorem to bound its co-dimension from above by $O(\alpha^{-2}\log\sigma^{-1}) = O(\alpha^{-2}k)$. On the other hand $\widehat{\mu_V}(\gamma) = 1$ if $\gamma \in \mathrm{Spec}_{\alpha/2}(1_S)$ and so by Parseval's theorem we have

$$\alpha^2(1 + 1/2) \leqslant \sum_{\gamma\in\widehat{G}} |\widehat{1_A}(\gamma)|^2|\widehat{\mu_V}(\gamma)|^2 = \|1_A * \mu_V\|^2_{L^2(G)}.$$

Now Hölder's inequality tells us that

$$\|1_A * \mu_V\|^2_{L^2(G)} \leqslant \|1_A * \mu_V\|_{L^1(G)}\|1_A * \mu_V\|_{L^\infty(G)} = \alpha\|1_A * \mu_V\|_{L^\infty(G)},$$

and the result follows on dividing by α. □

Proof of Theorem 3.6. We define a sequence of (linear) subspaces $(V_i)_i$ iteratively and write

$$\alpha_i := \|1_A * \mu_{V_i}\|_{L^\infty(G)} \text{ and } d_i := \mathrm{cod}_G(V_i).$$

Here $\mathrm{cod}_G(V)$ denotes the co-dimension of V in G, rather than any other superspace. We initiate the sequence with $V_0 := G$ and so $\alpha_0 = \alpha$.

Suppose that we are at stage i. Then let x_i be such that $1_A * \mu_{V_i}(x_i) = \alpha_i$ and apply Lemma 3.7 to $(A - x_i) \cap V_i$ and the linear space V_i. Then either $\dim V_i \leqslant k$; or $(A -$

$x_i) \cap V_i + (A - x_i) \cap V_i$ contains an affine subspace of dimension k; or there is a subspace $V_{i+1} \leqslant V_i$ such that

$$d_{i+1} \leqslant d_i + O(\alpha_i^{-2}k)$$

and

$$\alpha_i(1 + 1/2) \leqslant \|1_{(A-x_i)\cap V_i} * \mu_{V_{i+1}}\|_{L^\infty(V_i)} \leqslant \|1_A * \mu_{V_{i+1}}\|_{L^\infty(G)}.$$

Note that after i steps of the last case we have $\alpha_i \geqslant (1+1/2)^i\alpha$. However, this quantity can be at most 1 so we must have ended up in one of the first two cases inside $i_0 = O(\log \alpha^{-1})$ steps; stop the iteration at this point. Then

$$\begin{aligned} d_{i_0} &= O(\alpha_0^{-2}k + \alpha_1^{-2}k + \cdots + \alpha_{i_0-1}^{-2}k) \\ &= O(\alpha^{-2}k(1 + (1+1/2)^{-2} + (1+1/2)^{-4} + \dots)) = O(\alpha^{-2}k). \end{aligned}$$

If $\dim V_{i_0} \leqslant k$ then $n \leqslant k + d_{i_0} = O(\alpha^{-2}k)$, so we can pick $k = \Omega(\alpha^2 n)$ such that this is not so. In that case we have some (affine) subspace of dimension k inside

$$(A - x_i) \cap V_i + (A - x_i) \cap V_i \subset A + A,$$

and the result is proved. □

This sort of density increment argument is very common in additive combinatorics and originates with the work of Roth [Rot52, Rot53]. The above approach actually uses an energy-increment variant due to Heath-Brown [HB87] and Szemerédi [Sze90].

3.8. **Application: Grothendieck's inequality.** We now turn to an application of a more function-analytic nature. Suppose that M is an $n \times n$ matrix such that

$$|\sum_{i,j} M_{ij}x_iy_i| \leqslant \sup_{i,j} |x_i||y_j| \text{ for all real sequences } (x_i)_i, (y_j)_j. \tag{3.1}$$

(Another way of saying this is that $\|M\|_{\infty\to 1} \leqslant 1$.) The real numbers equipped with multiplication form a 1-dimensional (real) Hilbert space and we can ask to what extent the sequence of real numbers can be replaced with elements of a higher dimensional (real) Hibert space.

In particular, suppose that H is a (real) d-dimensional Hilbert space. We may certainly suppose[2] that $H \cong L^2(X)$ for some finite set X of size d, and hence that $H = L^2(X)$. Then

$$\sum_{i,j} M_{i,j}\langle v_i, w_j\rangle_{L^2(X)} = \int \sum_{i,j} M_{i,j}v_i(x)w_j(x)d\mu_X(x). \tag{3.2}$$

On the other hand the Cauchy-Schwarz inequality tells us that

$$\|v\|_{L^\infty(X)} \leqslant \sqrt{d}\|v\|_{L^2(X)} \text{ for all } v \in L^2(X), \tag{3.3}$$

so we get that

$$|\sum_{i,j} M_{i,j}\langle v_i, w_j\rangle_{L^2(X)}| \leqslant d \sup_{i,j} \|v_i\|_{L^2(X)}\|w_j\|_{L^2(X)}$$

[2]See the exercises.

from the hypothesis (3.1). It follows that we can write K_d for the smallest constant such that for all $n \times n$ matrices M satisfying (3.1) and all d-dimensional real Hilbert spaces H we have

$$|\sum_{i,j} M_{i,j}\langle v_i, w_j\rangle_H| \leqslant K_d \sup_{i,j} \|v_i\|_{L^2(X)}\|w_j\|_H.$$

Note that if we restrict to the (Hilbert) subspace generated by $(v_i)_i, (w_j)_j$, then none of the quantities of concern change and so we have $K_d \leqslant K_{2n}$.

In this notation our previous argument showed that $K_d \leqslant d$, whence $K_d \leqslant \min\{d, 2n\}$, and Grothendieck's inequality tells us that K_d is bounded by an absolute constant.

Theorem 3.9 (Grothendieck's inequality)**.** *We have that $K_d = O(1)$.*

Proof. We continue to assume, as we may, that $H = L^2(X)$. In general we cannot do better than (3.3). However, if the large values of the vectors v_i and w_j have small L^2-mass then we can. Let v_i and w_i be such that

$$K_d = |\sum_{i,j} M_{i,j}\langle v_i, w_j\rangle_{L^2(X)}| \text{ and } \|v_i\|_{L^2(X)}, \|w_j\|_{L^2(X)} \leqslant 1.$$

Decompose the v_is and w_js into their large and small parts: $v_i = v_i^L + v_i^S$ and $w_j = w_j^L + w_j^S$ where

$$v_i^L(x) := \begin{cases} v_i(x) & \text{if } |v_i(x)| \geqslant K \\ 0 & \text{otherwise.} \end{cases} \text{ and } w_j^L(x) := \begin{cases} w_j(x) & \text{if } |w_j(x)| \geqslant K \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$\begin{aligned} |\sum_{i,j} M_{i,j}\langle v_i, w_j\rangle_{L^2(X)}| &\leqslant |\sum_{i,j} M_{i,j}\langle v_i^S, w_j^S\rangle_{L^2(X)}| \\ &\quad + |\sum_{i,j} M_{i,j}\langle v_i^L, w_j\rangle_{L^2(X)}| + |\sum_{i,j} M_{i,j}\langle v_i^S, w_j^L\rangle_{L^2(X)}| \\ &\leqslant K^2 + K_d \max_i \|v_i^L\|_{L^2(X)} + K_d \max_j \|w_j^L\|_{L^2(X)}. \end{aligned}$$

Since the left hand side is just K_d, we are done if we can show that the two maxima on the right are small for some $K = O(1)$. Of course this is not true, but Rudin's inequality provides us with an isometric embedding to a space where it is.

Specifically, let $G = \mathbb{F}_2^d$ and $(\lambda_x)_{x\in X}$ be a set of d independent characters in $\widehat{G}$ and put

$$\widetilde{v_i} := \frac{1}{\sqrt{d}} \sum_{x\in X} v_i(x)\lambda_x \text{ and } \widetilde{w_j} := \frac{1}{\sqrt{d}} \sum_{x\in X} w_j(x)\lambda_x.$$

By Plancherel's theorem we have that

$$\langle \widetilde{v_i}, \widetilde{w_j}\rangle_{L^2(G)} = \langle v_i, w_j\rangle_{L^2(X)}.$$

Now, writing $\widetilde{v_i} = \widetilde{v_i}^L + \widetilde{v_i}^S$ and $\widetilde{w_j} = \widetilde{w_j}^L + \widetilde{w_j}^S$ in the same way as before, we see that

$$|\sum_{i,j} M_{i,j}\langle \widetilde{v_i}, \widetilde{w_j}\rangle_{L^2(G)}| \leqslant K^2 + K_{2^d} \max_i \|\widetilde{v_i}^L\|_{L^2(G)} + K_{2^d} \max_j \|\widetilde{w_j}^L\|_{L^2(G)}. \tag{3.4}$$

On the other hand, by Rudin's inequality for $p = 4$ we have that

$$K^2\|\widetilde{v_i}^L\|_{L^2(G)}^2 = \int |\widetilde{v_i}^L|^2 K^2 d\mu_G \leqslant \|\widetilde{v_i}\|_{L^4(G)}^4 = O(\|v_i\|_{L^2(X)}^4) = O(1),$$

and similarly for $\widetilde{w_j}^L$. It follows that there is a choice of $K = O(1)$ such that the maxima in (3.4) are each at most $1/4$, and hence

$$K_d = |\sum_{i,j} M_{i,j}\langle \widetilde{v_i}, \widetilde{w_j}\rangle_{L^2(G)}| \leqslant O(1) + \frac{1}{2}K_{2^d}.$$

Finally $K_{2^d} \leqslant 2n$ for all d whence

$$K_d \leqslant O(1) + \frac{1}{2}O(1) + \cdots + \frac{1}{2^l}O(1) + \frac{1}{2^{l+1}}n \leqslant O(1) + \frac{1}{2^l}n$$

for all l. Letting l tend to infinity completes the proof. □

The smallest universal constant above is called Grothendieck's constant for the reals, and determining its value is an open problem. One can prove a version for Hilbert spaces over the complex numbers by splitting into real and imaginary parts, and the constant there is different.

4. Voting, influence and Boolean functions

We are finally at a point where we shall introduce the definition of a Boolean function, although we have been using them for some time.

A *Boolean function* is a function $f : \{0,1\}^n \to \{0,1\}$. Such functions are the indicator functions of subsets of $\{0,1\}^n$, in particular $f = 1_{\operatorname{supp} f}$, and we shall think of Boolean functions and sets interchangeably.

Often Boolean functions are envisaged as voting schemes: one imagines having two candidates 0 and 1 and n voters each of whom votes for one of the candidates. The winner is the value of $f(x)$ when the ith voter votes for candidate x_i. With this is mind we have some examples.

Example (Dictatorships)**.** The function $f(x) = x_i$ for some i is called a dictatorship because the candidate chosen only depends on who the ith person voted for.

Example (Majority)**.** The majority function is the function

$$f(x) := \begin{cases} 1 & \text{if } x_1 + \cdots + x_n > n/2 \\ 0 & \text{otherwise.} \end{cases}$$

Example (Parity)**.** The parity function if the function

$$f(x) := \begin{cases} 1 & \text{if } x_1 + \cdots + x_n \text{ is odd} \\ 0 & \text{otherwise.} \end{cases}$$

Example (Juntas)**.** A function is a k-junta if there is some subset $S \subset \{1, \dots, n\}$ such that the value of $f(x)$ only depends on the values of $(x_i)_{i \in S}$. In particular a dictatorship is a 1-junta.

We shall be interested in questions such as how much influence a voter or block of voters has in a system, but we should remark that we make no assumptions about the voting system being particularly sensible. Indeed, it seems unlikely that one would choose to implement the parity system above.

4.1. **The connection to dyadic groups and Beckner's inequality.** The set $\{0,1\}^n$ may be endowed with a group (in fact vector space) structure in a natural way. We put

$$(x+y)_i := x_i + y_i \pmod 2 \text{ for all } x, y \in \{0,1\}^n.$$

The resulting group is (isomorphic to) $\mathbb{F}_2^n$, and we shall denote it by G. What is different about this chapter is that we have implicitly chosen a set of generators – the canonical basis of $\{0,1\}^n$, denoted $e_1, \dots, e_n$, so that e_i has zeros everywhere except the ith position where it has a 1. In particular, $x_i = x \cdot e_i$.

If G has the above form then we let $\gamma_1, \dots, \gamma_n$ be the maps $x \mapsto (-1)^{x_i}$. The vectors $\gamma_1, \dots, \gamma_n$ are independent and so form a basis for $\widehat{G}$, and we write $|\gamma|$ for the number of characters from $\{\gamma_1, \dots, \gamma_n\}$ in the expression for γ.

A rather powerful tool in this setting (where a basis is specified) is called Beckner's inequality. Suppose $G = \{0,1\}^n$ (thought of as a group of exponent 2) and $\epsilon \in (0,1]$. We define

$$p_\epsilon(x) := \prod_{i=1}^n (1 + \epsilon\gamma_i) = \prod_{i=1}^n (1 + \epsilon(-1)^{x_i}).$$

We can then calculate the Fourier transform and see that

$$\widehat{p_\epsilon}(\gamma) = \sum_{S \subset [n]} \epsilon^{|S|} \int \gamma \prod_{i \in S} \gamma_i d\mu_G = \epsilon^{|\gamma|}.$$

We can now state Beckner's inequality which should remind you of the dual of Rudin's inequality since $\widehat{p_\epsilon}(\gamma_i)$ is ϵ on the set $\{\gamma_1, \dots, \gamma_n\}$ and $o(\epsilon)$ everywhere else except the trivial character.

Theorem 4.2 (Beckner's inequality)**.** *Suppose that G and p_ϵ are as above. Then*

$$\|\widehat{p_\epsilon}\widehat{f}\|_{\ell^2(\widehat{G})} = \|p_\epsilon * f\|_{L^2(G)} \leqslant \|f\|_{L^{1+\epsilon^2}(G)} \text{ for all } f \in L^{1+\epsilon^2}(G).$$

We shall effectively prove this by taking tensor products of the $n = 1$ case which we begin with in the following lemma.

Lemma 4.3 (The two-point lemma)**.** *For all reals a, b we have*

$$a^2 + \epsilon^2 b^2 \leqslant \left(\frac{|a+b|^{1+\epsilon^2} + |a-b|^{1+\epsilon^2}}{2} \right)^{2/(1+\epsilon^2)}.$$

Proof. Clearly we may assume that $|a| \geqslant |b|$ since the right hand side is symmetric in a and b and so dividing by $|a|$ it follows that the result is proved if we can show that

$$1+\epsilon^2 y^2 \leqslant \left(\frac{(1+y)^{1+\epsilon^2}+(1-y)^{1+\epsilon^2}}{2}\right)^{2/(1+\epsilon^2)} \quad \text{for all } |y|\leqslant 1. \tag{4.1}$$

Both sides are continuous in y, so it suffices to prove the inequality for $|y|<1$ where we can use the binomial theorem to expand out the powers of $(1\pm y)$ on the right. Specifically, put $p=1+\epsilon^2$ and note that

$$\begin{aligned}\frac{1}{2}\left((1+y)^{1+\epsilon^2}+(1-y)^{1+\epsilon^2}\right) &= \sum_{r=0}^{\infty}\binom{p}{r}\frac{1}{2}(y^r+(-y)^r)\\ &= \sum_{l=0}^{\infty}\binom{p}{2l}y^{2l} \geqslant 1+\epsilon^2y^2\frac{p}{2}\end{aligned}$$

since $\binom{p}{2l}\geqslant 0$ for all $l\in\mathbb{N}_0$. On the other hand $(1+x)^\theta\leqslant 1+\theta x$ for all $\theta\in[0,1]$ and $x\geqslant 0$, which applied with $x=\epsilon^2y^2$ and $\theta=p/2$ gives us (4.1). The result is proved. □

Proof of Theorem 4.2. We proceed by induction on n. The base case is done by the two-point lemma. We write $G_j:=\{0,1\}^j$ and define the operator T_j as follows:

$$T_j : L^2(G_j)\to L^p(G_j); f\mapsto \int\prod_{i=1}^{j}(1+\epsilon(-1)^{x_i+y_i})f(y_1,\dots,y_j)dx_1\dots dx_j,$$

and suppose that we have established the jth case of the induction. Given $f\in L^2(G_{j+1})$ write $f=g+\gamma_{j+1}h$ for two functions $g,h\in L^2(G_j)$, and note that

$$T_{j+1}f = T_jg+\epsilon\gamma_{j+1}T_jh.$$

In particular

$$T_{j+1}f(x_1,\dots,x_{j+1}) = T_jg(x_1,\dots,x_j)+\epsilon(-1)^{x_{j+1}}T_jh(x_1,\dots,x_j).$$

Now, put $p=1+\epsilon^2$ and note that

$$\begin{aligned}\|T_{j+1}f\|^2_{L^2(G_{j+1})} &= \int |T_jg(x)|^2+\epsilon^2|T_jh(x)|^2d\mu_{G_j}(x)\\ &\leqslant \int\left(\frac{|T_j(g+h)(x)|^p+|T_j(g-h)(x)|^p}{2}\right)^{2/p}d\mu_{G_j}(x)\end{aligned}$$

by the two-point lemma. Now, put $X:=G_j$ and $Y:=\{0,1\}$ and define k on $X\times Y$ by

$$k(x,y):=\begin{cases}|T_j(g+h)(x)|^p & \text{if } y=0\\ |T_j(g-h)(x)|^p & \text{if } y=1\end{cases},$$

so that

$$\int\left(\frac{|T_j(g+h)(x)|^p+|T_j(g-h)(x)|^p}{2}\right)^{2/p}d\mu_{G_j}(x) = \int\left(\int|k(x,y)|d\mu_Y(y)\right)^{2/p}d\mu_X(x).$$

By the integral triangle inequality (see Appendix A for a statement and proof) with $q = 2/p$ we get that

$$\begin{aligned}\int \left(\int |k(x,y)|d\mu_Y(y)\right)^{2/p} d\mu_X(x) &\leqslant \left(\int \left(\int |k(x,y)|^{2/p} d\mu_X(x)\right)^{p/2} d\mu_Y(y)\right)^{2/p} \\ &= \left(\frac{1}{2}\left(\left(\int |T_j(g+h)(x)|^2 d\mu_{G_j}(x)\right)^{p/2}\right.\right. \\ &\quad \left.\left. + \left(\int |T_j(g-h)(x)|^2 d\mu_{G_j}(x)\right)^{p/2}\right)\right)^{2/p}.\end{aligned}$$

However, by the inductive hypothesis we have

$$\left(\int |T_j(g+h)(x)|^2 d\mu_{G_j}(x)\right)^{p/2} \leqslant \|g+h\|^p_{L^p(G_j)},$$

and similarly for $g-h$, hence (combining everything we have done so far)

$$\|T_{j+1}f\|^2_{L^2(G_{j+1})} \leqslant \left(\frac{1}{2}\left(\|(g+h)\|^p_{L^p(G_j)} + \|(g-h)\|^p_{L^p(G_j)}\right)\right)^{2/p} = \|f\|^2_{L^p(G_{j+1})}.$$

The result is proved. □

Again the name here is not quite right. Theorem 4.2 is more properly attributed to Bonami [Bon70], although some point to Nelson [Nel73]. The additive combinatorial literature often refers to it as Beckner's inequality and much of the computer science literature to it as the Bonami-Beckner inequality.

4.4. **Influence.** Given a Boolean function f, the *influence* of voter i is denoted $\sigma_i(f)$ and is defined to be the probability that i changing his vote effects the outcome if all voters vote uniformly at random:

$$\sigma_i(f) := \int |f_i|^2 d\mu_G(x) \text{ where } f_i(x) = f(x) - f(x+e_i).$$

In particular, notice by Parseval that

$$\sigma_i(f) = 4 \sum_{\gamma \in \widehat{G}: \gamma(e_i) = -1} |\widehat{f}(\gamma)|^2. \tag{4.2}$$

Example (Influence in dictatorships)**.** If f is a dictatorship with i the dictator then $\sigma_i(f) = 1$ and $\sigma_j(f) = 0$ for all $j \neq i$. This is, perhaps, not altogether surprising.

Example (Influence in parity)**.** If f is the parity function then it is easy to see that $\sigma_i(f) = 1$ for all i.

One of the central questions we want to ask is whether there is always a voter with large influence. In some sense this is obviously not the case: if f has very small variance then clearly no voter has a large influence. However, if we insist that the voting scheme is 'fair' in the sense that

$$\int f d\mu_G \sim 1/2$$

then we might hope to find an influential voter. A trivial estimate follows from (4.2): the total influence is

$$I(f) := \sum_{i=1}^n \sigma_i(f) = 4\sum_{\gamma\in\widehat{G}} |\gamma||\widehat{f}(\gamma)|^2 \geqslant 4\sum_{\gamma\neq 0_{\widehat{G}}} |\widehat{f}(\gamma)|^2 = 4\,\mathrm{Var}(f).$$

If f is 'fair' then this is asymptotically 1 and hence, by averaging, there is a voter with influence at least $\Omega(1/n)$. It turns out that there are examples where no voter has much more influence than this.

Example (The tribes example)**.** Suppose that

$$f(x) := x_1\dots x_k \vee x_{k+1}\dots x_{2k} \vee \dots \vee x_{(r-1)k+1}\dots x_{rk}$$

where $\vee$ denotes logical OR. First we need to determine a relationship between r and k that will make this function 'fair'. Specifically, then, we put

$$\frac{1}{2} \sim \int f d\mu_G = 1-(1-2^{-k})^r.$$

This tells us that we want to take $r \sim 2^k \log 2$, and of course putting $n := rk$ we find that

$$k = \log_2 n - \log_2\log n + O(1) \text{ and } r \sim \frac{n}{\log_2 n}.$$

Now, by symmetry all voters have the same influence: so, for example, x_1 influences the outcome iff $x_{(i-1)k+1}\dots x_{ik} = 0$ for all $i \in \{2,\dots,r\}$ and $x_i = 1$ for all $i \in \{2,\dots,k\}$. The probability that $x_{(i-1)k+1}\dots x_{ik} = 0$ for some i is $1-2^{-k}$, so the probability that x_1 influences the outcome is

$$2^{1-k}(1-2^{-k})^{r-1} \leqslant 2^{1-k}\exp(-2^{-k}(r-1)) = O\left(\frac{\log n}{n}\right).$$

That is to say, none of the voters has very much influence.

Interestingly, Beckner's inequality was used by Kahn, Kalai and Linial in [KKL88] to establish that the tribes upper bound has a matching lower bound although the argument is more complicated than the trivial averaging earlier.

Theorem 4.5 (KKL)**.** *Suppose that f is a Boolean function with $\mathrm{Var}(f) = \Omega(1)$. Then there is some i such that*

$$\sigma_i(f) = \Omega\left(\frac{\log n}{n}\right).$$

Proof. By Beckner's inequality with $\epsilon = 1/2$ we have that

$$\begin{aligned}\sum_{|\gamma|\leqslant d,\gamma(e_i)=-1} |\widehat{f}(\gamma)|^2 &\leqslant 4^d \sum_{\gamma(e_i)=-1} 2^{-2|\gamma|}|\widehat{f}(\gamma)|^2 \\ &= 4^{d-1}\|p_{1/2} * f_i\|^2_{L^2(G)} \leqslant 4^{d-1}\|f_i\|^2_{L^{5/4}(G)} = 4^{d-1}\sigma_i(f)^{8/5}.\end{aligned}$$

It follows that if $\sigma_i(f) \leqslant n^{-5/6}$ for all i (which we may as well assume since otherwise we'd be done) then

$$\sum_{|\gamma|\leqslant d} |\gamma||\widehat{f}(\gamma)|^2 \leqslant 4^{d-1}n.(n^{-5/6})^{8/5} \leqslant 4^{d-1}n^{-1/3}.$$

Now pick $d = \Omega(\log n)$ such that this is at most $\mathrm{Var}(f)/2$ for sufficiently large n. Then

$$\mathrm{Var}(f) = \sum_{\gamma\neq 0_{\widehat{G}}} |\widehat{f}(\gamma)|^2 \leqslant \sum_{\gamma:|\gamma|>d} |\widehat{f}(\gamma)|^2 + \mathrm{Var}(f)/2,$$

and it follows immediately that

$$I(f) = 4\sum_{\gamma} |\gamma||\widehat{f}(\gamma)|^2 > d \sum_{\gamma:|\gamma|>d} |\widehat{f}(\gamma)|^2 = \Omega(d) = \Omega(\log n),$$

and we are done by averaging. □

What we actually showed in this argument is that if the individual influences are small, then the total influence is rather large. It turns out that we can establish something stronger, namely that if f has small total influence then it is in some sense close to a junta.

Theorem 4.6 (Friedgut)**.** *Suppose that f is a Boolean function with $I(f) \leqslant C$ and $\eta \in (0,1]$ is a parameter. Then there is an $\exp(O(C\eta^{-1}))$-junta g such that $\|f-g\|^2_{L^2(G)} \leqslant \eta$.*

Proof. Suppose that V is a subspace of G and put

$$g(x) = \begin{cases} 1 & \text{if } f * \mu_V(x) \geqslant 1/2 \\ 0 & \text{otherwise.} \end{cases}.$$

Note that g is also Boolean, constant on cosets of V and we have the point-wise estimate

$$|f(x) - g(x)| \leqslant 2|f(x) - f * \mu_V(x)| \text{ for all } x \in G.$$

Hence

$$\|f-g\|^2_{L^2(G)} \leqslant 4\|f - f*\mu_V\|^2_{L^2(G)} = 4\sum_{\gamma\notin V^\perp} |\widehat{f}(\gamma)|^2$$

Of course control of the total influence implies that the energy is concentrated on low levels:

$$d\sum_{\gamma:|\gamma|\geqslant d} |\widehat{f}(\gamma)|^2 \leqslant \sum_{\gamma\in\widehat{G}} |\gamma||\widehat{f}(\gamma)|^2 \leqslant C,$$

and so

$$\sum_{\gamma\notin V^\perp} |\widehat{f}(\gamma)|^2 \leqslant 4^d \sum_{\gamma\notin V^\perp:|\gamma|\leqslant d} 2^{-2|\gamma|}|\widehat{f}(\gamma)|^2 + \frac{C}{d}.$$

Write $I := \{i : \sigma_i(f) \geqslant \tau\}$ (whence $|I| \leqslant \tau^{-1}C$) and let $V = \{\gamma_i : i \in I\}^\perp$, so that if $\gamma \notin V^\perp$ then there is some $i \notin I$ such that $\gamma(e_i) = -1$. It follows that

$$\begin{aligned}\sum_{\gamma\notin V^\perp:|\gamma|\leqslant d} 2^{-2|\gamma|}|\widehat{f}(\gamma)|^2 &\leqslant \sum_{i\notin I}\sum_{\gamma(e_i)=-1} 2^{-2|\gamma|}|\widehat{f}(\gamma)|^2\\ &= 4^{-1}\sum_{i\notin I}\|p_{1/2}*f_i\|^2_{L^2(G)} \leqslant \sum_{i\notin I}\sigma_i(f)\tau^{3/5} \leqslant C\tau^{3/5},\end{aligned}$$

where the f_is are as in the previous proof. Combining all this we conclude that

$$\|f-g\|^2_{L^2(G)} \leqslant C4^d\tau^{3/5} + \frac{C}{d}.$$

Putting $d = \lceil 2C\eta^{-1}\rceil$ we see that we can take $\tau = \exp(-O(C\eta^{-1}))$ and ensure that the difference is at most η. The result is proved since g is clearly an $O(\tau^{-1}C)$-junta. □

Note that in both proofs we were only interested in using Beckner's inequality for some $\epsilon < 1 - \Omega(1)$, rather than a very small value as we did in Chapter 3. On the other hand we needed the additional strength of Beckner to estimate the ℓ^2-mass of $\widehat{f}$ on the sets $\{\gamma : |\gamma| = d\}$.

5. Approximate structures: sets with small sumset

In this chapter we are interested in approximate substructures of $\mathbb{F}_2^n$. The reason for studying them is that approximate structures are often far more plentiful than exact ones (by virtue of having relaxed requirements), but frequently still support a lot of analysis making them almost as useful as their exact counterparts. Our basic aim is to tease out the structure of 'approximate subspaces'.

Suppose that H is an affine subspace of G. Then it is easy to see that $|H+H| = |H|$. Conversely, if $A \subset G$ has $|A+A| = |A|$ then it is not much harder to see that $A+A$ is a subspace of G, and hence A is an affine subspace. It turns out that if the sumset is only a bit bigger then much of this phenomenon persists.

Proposition 5.1. *Suppose that $A \subset G$ is a non-empty set with $|A+A| < 1.5|A|$. Then $A+A$ is a subspace of G.*

To prove this it will be convenient to introduce the notion of a symmetry set. The *symmetry set* of a (non-empty) set A at threshold η is

$$\mathrm{Sym}_\eta(A) := \{x \in G : 1_A * 1_A(x) \geqslant \eta\mu_G(A)\}.$$

Note that $\mathrm{Sym}_\eta(A)$ is a (symmetric) neighbourhood of 0_G contained in $A+A$; these sets are particularly useful because of the following trivial application of the pigeonhole principle.

Lemma 5.2. *Suppose that $A \subset G$ is a non-empty set and $\epsilon, \epsilon' \in [0,1)$ are parameters. Then*

$$\mathrm{Sym}_{1-\epsilon}(A) + \mathrm{Sym}_{1-\epsilon'}(A) \subset \mathrm{Sym}_{1-(\epsilon+\epsilon')}(A).$$

Proof. Suppose that $x \in \mathrm{Sym}_{1-\epsilon}(A)$ and $y \in \mathrm{Sym}_{1-\epsilon'}(A)$. Then

$$\begin{aligned} 1_A * 1_A(x+y) &= \mu_G((x+A) \cap (A+y)) \\ &\geqslant \mu_G(((x+A) \cap A) \cap ((A+y) \cap A)) \\ &\geqslant \mu_G((x+A) \cap A)) + \mu_G((A+y) \cap A) \\ &\quad -\mu_G(((x+A) \cap A) \cup ((A+y) \cap A)) \\ &\geqslant 1_A * 1_A(x) + 1_A * 1_A(y) - \mu_G(A) \\ &\geqslant (1-\epsilon+1-\epsilon'-1)\mu_G(A), \end{aligned}$$

and $x+y \in \mathrm{Sym}_{1-(\epsilon+\epsilon')}(A)$. The result follows. □

Notice, in particular, that $\mathrm{Sym}_1(A)$ is a subspace of G.

Proof of Proposition 5.1. Suppose that $a, a' \in A$ and write $K := \mu_G(A+A)/\mu_G(A)$. Then

$$\begin{aligned} 1_A * 1_A(a+a') &= \mu_G((a+A) \cap (A+a')) \\ &\geqslant \mu_G(a+A) + \mu_G(A+a') - \mu_G((a+A) \cup (A+a')) \\ &\geqslant 2\mu_G(A) - \mu_G(A+A) \geqslant (2-K)\mu_G(A). \end{aligned}$$

It follows that $A - A \subset \mathrm{Sym}_{2-K}(A)$ and hence, by the previous lemma, we have that

$$(A+A)+(A+A) \subset \mathrm{Sym}_{2-K}(A) + \mathrm{Sym}_{2-K}(A) \subset \mathrm{Sym}_{3-2K}(A) \subset A+A$$

since $K < 3/2$; it follows that $(A+A)+(A+A) = (A+A)$. On the other hand it is also a (symmetric) neighbourhood of the identity so it follows that it is a subspace. □

To make good use of Lemma 5.2 we needed the symmetry sets we found to be large. For large threshold values this will typically not be the case, however, a simple application of Cauchy-Schwarz actually shows that there are some not too small threshold values for which it is.

Lemma 5.3. *Suppose that A has $|A+A| \leqslant K|A|$. Then*

$$\mu_G(\mathrm{Sym}_{1/2K}(A)) \geqslant \mu_G(A)/2K.$$

Proof. First of all by the Cauchy-Schwarz inequality we have that

$$\int (1_A * 1_A)^2 d\mu_G \geqslant \frac{1}{\mu_G(A+A)} \left(\int 1_A * 1_A d\mu_G \right)^2 = \mu_G(A)^3/K.$$

On the other hand

$$\int_{\mathrm{Sym}_{1/2K}(A)^c} (1_A * 1_A)^2 d\mu_G \leqslant \frac{\mu_G(A)}{2K} \int 1_A * 1_A d\mu_G = \frac{\mu_G(A)^3}{2K},$$

and it follows by the triangle inequality that

$$\frac{\mu_G(A)^3}{2K} \leqslant \int_{\mathrm{Sym}_{1/2K}(A)} (1_A * 1_A)^2 d\mu_G \leqslant \mu_G(A)^2 \mu_G(\mathrm{Sym}_{1/2K}(A))$$

and the result follows on dividing out $\mu_G(A)^2$. □

At a certain point the phenomenon in Proposition 5.1 starts to fail. Indeed, if $A = \{0_G, e_1, e_2, e_3\}$ then

$$A + A = \{0_G, e_1, e_2, e_3, e_1 + e_2, e_2 + e_3, e_1 + e_3\},$$

and so $|A + A| \leqslant 1.75|A|$, but $A + A$ is not a subspace of G.

Despite this example it turns out that a covering argument of Ruzsa [Ruz99] shows that A is contained in a subspace which is not too large. This is the $\mathbb{F}_2^n$ analogue of Freĭman's theorem [Fre73] for subsets of $\mathbb{Z}$ with small sumset.

Theorem 5.4 (Freĭman's theorem for $\mathbb{F}_2^n$)**.** *Suppose that $A \subset G$ is non-empty with $|A + A| \leqslant K|A|$. Then $\langle A\rangle$, the group generated by A, has size at most $\exp(O(K^{O(1)}))|A|$.*

The constant K is called the doubling constant of A, and roughly we think of A as having 'small' doubling if $K = O(1)$.

The theorem is not far from best possible. Indeed, suppose that A is a set of K linearly independent elements. Then $|A + A| \leqslant K|A|$, but $|\langle A\rangle| \geqslant 2^K$, whence the bound on the size of $\langle A\rangle$ cannot be $\exp(o(K))|A|$. In fact a rather precise answer has been given by Green and Tao in [GT09] using compressions.

The core Ruzsa covering argument takes the following form.

Lemma 5.5 (Ruzsa's covering lemma)**.** *Suppose that $A, S \subset G$ are non-empty with $|A + S| \leqslant K|S|$. Then there is a set X with $|X| \leqslant K$ such that $A \subset X + S + S$.*

Proof. Let $X \subset A$ be a maximal S-separated subset meaning maximal such that if $x, x' \in X$ are distinct then $(x + S) \cap (x' + S) = \emptyset$. By separation we see that the sets $x + S$ are disjoint subsets of $A + S$ whence $|X|.|S| \leqslant |A + S| \leqslant K|S|$ and $|X| \leqslant K$.

On the other hand, by maximality we see that if $x \in A$ then either $x \in X \subset X + S + S$, or else $x \notin X$ and so there is some $x' \in X$ such that $(x + S) \cap (x' + S) \neq \emptyset$ and hence $x \in x' + S + S \subset X + S + S$. The lemma is proved. □

Corollary 5.6. *Suppose that $S \subset G$ is non-empty with $|4S| \leqslant K|S|$. Then $\langle S\rangle$ has size at most $K2^K|S|$.*

Proof. We apply the previous lemma with $A = 3S$ to get a set X with $|X| \leqslant K$ such that $3S \subset X + 2S$. It follows by induction that $nS \subset (n-2)X + 2S$, and hence $\langle S\rangle \subset \langle X\rangle + 2S$. On the other hand $|\langle X\rangle| \leqslant 2^K$, and we certainly have $|\langle S\rangle| \leqslant |\langle X\rangle||2S|$ so the result follows. □

Ideally we should now like to prove that $|A + A| \leqslant K|A|$ implies that $|4A| \leqslant K^{O(1)}|A|$. This is true and a special case of the Plünnecke-Ruzsa inequalities which show that

$$|A + A| \leqslant K|A| \Rightarrow |nA| \leqslant K^n|A|.$$

The proofs use graph theoretic methods and takes some time, so we shall follow an easier route. Those interested in the Plünnecke-Ruzsa arguments may wish to consult [TV06, Chapter 6].

We shall work with symmetry sets again because they are somewhat smoother than arbitrary sets with small doubling. This means that the next argument of Tao [Tao08, Proposition 4.5] will give us a Plünnecke-type handle on the growth of symmetry sets.

Proposition 5.7. *Suppose that A has $|A+A| \leqslant K|A|$. Then*

$$|n\,\mathrm{Sym}_c(A) + A + A| \leqslant c^{-n}K^{n+1}|A|.$$

Proof. Suppose that $s_1, \dots, s_n \in \mathrm{Sym}_c(A)$ and $a, a' \in A$, and note that

$$1_{2A}^{(n+1)}(s_1 + \cdots + s_n + a + a') = \frac{1}{|G|^n} \sum_{z_1+\cdots+z_{n+1}=s_1+\cdots+s_n+a+a'} 1_{2A}(z_1)\dots 1_{2A}(z_{n+1}).$$

On the other hand if $(b_1, \dots, b_n) \in G^n$ and we put

$$z_1 = a + b_1, z_2 = b_1 + s_1 + b_2, \dots, z_n = b_{n-1} + s_{n-1} + b_n, z_{n+1} = b_n + s_n + a'$$

then

$$z_1 + \cdots + z_{n+1} = s_1 + \cdots + s_n + a + a'$$

and each vector $b \in G^n$ determines a unique vector $z \in G^{n+1}$. It follows that

$$\begin{aligned} 1_{2A}^{(n+1)}(s_1 + \cdots + s_n + a + a') &\geqslant \frac{1}{|G|^n} \sum_{b_1,\dots,b_n} 1_{2A}(a+b_1)\dots 1_{2A}(b_n + s_n + a') \\ &\geqslant \frac{1}{|G|^n} \sum_{b_1,\dots,b_n} 1_A(a)1_A(b_1)\dots 1_A(b_n + s_n)1_A(a') \\ &= \prod_{i=1}^{n} 1_A * 1_A(s_i) \geqslant (c\mu_G(A))^n. \end{aligned}$$

Thus

$$\mu_G(n\,\mathrm{Sym}_c(A) + A + A)(c\mu_G(A))^n \leqslant \int 1_{2A}^{(n+1)} d\mu_G = K^{n+1}\mu_G(A)^{n+1}$$

and the result follows on rearrangement. □

Proof of Theorem 5.4. By Lemma 5.3 we see that $S := \mathrm{Sym}_{1/2K}(A)$ has $|S| \geqslant |A|/2K$. Then by Proposition 5.7 we have that

$$|4S| \leqslant |4S + 2A| \leqslant (2K)^4K^5|A| = O(K^{O(1)}|S|).$$

It follows by Corollary 5.6 that the group generated by S has size at most

$$|\langle S\rangle| \leqslant \exp(O(K^{O(1)}))|S| \leqslant \exp(O(K^{O(1)}))|4S + 2A| \leqslant \exp(O(K^{O(1)}))|A|.$$

Finally, by Ruzsa's covering lemma we have a set X of size $O(K^{O(1)})$ such that

$$A \subset X + S + S \subset \langle X\rangle + \langle S\rangle$$

and the result follows. □

6. Correlation with approximate structures

A quantity which is closely connected to the doubling constant, and also to symmetry sets is the additive energy of a set. Specifically, if $A \subset G$ then the *additive energy* of A is defined to be

$$E(A) := \sum_{x+y=z+w} 1_A(x)1_A(y)1_A(z)1_A(w) = \sum_u \left(\sum_{x+y=u} 1_A(x)1_A(y) \right)^2.$$

In our usual notation this is

$$E(A) = |G|^3 \int (1_A * 1_A)^2 d\mu_G,$$

and in words it is the number of *additive quadruples* in A, that is the number of quadruples $(x,y,z,w) \in A^4$ such that $x+y=z+w$. It is easy to see that

$$E(A) \leqslant \sup_u \sum_{x+y=u} 1_A(x)1_A(y) \cdot \sum_u \sum_{x+y=u} 1_A(x)1_A(y) \leqslant |A| \cdot |A|^2 = |A|^3. \tag{6.1}$$

On the other hand, if $|A+A| \leqslant K|A|$ then by the same application of Cauchy-Schwarz as in Lemma 5.3 we get that

$$E(A) = \sum_u \left(\sum_{x+y=u} 1_A(x)1_A(y) \right)^2 \geqslant \frac{1}{|A+A|} \left(\sum_u \sum_{x+y=u} 1_A(x)1_A(y) \right)^2 \geqslant |A|^3/K,$$

which is to say it is close to the maximum value in (6.1). Of course this maximum in (6.1) is achieved by affine subspaces since another way of thinking of the quantity $E(A)$ is as the number of triples $(x,y,z) \in A^3$ which have $x+y-z \in A$; if a non-empty set A has this property, it is well known that it is a coset of a subgroup.

The additive energy is a particularly useful quantity because it is very stable under small perturbations of the underlying set. Indeed, if we add or remove $o(|A|)$ elements from A then the additive energy changes by $o(|A|^3)$ which is not much if $E(A) = \Omega(|A|^3)$. On the other hand, because of this stability under small perturbations, large additive energy does *not* imply small doubling.

To see this concretely consider, the example of A as a subspace V union $\eta|A|$ independent elements of G whose span intersects V in the trivial vector. It is easy to see that

$$E(A) \geqslant (1-O(\eta))|A|^3 \text{ and } |A+A| = \Omega(\eta|A|^2).$$

In this example, A nevertheless has a large structured part and fortunately this can be recovered.

Theorem 6.1 (Balog-Szemerédi-Gowers)**.** *Suppose that $A \subset G$ satisfies $E(A) \geqslant c|A|^3$. Then there is a subset $A' \subset A$ with $|A'| \geqslant c^{O(1)}|A|$ such that $|A'+A'| \leqslant c^{-O(1)}|A'|$.*

We shall prove this using symmetry sets because in this regard it does turn out that sets with large additive energy behave in a similar way to sets with small sumset. Indeed, the reader may wish to compare the next lemma with Lemma 5.3.

Lemma 6.2. *Suppose that $A \subset G$ has $E(A) \geqslant c|A|^3$. Then*

$$c^{-1}\mu_G(A) \geqslant \mu_G(\mathrm{Sym}_{c/2}(A)) \geqslant c\mu_G(A)/2 \text{ and } \langle 1_A * 1_A, 1_{\mathrm{Sym}_{c/2}(A)}\rangle_{L^2(G)} \geqslant c\mu_G(A)^2/2.$$

Proof. The first inequality is a simple application of the triangle inequality:

$$\frac{c\mu_G(A)}{2}\mu_G(\mathrm{Sym}_{c/2}(A)) \leqslant \int_{\mathrm{Sym}_{c/2}(A)} 1_A * 1_A d\mu_G = \mu_G(A)^2.$$

The third inequality follows since

$$\begin{aligned}
\langle 1_A * 1_A, 1_{\mathrm{Sym}_{c/2}(A)}\rangle_{L^2(G)} &\geqslant \frac{1}{\mu_G(A)}\langle (1_A * 1_A)^2, 1_{\mathrm{Sym}_{c/2}(A)}\rangle_{L^2(G)} \\
&= \frac{1}{\mu_G(A)}\int_{\mathrm{Sym}_{c/2}(A)} (1_A * 1_A)^2 d\mu_G \\
&= \frac{1}{\mu_G(A)}\left(\int (1_A * 1_A)^2 d\mu_G - \int_{\mathrm{Sym}_{c/2}(A)^c} (1_A * 1_A)^2 d\mu_G\right) \\
&\geqslant c\mu_G(A)^2 - \frac{1}{\mu_G(A)}\left(\frac{c\mu_G(A)}{2}\right)\mu_G(A)^2 \\
&\geqslant c\mu_G(A)^2/2,
\end{aligned}$$

and hence the second inequality follows immediately by Hölder's inequality. □

The aim now is to prove Theorem 6.1 in two parts following Sudakov, Szemerédi and Vu [SSV05]. The result was first proved by Balog and Szemerédi in [BS94], and with good bounds by Gowers in [Gow98].

In the next lemma we shall find a large subset A' of A such that almost all pairs $(x, y) \in A'^2$ have $x + y$ in a symmetry set of A. In particular this means that almost all of the pairs in A'^2 represent one of at most $O(|A|)$ elements. This is close to having small doubling and we then complete the proof by a pigeonhole argument.

Lemma 6.3. *Suppose that $A \subset G$ has $E(A) \geqslant c|A|^3$ and $\epsilon \in (0, 1]$ is a parameter. Then there is a subset $A' \subset A$ with $|A'| = \Omega(c|A|)$ such that*

$$|\{(x, y) \in A'^2 : x + y \in \mathrm{Sym}_{\epsilon c^2/2}(A)\}| \geqslant (1 - \epsilon)|A'|^2.$$

Proof. We let X be a random variable such that

$$\mathbb{P}(X = z) = \frac{1_A * 1_A(z)}{\mu_G(A)|A|},$$

and put $A' := A \cap (X + A)$. (Note that this is a valid probability distribution for X.) Then

$$\mathbb{E}|A'| = \mathbb{E}|G|1_A * 1_A(X) = \frac{1}{\mu_G(A)^2}\sum_{z \in G} 1_A * 1_A(z)^2 = \frac{1}{|A|^2}E(A) \geqslant c|A|,$$

and so, by the Cauchy-Schwarz inequality, we have that $\mathbb{E}|A'|^2 \geqslant c^2|A|^2$.

On the other hand for $x, y \in A$ we have

$$\begin{aligned}\mathbb{P}((x,y) \in A'^2) &= \mathbb{P}(x \in X + A, y \in X + A)\\ &= \mathbb{P}(X \in (x+A) \cap (y+A))\\ &\leqslant \sup_{z \in G} \mathbb{P}(X = z).|(x+A) \cap (y+A)|\\ &\leqslant \frac{1_A * 1_A(x+y)}{\mu_G(A)}.\end{aligned}$$

If we now write $B := \{(x,y) \in A'^2 : x + y \notin \mathrm{Sym}_{\epsilon c^2/2}(A)\}$ then

$$\mathbb{E}|B| = \sum_{\substack{x,y \in A\\ x+y \notin \mathrm{Sym}_{\epsilon c^2/2}(A)}} \mathbb{P}((x,y) \in A'^2) \leqslant |A|^2.\epsilon c^2/2 \leqslant \frac{\epsilon}{2}\mathbb{E}|A'|^2.$$

It follows that

$$\mathbb{E}(|A'|^2 - \epsilon^{-1}|B|) \geqslant \frac{1}{2}\mathbb{E}|A'|^2 \geqslant \frac{1}{2}c^2|A|^2$$

and we may pick X such that $|A'|^2 - \epsilon^{-1}|B|) \geqslant c^2|A|^2/2$, and hence $|A'| \geqslant c|A|/\sqrt{2}$ and $|B| \leqslant \epsilon|A'|^2$. The result is proved. □

Proof of Theorem 6.1. Apply Lemma 6.3 with[3] $\epsilon = 1/6$ to see that there is a set $A'' \subset A$ with $|A''| = \Omega(c|A|)$ such that

$$|\{(x,y) \in A''^2 : x + y \in \mathrm{Sym}_{c^2/12}(A)\}| \geqslant (1 - 1/6)|A''|^2. \tag{6.2}$$

In words this says that for most pairs $(x,y) \in A''^2$ we have $x + y \in \mathrm{Sym}_{c^2/12}(A)$. We let A' be the set of xs in A'' such that there are many ys in A'' with $x + y \in \mathrm{Sym}_{c^2/12}(A)$. In particular, for each $x \in A''$ write

$$N_x := \{y \in A'' : x + y \in \mathrm{Sym}_{c^2/12}(A)\}$$

and put

$$A' := \{x \in A'' : |N_x| \geqslant 2|A''|/3\}.$$

The total number of pairs $(x,y) \in A''$ such that $x + y \in \mathrm{Sym}_{c^2/12}(A)$ is then at most

$$\begin{aligned}|A'|.\max_{x \in A'}|N_x| + |A'' \setminus A'|.\max_{x \in A'' \setminus A'}|N_x| &\leqslant |A'|.|A''| + (|A''| - |A'|).2|A''|/3\\ &= |A'||A''|/3 + 2|A''|^2/3.\end{aligned}$$

On the other hand combining this with (6.2) we see that

$$|A'||A''|/3 + 2|A''|^2/3 \geqslant (1 - 1/6)|A''|^2,$$

and hence $|A'| \geqslant |A''|/2$.

Now, if $(x,y) \in A'^2$ then $|N_x|, |N_y| \geqslant 2|A''|/3$ and $N_x, N_y \subset A''$. It follows that

$$|N_x \cap N_y| = |N_x| + |N_y| - |N_x \cup N_y| \geqslant |N_x| + |N_y| - |A''| \geqslant |A''|/3;$$

[3]The reason for the choice of ϵ will become clear later.

if words, there are at least $|A''|/3$ elements $z \in A''$ such that $x + z \in \mathrm{Sym}_{c^2/12}(A)$ and $y + z \in \mathrm{Sym}_{c^2/12}(A)$. We conclude that if $x, y \in A'$ then

$$\begin{aligned} 1_A * 1_A * 1_A * 1_A(x+y) &= \int 1_A * 1_A(x+y+z) 1_A * 1_A(z) d\mu_G(z) \\ &= \int 1_A * 1_A(x+z) 1_A * 1_A(y+z) d\mu_G(z) \\ &\geqslant \int 1_A * 1_A(x+z) 1_A * 1_A(y+z) 1_{A''}(z) d\mu_G(z) \\ &\geqslant (c^2\mu_G(A)/12)^2 \mu_G(A'')/3, \end{aligned}$$

whence

$$\mu_G(A'+A').(c^2\mu_G(A)/12)^2\mu_G(A'')/3 \leqslant \mu_G(A)^4.$$

This rearranges to give

$$\mu_G(A'+A') \leqslant O(c^{-4}\mu_G(A)^2\mu_G(A'')^{-1}) = O(c^{-6}\mu_G(A')),$$

and the result is proved. □

We now turn our attention to rough morphisms. A map $\phi : G \to G$ is a morphism (linear map) if

$$\phi(x+y) = \phi(x) + \phi(y) \text{ for all } x, y \in G.$$

Our next result combines the work of this section and the last to show that if a function ϕ satisfies this relationship for many pairs (x, y) then it is equal to a genuine morphism on a large set. The result itself is due to Samorodnitsky [Sam07] by an argument which was originally developed for the integers by Gowers in [Gow98].

Theorem 6.4 (Rough morphisms)**.** *Suppose that $\phi : G \to G$ is such that*

$$\mu_{G^2}(\{(x,y) : \phi(x+y) = \phi(x) + \phi(y)\}) \geqslant c.$$

Then there is a homomorphism $\theta : G \to G$ such that

$$\mu_G(\{x : \phi(x) = \theta(x)\}) \geqslant \exp(-O(c^{-O(1)})).$$

Proof. We examine the set $A := \{(x, \phi(x)) : x \in G\} \subset G^2$ which has size $|G|$. It turns out that this has large additive energy:

$$\begin{aligned} E(A) = \sum_{\substack{x+y=z+w \\ \phi(x)+\phi(y)=\phi(z)+\phi(w)}} 1 &\geqslant \sum_u \left(\sum_{\substack{x+y=u \\ \phi(x)+\phi(y)=\phi(u)}} 1 \right)^2 \\ &\geqslant \frac{1}{|G|} \left(\sum_u \sum_{\substack{x+y=u \\ \phi(x)+\phi(y)=\phi(u)}} 1 \right)^2 \geqslant c^2|G|^3 = c^2|A|^3. \end{aligned}$$

By the Balog-Szemerédi-Gowers lemma there is a set $A' \subset A$ with $|A'| \geqslant c^{O(1)}|A|$ and $|A' + A'| \leqslant c^{-O(1)}|A'|$.

Let $\pi : G^2 \to G$ be the natural co-ordinate projection map $(x, y) \mapsto x$ so that $\pi(A) = G$ and $|\pi(A')| \geqslant c^{O(1)}|G|$, and let $b_1, \dots, b_m$ be a maximal[4] set of linearly independent elements in $\pi(A')$. Now, by maximality if $a \in \pi(A')$ then either $a = b_i$ for some i or $\{a, b_1, \dots, b_m\}$ is linearly dependent. In the second case

$$\mu.a + \sum_{i=1}^{m} \mu_i.b_i = 0_G \text{ for some } \mu, \mu_1, \dots, \mu_m \in \mathbb{F}_2$$

not all zero. Since $b_1, \dots, b_m$ are linearly independent we see that $\mu \neq 0$, hence $\mu = 1$ and $a \in \langle b_1, \dots, b_m\rangle$. Thus in either case $a \in \langle b_1, \dots, b_m\rangle$ and so it follows that $\pi(A') \subset \langle b_1, \dots, b_m\rangle$, and hence

$$c^{O(1)}|G| \leqslant 2^m,$$

so $m \geqslant n - O(\log c^{-1})$. Now let $v_1, \dots, v_m \in G$ be such that $(b_i, v_i) \in \langle A'\rangle$ for all $i \in \{1, \dots, m\}$ and put $H := \langle (b_1, v_1), \dots, (b_m, v_m)\rangle$.

It is easy to see that $\pi(H) = \pi(\langle A'\rangle)$ and so $|H| \geqslant c^{O(1)}|G|$. Moreover, by Freĭman's theorem $|\langle A'\rangle| \leqslant \exp(O(c^{-O(1)}))|G|$, and hence $|\langle A'\rangle/H| \leqslant \exp(O(c^{-O(1)}))$. The cosets of H in $\langle A'\rangle$ partition $\langle A'\rangle$ and hence

$$|A'| = |A' \cap \langle A'\rangle| = \sum_{W \in \langle A'\rangle/H} |A' \cap W| \leqslant |\langle A'\rangle/H| \sup_{W \in \langle A'\rangle/H} |A' \cap W|.$$

It follows that there is some coset $W = (w, z) + H$ with $(w, z) \in \langle A'\rangle$ such that

$$|((w, z) + H) \cap A'| \geqslant \exp(-O(c^{-O(1)}))|A'| = \exp(-O(c^{-O(1)}))|G|.$$

Since $\pi(H) = \pi(\langle A'\rangle)$ we see that $w + \pi(H) = \pi(H)$, so we may assume that $w = 0_G$ and hence

$$|((0_G, z) + H) \cap A| \geqslant |((0_G, z) + H) \cap A'| \geqslant \exp(-O(c^{-O(1)}))|G|.$$

If z were also to be 0_G we'd be done as we could define θ to be the linear extension of $\theta(b_i) = v_i$; unfortunately this is not true, but we can correct the situation.

Claim. *We may assume that there is some j with $1 \leqslant j \leqslant m$ such that for at least $1/4$ of elements $a \in \pi(((0_G, z) + H) \cap A)$ we have $a \cdot b_j = 1$.*

Proof. Write $P := \pi(((0_G, z) + H) \cap A)$ which has $\mu_G(P) \geqslant \exp(-O(c^{-O(1)}))$, and define characters $\lambda_i(x) := (-1)^{x \cdot b_i}$. Suppose that $|\widehat{1_P}(\lambda_i)| \geqslant \mu_G(P)/2$ for all $1 \leqslant i \leqslant m$. Then by Chang's theorem, since the $b_1, \dots, b_m$ are independent, we have

$$n - O(\log c^{-1}) \leqslant m = O(\log \mu_G(P)^{-1}) = O(c^{-O(1)}).$$

It follows that $n = O(c^{-O(1)})$ and we are trivially done since the lower bound in the conclusion simply asserts that $\phi(x) = \theta(x)$ for at least one x; defining such a θ is trivial. It follows that we may assume there is some $1 \leqslant j \leqslant m$ such that $|\widehat{1_P}(\lambda_j)| \leqslant \mu_G(P)/2$. Now, write

$$P^+ := \{x \in P : x \cdot b_j = 1\} \text{ and } P^- := \{x \in P : x \cdot b_j = 0\}.$$

[4]The reader may care to compare this use of maximality with that in Ruzsa's covering lemma, Lemma 5.5.

From the definition of the Fourier transform and P as a disjoint union of P^+ and P^- we have

$$|\mu_G(P^+) - \mu_G(P^-)| \leqslant \mu_G(P)/2 \text{ and } \mu_G(P^+) + \mu_G(P^-) = \mu_G(P).$$

It follows that $\mu_G(P^+) \geqslant \mu_G(P)/4$, that is $a \cdot b_j = 1$ for at least $1/4$ of the elements $a \in P$. □

Finally, we extend the vectors $b_1, \dots, b_m$ by $b_{m+1}, \dots, b_n$ such that $b_1, \dots, b_n$ is a basis for G and define θ by linear extension from its definition on the basis $(b_i)_i$:

$$\theta(b_i) = \begin{cases} v_i & \text{if } 1 \leqslant i \leqslant m, i \neq j \\ v_j + z & \text{if } i = j \\ 0_G & \text{otherwise.} \end{cases}$$

Now, suppose that $a \in \pi(((0_G, z) + H) \cap A)$ has $a \cdot b_j = 1$. Then

$$a = \sum_{i : a \cdot b_i = 1} b_i \text{ and } \phi(a) = z + \sum_{i : a \cdot b_i = 1} v_i.$$

On the other hand

$$\theta(a) = \sum_{i : a \cdot b_i = 1} \theta(b_i) = z + \sum_{i : a \cdot b_i = 1} v_i = \phi(a),$$

and there are many such a by the claim and the lower bound on the size of $\pi(((0_G, z) + H) \cap A)$. □

7. Polynomial testing

In this section we shall consider polynomials $p : \mathbb{F}_2^n \to \mathbb{F}_2$. These are, of course, the same as Boolean functions in the sense that if we are given $A \subset \mathbb{F}_2^n$ then there is a polynomial $p : \mathbb{F}_2^n \to \mathbb{F}_2$ such that $\{x : p(x) = 1\} = A$. Indeed, we simply define

$$p(x) = \sum_{a \in A} \prod_{i : a \cdot e_i = 1} x_i \prod_{i : a \cdot e_i = 0} (1 - x_i).$$

We are interested in determining whether or not a given function $f : \mathbb{F}_2^n \to \mathbb{F}_2$ correlates with a low degree polynomial.

7.1. **Correlation with linear polynomials.** In the first instance we should like to see if f has a large (affine) linear part, meaning whether it correlates with a function of the form

$$x \mapsto a \cdot x + b$$

where $a \in \mathbb{F}_2^n$, $b \in \mathbb{F}_2$ and $a \cdot x = a_1 x_1 + \cdots + a_n x_n$. If f has this form then it is easy to see (by linearity) that

$$f(x) + f(x + y) + f(x + z) + f(x + y + z) = 0_{\mathbb{F}_2} \tag{7.1}$$

for all $x,y,z \in \mathbb{F}_2^n$. Bearing in mind the Rough Morphism Theorem (Theorem 6.4), we shall be interested in what we can say if this equality is satisfied an unusually large amount of the time. We define

$$\|g\|_{U^2}^4 := \int g(x)g(x+y)g(x+z)g(x+y+z)d\mu_G(x)d\mu_G(y)d\mu_G(z),$$

which it turns out is a norm:

Lemma 7.2. *We have the identity,*

$$\|g\|_{U^2} = \|\widehat{g}\|_{\ell^4(\widehat{G})} \text{ for all } g \in L^\infty(G),$$

so that, in particular, $\|\cdot\|_{U^2}$ is a norm.

Proof. This is an easy calculation using Parseval's theorem and the change of variables $w = x+z$:

$$\begin{aligned}\|g\|_{U^2}^4 &= \int g(x)g(x+y)g(x+z)g(x+y+z)d\mu_G(x)d\mu_G(y)d\mu_G(z)\\ &= \int g(x)g(y+x)g(w)g(y+w)d\mu_G(x)d\mu_G(w)d\mu_G(y)\\ &= \langle g*g, g*g\rangle_{L^2(G)} = \|\widehat{g}\|_{\ell^4(\widehat{G})}^4.\end{aligned}$$

□

Now, the proportion of triples for which (7.1) holds is

$$P(f) := \int \frac{1}{2}(1+(-1)^{f(x)+f(x+y)+f(x+z)+f(x+y+z)})d\mu_G(x)d\mu_G(y)d\mu_G(z) = \frac{1}{2}+\frac{1}{2}\|g\|_{U^2}^4$$

where $g = (-1)^f$. Since $\|\cdot\|_{U^2}$ is a norm we see that (7.1) holds at least half the time for any function $f : \mathbb{F}_2^n \to \mathbb{F}_2$; it turns out if it holds an absolute proportion more of the time then f correlates with a linear polynomial.

Theorem 7.3 (U^2-inverse theorem)**.** *Suppose that $f : \mathbb{F}_2^n \to \mathbb{F}_2$ has $\|(-1)^f\|_{U^2} \geqslant \epsilon$. Then there is a linear polynomial $p : \mathbb{F}_2^n \to \mathbb{F}_2$ such that*

$$\langle (-1)^f, (-1)^p\rangle_{L^2(G)} \geqslant \epsilon^{O(1)}.$$

Proof. We put $g := (-1)^f$ and note that by hypotheses, the previous Lemma and Parseval's theorem we have

$$\epsilon^4 \leqslant \|g\|_{U^2}^4 = \|\widehat{g}\|_{\ell^4(\widehat{G})}^4 \leqslant \sup_{\gamma\in\widehat{G}}|\widehat{g}(\gamma)|^2\|\widehat{g}\|_{\ell^2(\widehat{G})}^2 = \sup_{\gamma\in\widehat{G}}|\widehat{g}(\gamma)|^2\|g\|_{L^2(G)}^2.$$

Of course $\|g\|_{L^2(G)}^2 = 1$ and so there is some character $\gamma \in \widehat{G}$ and phase $\sigma \in \{-1,1\}$ such that

$$\sigma\widehat{g}(\gamma) \geqslant \epsilon^2$$

Now, since γ is a character there is some $a \in \mathbb{F}_2^n$ such that $\gamma(x) = (-1)^{a\cdot x}$; since $\sigma \in \{-1, 1\}$ there is some $b \in \mathbb{F}_2$ such that $\sigma = (-1)^b$. It follows that the linear polynomial p defined by $p(x) = a \cdot x + b$ satisfies

$$\langle g, (-1)^p\rangle_{L^2(G)} = \langle g, \gamma.(-1)^b\rangle_{L^2(G)} = \sigma\widehat{g}(\gamma) \geqslant \epsilon^2,$$

and the result is proved. □

The point is that if we are given a black box into which we can input values of x and which outputs $f(x)$, then in a number of steps independent of the size of the underlying group, we can determine (with, say, 99% reliability) if $\|(-1)^f\|_{U^2}$ is large or not and hence whether f correlates with a linear polynomial.

7.4. **Characterising higher degree polynomials.** In the previous subsection we encoded the idea of f being linear by differencing (differentiating). If f is a linear polynomial, then for each fixed y, the map $x \mapsto f(x + y) - f(x)$ is constant and so, if we difference again we get that

$$(f(x + y + z) - f(x + z)) - (f(x + y) - f(x)) = 0_{\mathbb{F}_2}.$$

This can be rewritten as (7.1). With higher order polynomials we have to keep differencing, so that if f is a polynomial of degree d then

$$\sum_{\omega\in\{0,1\}^{d+1}} f(x + \omega \cdot h) = 0_{\mathbb{F}_2} \tag{7.2}$$

for all $x \in G, h \in G^{d+1}$. As with the previous section we define

$$\|g\|_{U^k}^{2^k} := \int \prod_{\omega\in\{0,1\}^k} g(x + \omega \cdot h) d\mu_G(x) \prod_{i=1}^{k} d\mu_G(h_i).$$

It is also helpful to formulate this inductively: we define a differencing operator on functions $g \in L^\infty(G)$ by

$$\partial_y(g)(x) := g(x + y)g(x),$$

and then note that

$$\|g\|_{U^k}^{2^k} = \int \prod_{\omega\in\{0,1\}^{k-1}} \partial_{h_k}(g)(x + \omega \cdot h) d\mu_G(x) \prod_{i=1}^{k} d\mu_G(h_i) = \int \|\partial_{h_k}(g)\|_{U^{k-1}}^{2^{k-1}} d\mu_G(h_k). \tag{7.3}$$

It turns out that $\|\cdot\|_{U^k}$ is a norm for $k \geqslant 2$, and to prove this we need an analogue of the Cauchy-Schwarz inequality. Given a family of functions $(f_\omega)_{\omega\in\{0,1\}^k}$ we define the Gowers inner product to be

$$\langle (f_\omega)_{\omega\in\{0,1\}^k}\rangle_{U^k} := \int \prod_{\omega\in\{0,1\}^k} f_\omega(x + \omega \cdot h) d\mu_G(x) \prod_{i=1}^{k} d\mu_G(h_i).$$

Lemma 7.5 (Gowers-Cauchy-Schwarz inequality)**.** *For all families of functions $(f_\omega)_{\omega\in\{0,1\}^k}$ we have*

$$|\langle (f_\omega)_{\omega\in\{0,1\}^k}\rangle_{U^k}| \leqslant \prod_{\omega\in\{0,1\}^k} \|f_\omega\|_{U^k}.$$

The proof here is notationally heavy; readers interested in an alternative source may wish to consult [TV06, p419] or [Gow01, Lemma 3.8].

Proof. Note that the inner product is equal to

$$\int\int \prod_{\omega\in\{0,1\}^{k-1}} f_{\omega,0}(x+\omega\cdot h)f_{\omega,1}(x+\omega\cdot h+h_k)d\mu_G(x)d\mu_G(h_k)d\mu_{G^{k-1}}(h).$$

Make the change of variables $y=x+h_k$ so that this is, in turn, equal to

$$\int \left(\int \prod_{\omega\in\{0,1\}^{k-1}} f_{\omega,0}(x+\omega\cdot h)d\mu_G(x)\right)\left(\int \prod_{\omega\in\{0,1\}^{k-1}} f_{\omega,1}(y+\omega\cdot h)d\mu_G(y)\right)d\mu_{G^{k-1}}(h).$$

By Cauchy-Schwarz this is at most

$$\left(\int \left(\int \prod_{\omega\in\{0,1\}^{k-1}} f_{\omega,0}(x+\omega\cdot h)d\mu_G(x)\right)^2 d\mu_{G^{k-1}}(h)\right)^{1/2}$$
$$\times \left(\int \left(\int \prod_{\omega\in\{0,1\}^{k-1}} f_{\omega,1}(y+\omega\cdot h)d\mu_G(y)\right)^2 d\mu_{G^{k-1}}(h)\right)^{1/2}.$$

Changing variables back we see that this is equal to

$$\langle (f_{\omega',0})_{\omega\in\{0,1\}^k}\rangle_{U^k}^{1/2}\langle (f_{\omega',1})_{\omega\in\{0,1\}^k}\rangle_{U^k}^{1/2}$$

where ω' is ω projected onto the first $k-1$ co-ordinates. By symmetry the same is true for all the other co-ordinates and so after k applications we get that

$$|\langle (f_\omega)_{\omega\in\{0,1\}^k}\rangle_{U^k}| \leqslant \prod_{\rho\in\{0,1\}^k} \langle (f_\rho)_{\omega\in\{0,1\}^k}\rangle_{U^k}^{1/2^k},$$

where $(f_\rho)_{\omega\in\{0,1\}^k}$ just means that all 2^k functions in the vector are the same function, f_ρ. □

Lemma 7.6. *We have the nesting property*

$$|\int f d\mu_G| = \|f\|_{U^1} \leqslant \|f\|_{U^2} \leqslant \ldots \leqslant \|f\|_{U^k} \leqslant \ldots,$$

and $\|\cdot\|_{U^k}$ is a norm for all $k\geqslant 2$.

Proof. The nesting follows immediately from the Gowers-Cauchy-Schwarz inequality: given $f \in L^\infty(G)$ write $f_\omega = f$ if $\omega_k = 1$ and $f_\omega \equiv 1$ if $\omega_k = 0$. Then

$$\|f\|_{U^{k-1}}^{2^{k-1}} = \langle (f_\omega)_{\omega\in\{0,1\}^k}\rangle_{U^k} \leqslant \|f\|_{U^k}^{2^{k-1}} \|1\|_{U^k}^{2^{k-1}} = \|f\|_{U^k}^{2^{k-1}}$$

and we are done.

It is immediate that $\|\cdot\|_{U^k}$ is homogenous and zero on the zero function. Since $\|\cdot\|_{U^2}$ is a norm by Lemma 7.2 we see that $\|f\|_{U^k} = 0$ must imply $f = 0$ for $k \geqslant 2$. It remains to check the triangle inequality: First note that

$$\|f_0 + f_1\|_{U^k}^{2^k} = \langle (f_0 + f_1)_{\omega\in\{0,1\}^k}\rangle_{U^k} = \sum_{I\subset\{0,1\}^k} \langle (f_{1_{\omega\in I}})_{\omega\in\{0,1\}^k}\rangle_{U^k},$$

where we are using the multi-linearity of the Gowers inner product. In particular the Gowers inner product takes 2^k terms indexed by $\omega \in \{0,1\}^k$ and the product $\langle (f_{1_{\omega\in I}})_{\omega\in\{0,1\}^k}\rangle_{U^k}$ simply denotes the Gowers inner product of some copies of f_0 and f_1, with f_0 in the position indexed by ω if $\omega \in \{0,1\}^k \setminus I$, and f_1 in the position indexed by ω if $\omega \in I$.

Now, by the Gowers-Cauchy-Schwarz inequality we see that

$$|\langle (f_{1_{\omega\in I}})_{\omega\in\{0,1\}^k}\rangle_{U^k}| \leqslant \|f_0\|_{U^k}^{2^k-|I|} \|f_1\|_{U^k}^{|I|},$$

and hence

$$\|f_0 + f_1\|_{U^k}^{2^k} \leqslant \sum_{I\subset\{0,1\}^k} \|f_0\|_{U^k}^{2^k-|I|} \|f_1\|_{U^k}^{|I|} = (\|f_0\|_{U^k} + \|f_1\|_{U^k})^{2^k}.$$

The inequality follows on taking 2^kth roots, and the lemma is proved. □

Finally, as before, the proportion of tuples satisfying (7.2) is

$$P_k(f) = \frac{1}{2} + \frac{1}{2}\|(-1)^f\|_{U^k}^{2^k},$$

which again means that (7.2) holds at least half the time, and so again we should like to show that $\|(-1)^f\|_{U^k} \geqslant \epsilon$ implies correlation with a degree k polynomial. There are subtleties to this request as it is stated and we shall only be able to deal with the quadratic case here.

7.7. **Correlating with quadratic polynomials.** Our object here is to prove the following theorem due to Samorodnitsky [Sam07]. In other groups the U^3-inverse theorem is more complicated to state and is due to Green and Tao [GT08].

Theorem 7.8 (U^3-inverse theorem)**.** *Suppose that $f : \mathbb{F}_2^n \to \mathbb{F}_2$ has $\|(-1)^f\|_{U^3} \geqslant \epsilon$. Then there is a quadratic polynomial $p : \mathbb{F}_2^n \to \mathbb{F}_2$ such that*

$$\langle (-1)^f, (-1)^p\rangle_{L^2(G)} \geqslant \exp(-O(\epsilon^{-O(1)})).$$

If $f(x)$ is a quadratic polynomial then it has the form

$$x \mapsto \langle Ax, x\rangle + \langle a, x\rangle + b$$

were A is an $\mathbb{F}_2$-valued matrix, $a \in \mathbb{F}_2^n$ and $b \in \mathbb{F}_2$. By symmetry and since the diagonal term $\langle a, x\rangle$ can be absorbed into A by replacing A_{ii} by $A_{ii}+a_i$, we may assume that $a = 0_G$ and A is upper triangular. (Note that $x_i^2 = x_i$ in $\mathbb{F}_2$.)

Differencing once we have

$$f(x+y) - f(x) = \langle (A+A^t)y, x\rangle + \langle Ay, y\rangle,$$

which is a linear polynomial in x for fixed y. It follows that

$$(f(x+y+z) - f(x+z)) - (f(x+y) - f(x)) = \langle (A+A^t)y, z\rangle$$

which is constant in x for fixed y and z. Crucially $A + A^t$ is symmetric and has zero diagonal; we shall now set about establishing a converse.

For the remainder of this section it will be convenient to identify G with $\widehat{G}$ via the map

$$r \mapsto (x \mapsto (-1)^{r\cdot x}).$$

Lemma 7.9. *Suppose that $f : \mathbb{F}_2^n \to \mathbb{F}_2$, S is a symmetric matrix with zero diagonal and $g := (-1)^f$ has*

$$\int \widehat{\partial_y g}(Sy)^2 d\mu_G(y) \geqslant \delta.$$

Then there is a quadratic polynomial p such that $\langle (-1)^f, (-1)^p\rangle_{L^2(G)} \geqslant \delta^{O(1)}$.

Proof. Let A be a matrix such that $A + A^t = S$, and consider $h(x) := (-1)^{\langle Ax, x\rangle}$. By our earlier calculation we have

$$\partial_y h(x) = (-1)^{\langle (A+A^t)y, x\rangle + \langle Ay, y\rangle},$$

and so

$$\langle \partial_y g, \partial_y h\rangle_{L^2(G)} = (-1)^{\langle Ay, y\rangle}\widehat{\partial_y g}((A+A^t)y) = (-1)^{\langle Ay, y\rangle}\widehat{\partial_y g}(Sy).$$

It follows that

$$\int \langle \partial_y g, \partial_y h\rangle_{L^2(G)}^2 d\mu_G(y) \geqslant \delta.$$

On the other hand the integral is equal to

$$\int g(x)g(x+y)g(z)g(z+y)h(x)h(x+y)h(z)h(z+y)d\mu_G(x)d\mu_G(z)d\mu_G(y),$$

which is in turn equal to

$$\int (gh)(x)(gh)(x+y)(gh)(z)(gh)(z+y)d\mu_G(x)d\mu_G(z)d\mu_G(y) = \|gh\|_{U^2}^4.$$

It follows by the U^2-inverse theorem that there is some $a \in \mathbb{F}_2^n$ and $b \in \mathbb{F}_2$ such that $l(x) := \langle a, x\rangle + b$ has

$$\langle gh, (-1)^{l(x)}\rangle_{L^2(G)} \geqslant \delta^{O(1)}.$$

The result follows on putting $p(x) = \langle Ax, x\rangle + \langle a, x\rangle + b$. □

Our problem now is to find a suitable linear map, and to do that we shall find a roughly linear choice function.

Lemma 7.10. *Suppose that $g : G \to \{-1,1\}$ has $\|g\|_{U^3} \geqslant \epsilon$. Then there is a function $\phi : G \to \widehat{G}$ and a set A of density at least $\epsilon^{O(1)}$ such that $\widehat{\partial_x g}(\phi(x))^2 \geqslant \epsilon^{O(1)}$ for all $x \in A$ and*

$$\mu_{G^2}(\{(x,y) \in G^2 : \phi(x) + \phi(y) = \phi(x+y), x, y, x+y \in A\}) \geqslant \epsilon^{O(1)}.$$

Proof. Pick a function $\phi : G \to \widehat{G}$ randomly with

$$\mathbb{P}(\phi(x) = \gamma) = \widehat{\partial_x g}(\gamma)^2,$$

such that the choices are independent for distinct xs. Note that

$$\sum_{\gamma \in \widehat{G}} \mathbb{P}(\phi(x) = \gamma) = \sum_{\gamma \in \widehat{G}} \widehat{\partial_x g}(\gamma)^2 = g^2 * g^2(x) = 1$$

by Parseval's theorem so these are genuine probability distributions.

Now, write

$$A(\phi) := \{x \in G : \widehat{\partial_x g}(\phi(x))^2 \geqslant \epsilon^{16}/6\},$$

and

$$L(\phi) := \mu_{G^2}(\{(x,y) \in G^2 : \phi(x) + \phi(y) = \phi(x+y), x, y, x+y \in A\}).$$

Then

$$\begin{aligned}
\mathbb{E}L(\phi) &= \int_{x,y,x+y \in A(\phi)} \sum_{\gamma,\gamma'} \widehat{\partial_x g}(\gamma)^2 \widehat{\partial_y g}(\gamma')^2 \widehat{\partial_{x+y} g}(\gamma+\gamma')^2 d\mu_G(x) d\mu_G(y) \\
&\geqslant \int_{x,y,x+y} \sum_{\gamma,\gamma'} \widehat{\partial_x g}(\gamma)^2 \widehat{\partial_y g}(\gamma')^2 \widehat{\partial_{x+y} g}(\gamma+\gamma')^2 d\mu_G(x) d\mu_G(y) \\
&\quad -3.\frac{\epsilon^{16}}{6} \int_{x,y} \sum_{\gamma,\gamma'} \widehat{\partial_x g}(\gamma)^2 \widehat{\partial_y g}(\gamma')^2 d\mu_G(x) d\mu_G(y).
\end{aligned}$$

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

$$\int \sum_{\gamma \in \widehat{G}} \widehat{\partial_z g}(\gamma)^2 d\mu_G(z) = \int \int \partial_z(g)(x)^2 d\mu_G(x) d\mu_G(z) = \|g\|_{L^2(G)}^4 = 1, \tag{7.4}$$

whence

$$\mathbb{E}L(\phi) \geqslant \int \sum_{\gamma,\gamma'} \widehat{\partial_x g}(\gamma)^2 \widehat{\partial_y g}(\gamma')^2 \widehat{\partial_{x+y} g}(\gamma+\gamma')^2 d\mu_G(x) d\mu_G(y) - \epsilon^{16}/2.$$

It follows that if we can show that

$$T := \int \sum_{\gamma,\gamma'} \widehat{\partial_x g}(\gamma)^2 \widehat{\partial_y g}(\gamma')^2 \widehat{\partial_{x+y} g}(\gamma+\gamma')^2 d\mu_G(x) d\mu_G(y) \geqslant \epsilon^{16}, \tag{7.5}$$

then by averaging we can pick a ϕ such that $L(\phi) \geqslant \epsilon^{16}/2$. In particular $\mu_G(A(\phi))^2 \geqslant L(\phi)$ so $\mu_G(A(\phi)) \geqslant \epsilon^8/\sqrt{2}$, and so, putting $A := A(\phi)$ we have

$$A \subset \{x \in G : \widehat{\partial_x g}(\phi(x))^2 \geqslant \epsilon^{16}/6\}$$

and

$$\mu_{G^2}(\{(x,y)\in G^2 : \phi(x)+\phi(y)=\phi(x+y), x, y, x+y\in A\}) \geqslant \epsilon^{16}/2.$$

It will follow that we are done, and it remains to establish (7.5). By the definition of the Fourier transform

$$\widehat{\partial_w g}(\lambda)^2 = \int (\partial_w g) * (\partial_w g)(z)\lambda(z)d\mu_G(z) \text{ for all } w\in G, \lambda\in\widehat{G},$$

which, inserted in place of all the Fourier transforms in (7.5) gives

$$T = \int\int (\partial_x g) * (\partial_x g)(z)(\partial_y g) * (\partial_y g)(z)(\partial_{x+y} g) * (\partial_{x+y} g)(z)d\mu_G(z)d\mu_G(x)d\mu_G(y).$$

Now, note that

$$\partial_x g * \partial_x g(z) = \partial_z g * \partial_z g(x) \text{ for all } x, y\in G, \tag{7.6}$$

so writing h_z for the function $(\partial_z g) * (\partial_z g)$ we get

$$T = \int\int h_z(x)h_z(y)h_z(x+y)d\mu_G(z)d\mu_G(x)d\mu_G(y) = \int \sum_{\gamma\in\widehat{G}} \widehat{h}_z(\gamma)^3 d\mu_G(z),$$

where the equality is by Fourier inversion. We conclude that

$$\int \sum_{\gamma,\gamma'} \widehat{\partial_x g}(\gamma)^2\widehat{\partial_y g}(\gamma')^2\widehat{\partial_{x+y} g}(\gamma+\gamma')^2 d\mu_G(x)d\mu_G(y) = \int \sum_{\gamma\in\widehat{G}} \widehat{\partial_z g}(\gamma)^6 d\mu_G(z),$$

an identity called Samorodnitzky's identity. By Cauchy-Schwarz applied to $\widehat{\partial_z g}(\gamma)^3\times\widehat{\partial_z g}(\gamma)$ (or log-covexity of L^p-norms) we have

$$\left(\int \sum_{\gamma\in\widehat{G}} \widehat{\partial_z g}(\gamma)^2 d\mu_G(z)\right)^{1/2} \left(\int \sum_{\gamma\in\widehat{G}} \widehat{\partial_z g}(\gamma)^6 d\mu_G(z)\right)^{1/2} \geqslant \int \sum_{\gamma\in\widehat{G}} \widehat{\partial_z g}(\gamma)^4 d\mu_G(z).$$

Of course, the term on the right is equal to $\|g\|_{U^3}^8 \geqslant \epsilon^8$ by the inductive definition of the U^3-norm and Lemma 7.2, and so inserting (7.4) into the first term on the left we get

$$\int \sum_{\gamma,\gamma'} \widehat{\partial_x g}(\gamma)^2\widehat{\partial_y g}(\gamma')^2\widehat{\partial_{x+y} g}(\gamma+\gamma')^2 d\mu_G(x)d\mu_G(y) \geqslant \epsilon^{16},$$

and the result is proved. □

We should like to couple the previous lemma with the Rough Morphism Theorem, but the set A presents a problem. It could be that an application of the theorem gives us a morphism which coincides with ϕ on a set disjoint from A. The proof of the Rough Morphism Theorem can be adapted to ensure that the coincidences happen on A; we shall establish this strengthening as a consequence.

We shall take ϕ to be highly random on the complement of A which will force any correlation with a morphism to be largely on A. The following trivial counting argument will let us do this.

Lemma 7.11. *There is a function $\nu : G \to \widehat{G}$ such that for any morphism $\theta : G \to \widehat{G}$ we have*

$$|\{x : \nu(x) = \theta(x)\}| = O(n^2).$$

Proof. Write $N := 2^n$ and note that there are N^N functions $G \to \widehat{G}$. On the other hand there are at most N^n morphisms $G \to \widehat{G}$, and hence at most

$$N^n.\binom{N}{N-m}.N^{N-m} = \binom{N}{m}.N^{N+n-m} \leqslant \left(\frac{eN}{m}\right)^m N^{N+n-m} = \frac{e^m N^{N+n}}{m^m} =: M$$

functions $G \to \widehat{G}$ which differ from a morphism in at most $N-m$ places. Thus, if $M < N^N$ then there is a function which differs from every morphism in at least $N - m$ places. We can take $m = O(n^2)$ to guarantee this and the result is proved. □

We could actually have taken $m = O(n^2/\log n)$, but this slight improvement is not important.

Lemma 7.12. *Suppose that $g : G \to \{-1, 1\}$ has $\|g\|_{U^3} \geqslant \epsilon$. Then there is a symmetric matrix S with zero diagonal such that*

$$\int \widehat{\partial_x g}(Sx)^2 d\mu_G(x) \geqslant \exp(-O(\epsilon^{-O(1)})).$$

Proof. Apply Lemma 7.10 to get a function ϕ and set A and let $\tilde{\phi}$ be equal to ϕ on A, and equal to ν (from Lemma 7.11) on A^c. By the Rough Morphism theorem there is a morphism θ such that

$$\mu_G(x \in G : \theta(x) = \tilde{\phi}(x)) \geqslant \exp(-O(\epsilon^{-O(1)})).$$

It follows that either $\exp(-O(\epsilon^{-O(1)})) = 2^{-n}.O(n^2)$, or else

$$\mu_G(x \in A : \theta(x) = \phi(x) \geqslant \exp(-O(\epsilon^{-O(1)})).$$

In the first case the conclusion is trivial by taking $S \equiv 0$, since

$$\widehat{\partial_{0_G} g}(S0_G) = \widehat{\partial_{0_G} g}(0_{\widehat{G}}) = g * g(0_G) = \|g\|^2_{L^2(G)} = 1,$$

and so the integral is at least 2^{-n} which has the required size in this case. Thus we assume we are in the second case.

We write M for the matrix corresponding to θ under the identification $r \mapsto (x \mapsto (-1)^{r \cdot x})$ of G and $\widehat{G}$, so that

$$\int \widehat{\partial_x g}(Mx)^2 d\mu_G(x) \geqslant \exp(-O(\epsilon^{-O(1)})).$$

Write $h(x) := (-1)^{\langle Mx, x\rangle}$ and suppose that k is any function.

Claim.

$$\int k(x)\widehat{\partial_x g}(Mx)^2 d\mu_G(x) = \sum_r \widehat{k}(r) \int \widehat{\partial_y g}(M^t y + r)^2 d\mu_G(y)$$

Proof. This is a calculation following from the Fourier inversion formula for k and (7.6). The left hand side is equal to

$$\begin{aligned}
&\int \sum_r \widehat{k}(r)(-1)^{\langle r,x\rangle} \int (\partial_x g) * (\partial_x g)(y)(-1)^{\langle Mx,y\rangle} d\mu_G(y) d\mu_G(x)\\
&= \int \sum_r \widehat{k}(r) \int (\partial_y g) * (\partial_y g)(x)(-1)^{\langle Mx,y\rangle}(-1)^{\langle r,x\rangle} d\mu_G(y) d\mu_G(x)\\
&= \int \sum_r \widehat{k}(r) \int (\partial_y g) * (\partial_y g)(x)(-1)^{\langle x, M^t y + r\rangle} d\mu_G(y) d\mu_G(x)\\
&= \sum_r \widehat{k}(r) \int \widehat{\partial_y g}(M^t y + r)^2 d\mu_G(y).
\end{aligned}$$

□

Now, apply the claim with $k = h * h$ to get

$$\int h * h(x)\widehat{\partial_x g}(Mx)^2 d\mu_G(x) = \sum_r \widehat{h}(r)^2 \int \widehat{\partial_x g}(M^t x + r)^2 d\mu_G(x). \tag{7.7}$$

On the other hand

$$\sum_r \int \widehat{\partial_x g}(M^t x + r)^2 d\mu_G(x) = \sum_{\gamma \in \widehat{G}} \widehat{\partial_x g}(\gamma)^2 = g^2 * g^2(x) = 1,$$

and so the measure $\mathbb{P}$ defined by

$$\mathbb{P}(R) := \sum_{r \in R} \int \widehat{\partial_x g}(M^t x + r)^2 d\mu_G(x)$$

is a probability measure. Hence we can apply Cauchy-Schwarz in (7.7) to get that

$$\begin{aligned}
\sum_r \widehat{h}(r)^2 \int \widehat{\partial_x g}(M^t x + r)^2 d\mu_G(x) &= \mathbb{E}(\widehat{h}(r)^2)\\
&\geqslant (\mathbb{E}\widehat{h}(r))^2 = \left(\sum_r \widehat{h}(r) \int \widehat{\partial_x g}(M^t x + r)^2 d\mu_G(x)\right)^2,
\end{aligned}$$

where $\mathbb{E}$ denotes integration against the probability measure $\mathbb{P}$.

What we have shown, then, is that

$$\begin{aligned}
\int h * h(x)\widehat{\partial_x g}(Mx)^2 d\mu_G(x) &\geqslant \left(\sum_r \widehat{h}(r) \int \widehat{\partial_x g}(M^t x + r)^2 d\mu_G(x)\right)^2\\
&= \left(\int h(x)\widehat{\partial_x g}(Mx)^2 d\mu_G(x)\right)^2,
\end{aligned}$$

where the last equality is from the claim taken with $k = h$.

Now note that

$$1_{\{y:M^ty=My\}}(x) = h*h(x)h(x) \geqslant h*h(x),$$

so it follows that

$$\begin{aligned}\int_{M^tx=Mx} \widehat{\partial_x g}(Mx)^2 d\mu_G(x) &\geqslant \int h*h(x)\widehat{\partial_x g}(Mx)^2 d\mu_G(x)\\ &\geqslant \left(\int h(x)\widehat{\partial_x g}(Mx)^2 d\mu_G(x)\right)^2.\end{aligned}$$

To lower bound the right hand side we have the following claim.

Claim.

$$\widehat{\partial_x g}(r) = 0 \text{ } \textit{unless } \langle r, x\rangle = 0_{\mathbb{F}_2}. \tag{7.8}$$

Proof. By definition we have

$$\begin{aligned}\widehat{\partial_x g}(r) &= \int g(y)g(x+y)(-1)^{\langle r,y\rangle} d\mu_G(y)\\ &= (-1)^{\langle r,x\rangle}\int g(z+x)g(z)(-1)^{\langle r,z\rangle} d\mu_G(z) = (-1)^{\langle r,x\rangle}\widehat{\partial_x g}(r),\end{aligned}$$

and the claim follows. □

In particular, $h(x) = 1$ if $\widehat{\partial_x g}(Mx) \neq 0$ and so

$$\int h(x)\widehat{\partial_x g}(Mx)^2 d\mu_G(x) = \int \widehat{\partial_x g}(Mx)^2 d\mu_G(x) \geqslant \exp(-O(\epsilon^{-O(1)})).$$

Let U be the space generated by $\{x : M^tx = Mx\}$, and let $Bx = Mx$ for all $x \in U$. B is symmetric by construction and can be extended to the whole space while preserving this symmetry, whence

$$\int \widehat{\partial_x g}(Bx)^2 d\mu_G(x) \geqslant \exp(-O(\epsilon^{-O(1)})).$$

Finally, note that $\widehat{\partial_x g}(Bx) = 0$ unless $\langle Bx, x\rangle = 0_{\mathbb{F}_2}$ by (7.8), but since B is symmetric $\langle Bx, x\rangle = \langle r, x\rangle$, where r is the vector on the diagonal of B, whence

$$\int_{\langle r,x\rangle = 0_{\mathbb{F}_2}} \widehat{\partial_x g}(Bx)^2 d\mu_G(x) \geqslant \exp(-O(\epsilon^{-O(1)})).$$

We now define $Sx = Bx + \langle x, r\rangle r$, which is symmetric and $Sx = Bx$ if $\langle r, x\rangle = 0_{\mathbb{F}_2}$. Finally S has zero diagonal: write $r'$for the diagonal of S, so that $\langle r', x\rangle = \langle Sx, x\rangle$ for all x. So

$$\langle r', x\rangle = \langle Sx, x\rangle = \langle Bx, x\rangle + \langle x, r\rangle^2 = \langle x, r\rangle + \langle x, r\rangle^2 = 0_{\mathbb{F}_2} \text{ for all } x \in G,$$

and hence $r' = 0_G$. The result is proved. □

The argument for converting the matrix M into the symmetric matrix S with zero diagonal is, for obvious reasons, called the symmetrisation argument.

Proof of Theorem 7.8. This follows immediately on combining Lemma 7.12 with Lemma 7.9. □

Acknowledgement

The author should like to thank all those who have provided feedback, comments and corrections, and particularly the students who took the course in the academic year 2010–2011.

Appendix A. The integral triangle inequality

In this brief section we shall prove the integral triangle inequality, more commonly referred to as the integral Minkowski inequality which is used in the proof of Beckner's inequality (Theorem 4.2).

Lemma A.1 (Integral triangle inequality)**.** *Suppose X and Y are finite sets, $q \in [1,\infty]$ and $f : X \times Y \to \mathbb{C}$. Then we have*

$$\left(\int \left(\int |f(x,y)|d\mu_Y(y)\right)^q d\mu_X(x)\right)^{1/q} \leqslant \int \left(\int |f(x,y)|^q d\mu_X(x)\right)^{1/q} d\mu_Y(y).$$

Proof. Define auxiliary functions g_y by $g_y(x) := |f(x,y)|$, and then note that the inequality is simply the statement

$$\| \int g_y d\mu_Y(y)\|_{L^q(X)} \leqslant \int \|g_y\|_{L^q(X)} d\mu_Y(y).$$

On the other hand since Y is finite this is just a weighted sum and hence follows from the usual triangle inequality for $L^q(X)$. □

References

[Bon70] A. Bonami. Étude des coefficients de Fourier des fonctions de $L^p(G)$. *Ann. Inst. Fourier (Grenoble)*, 20(fasc. 2):335–402 (1971), 1970.

[BS94] A. Balog and E. Szemerédi. A statistical theorem of set addition. *Combinatorica*, 14(3):263–268, 1994.

[Cha02] M.-C. Chang. A polynomial bound in Freĭman's theorem. *Duke Math. J.*, 113(3):399–419, 2002.

[Fox11] J. Fox. A new proof of the graph removal lemma. *Ann. of Math. (2)*, 174(1):561–579, 2011, arXiv:1006.1300.

[Fre73] G. A. Freĭman. *Foundations of a structural theory of set addition*. American Mathematical Society, Providence, R. I., 1973. Translated from the Russian, Translations of Mathematical Monographs, Vol 37.

[Gow98] W. T. Gowers. A new proof of Szemerédi's theorem for arithmetic progressions of length four. *Geom. Funct. Anal.*, 8(3):529–551, 1998.

[Gow01] W. T. Gowers. A new proof of Szemerédi's theorem. *Geom. Funct. Anal.*, 11(3):465–588, 2001.

[Gre02a] B. J. Green. Arithmetic progressions in sumsets. *Geom. Funct. Anal.*, 12(3):584–597, 2002.

[Gre02b] B. J. Green. Restriction and Kakeya phenomena. Available at `www.dpmms.cam.ac.uk/~bjg23`, 2002.

[Gre05] B. J. Green. A Szemerédi-type regularity lemma in abelian groups, with applications. *Geom. Funct. Anal.*, 15(2):340–376, 2005.

[GT08] B. J. Green and T. C. Tao. An inverse theorem for the Gowers $U^3(G)$ norm. *Proc. Edinb. Math. Soc. (2)*, 51(1):73–153, 2008.

[GT09] B. J. Green and T. C. Tao. Freĭman's theorem in finite fields via extremal set theory. *Combin. Probab. Comput.*, 18(3):335–355, 2009.

[HB87] D. R. Heath-Brown. Integer sets containing no arithmetic progressions. *J. London Math. Soc. (2)*, 35(3):385–394, 1987.

[KKL88] J. Kahn, G. Kalai, and N. Linial. The influence of variables on Boolean functions. In *Proceedings of the 29th Annual Symposium on Foundations of Computer Science*, pages 68–80, Washington, DC, USA, 1988. IEEE Computer Society.

[Nel73] E. Nelson. The free Markoff field. *J. Functional Analysis*, 12:211–227, 1973.

[Rot52] K. F. Roth. Sur quelques ensembles d'entiers. *C. R. Acad. Sci. Paris*, 234:388–390, 1952.

[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.

[Ruz99] I. Z. Ruzsa. An analog of Freĭman's theorem in groups. *Astérisque*, (258):xv, 323–326, 1999. Structure theory of set addition.

[Sam07] A. Samorodnitsky. Low-degree tests at large distances. In *STOC'07—Proceedings of the 39th Annual ACM Symposium on Theory of Computing*, pages 506–515. ACM, New York, 2007.

[SSV05] B. Sudakov, E. Szemerédi, and V. H. Vu. On a question of Erdős and Moser. *Duke Math. J.*, 129(1):129–155, 2005.

[Sze90] E. Szemerédi. Integer sets containing no arithmetic progressions. *Acta Math. Hungar.*, 56(1-2):155–158, 1990.

[Tao08] T. C. Tao. Product set estimates for non-commutative groups. *Combinatorica*, 28(5):547–594, 2008.

[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