What Is Quadratic Fourier Analysis?

W. T. Gowers

University of Cambridge

April 9, 2008

Szemerédi's theorem

Theorem

For every $\delta > 0$ and every positive integer k there exists N such that every subset $A \subset \{1, 2, \ldots, N\}$ of cardinality at least δN contains an arithmetic progression of length k.

An equivalent formulation of Szemerédi's theorem

Let $\mathbb{Z}_N$ stand for $\mathbb{Z}/N\mathbb{Z}$.

Theorem

For every $\delta > 0$ and every positive integer k there exists $c = c(\delta, k) > 0$ such that if f is any function from $\mathbb{Z}_N$ to $[0,1]$ and $\mathbb{E}_x f(x) \geq \delta$ then

$$\mathbb{E}_{x,d} f(x) f(x+d) f(x+2d) \dots f(x+(k-1)d) \geq c.$$

If $A \subset \mathbb{Z}_N$ and $f = \chi_A$ then the above expectation is closely related to the number of arithmetic progressions of length k in A.

Discrete Fourier analysis

The *discrete Fourier transform* for functions $f : \mathbb{Z}_N \to \mathbb{C}$ is defined by the following formula:

$$\hat{f}(r) = \mathbb{E}_x f(x) e_N(-rx),$$

where $e_N(y)$ stands for $\exp 2\pi i y/N$.

Compare with the formula for functions defined on [0,1]:

$$\hat{f}(\alpha) = \int f(x) e(-\alpha x) dx,$$

where $e(y)$ stands for $\exp(2\pi i y)$.

Basic facts about the discrete Fourier transform.

$\langle \hat{f}, \hat{g} \rangle = \langle f, g \rangle$ *Parseval's identity*

$\widehat{f * g}(r) = \hat{f}(r)\hat{g}(r)$ *Convolution identity*

Here, $\langle f, g \rangle$ means $\mathbb{E}_x f(x)\overline{g(x)}$, $\langle \hat{f}, \hat{g} \rangle$ means $\sum_r \hat{f}(r)\overline{\hat{g}(r)}$, and $f * g(u)$ is defined to be $\mathbb{E}_{x+y=u} f(x)g(y)$.

Important observation:

If $\|\hat{f} - \hat{g}\|_\infty$ is small then so is

$$\mathbb{E}_{x,d} f(x)f(x+d)f(x+2d) - \mathbb{E}_{x,d} g(x)g(x+d)g(x+2d)$$

Here we are assuming that both f and g are functions from $\mathbb{Z}_N$ to $[0,1]$.

Sketch of proof (1)

We can rewrite the expression as a sum of three terms as follows:

$$\begin{aligned}&\mathbb{E}_{x,d}(f(x)-g(x))f(x+d)f(x+2d)\\&+\mathbb{E}_{x,d}g(x)(f(x+d)-g(x+d))f(x+2d)\\&+\mathbb{E}_{x,d}g(x)g(x+d)(f(x+2d)-g(x+2d))\end{aligned}$$

Each one of these terms involves the function $f-g$.

Sketch of proof (2)

$$\begin{aligned}\mathbb{E}_{x,d} f(x)g(x+d)h(x+2d) &= \mathbb{E}_{x+z=2y} f(x)h(z)g(y) \\ &= \langle f * h, g_2\rangle\end{aligned}$$

where $g_2(y) = \overline{g(y/2)}$.

By Parseval + convolution this equals

$$\langle \hat{f}\hat{h}, \hat{g_2}\rangle = \sum_r \hat{f}(r)\hat{h}(r)\hat{g}(-2r),$$

which is, for example, at most

$$\|\hat{g}\|_\infty \sum_r |\hat{f}(r)||\hat{h}(r)| \leq \|\hat{g}\|_\infty \|\hat{f}\|_2 \|\hat{g}\|_2 \leq \|\hat{g}\|_\infty.$$

A second important observation

If $f : \mathbb{Z}_N \to [-1,1]$ *then* $\|\hat{f}\|_\infty \leq \|\hat{f}\|_4 \leq \|\hat{f}\|_\infty^{1/2}$.

In other words, $\|\hat{f}\|_\infty$ is "roughly equivalent" to $\|\hat{f}\|_4$.

Proof.

$\max_r |\hat{f}(r)|^4 \leq \sum_r |\hat{f}(r)|^4 \leq (\max_r |\hat{f}(r)|^2) \sum_r |\hat{f}(r)|^2$. □

This is useful because $\|\hat{f}\|_4$ has a *non-Fourier interpretation.*

The non-Fourier interpretation of $\|\hat{f}\|_4$.

$$\begin{aligned}\|\hat{f}\|_4^4 &= \sum_r |\hat{f}(r)|^4 = \langle \hat{f}^2, \hat{f}^2\rangle = \langle f * f, f * f\rangle \\ &= \mathbb{E}_{x+y=z+w} f(x)f(y)\overline{f(z)f(w)} \\ &= \mathbb{E}_{x,a,b} f(x)\overline{f(x+a)f(x+b)}f(x+a+b).\end{aligned}$$

This is useful because it generalizes.

$$\|f * g\|_2^2 = \langle f * g, f * g\rangle = \langle f * f^*, g * g^*\rangle \leq \|f * f^*\|_2 \|g * g^*\|_2,$$

where $f^*(x) = \overline{f(-x)}$. But

$$\|f * f^*\|_2^2 = \langle f * f^*, f * f^*\rangle = \langle f * f, f * f\rangle = \|f * f\|_2^2$$

and similarly for g. So

$$\|f * g\|_2^2 \leq \|f * f\|_2 \|g * g\|_2$$

Therefore,

$$\begin{aligned} \mathbb{E}_{x+z=2y} f(x)h(z)g_2(y) &= \langle f * h, g_2\rangle \\ &\leq \|f * h\|_2 \|g_2\|_2 \\ &\leq \|f * f\|_2^{1/2} \|h * h\|_2^{1/2} \end{aligned}$$

The morals of the previous slide

With an eye to later generalizations, let us write $\|f\|_{U^2}$ for the norm defined by the formula

$$\|f\|_{U^2}^4 = \|f * f\|^2 = \mathbb{E}_{x,a,b} f(x)\overline{f(x+a)f(x+b)}f(x+a+b).$$

The following facts summarize why the U^2 norm is important.

The morals of the previous slide

With an eye to later generalizations, let us write $\|f\|_{U^2}$ for the norm defined by the formula

$$\|f\|_{U^2}^4 = \|f * f\|^2 = \mathbb{E}_{x,a,b} f(x)\overline{f(x+a)f(x+b)}f(x+a+b).$$

The following facts summarize why the U^2 norm is important.

- The expression $\mathbb{E}_{x,d} f(x)f(x+d)f(x+2d)$ is approximately invariant under small perturbations of f in the U^2 norm.

The morals of the previous slide

With an eye to later generalizations, let us write $\|f\|_{U^2}$ for the norm defined by the formula

$$\|f\|_{U^2}^4 = \|f * f\|^2 = \mathbb{E}_{x,a,b} f(x)\overline{f(x+a)f(x+b)}f(x+a+b).$$

The following facts summarize why the U^2 norm is important.

- The expression $\mathbb{E}_{x,d} f(x)f(x+d)f(x+2d)$ is approximately invariant under small perturbations of f in the U^2 norm.
- In particular, if f is close in the U^2 norm to the constant function $\delta\mathbf{1}$, then $\mathbb{E}_{x,d} f(x)f(x+d)f(x+2d) \approx \delta^3$.

The morals of the previous slide

With an eye to later generalizations, let us write $\|f\|_{U^2}$ for the norm defined by the formula

$$\|f\|_{U^2}^4 = \|f * f\|^2 = \mathbb{E}_{x,a,b} f(x)\overline{f(x+a)f(x+b)}f(x+a+b).$$

The following facts summarize why the U^2 norm is important.

- The expression $\mathbb{E}_{x,d} f(x)f(x+d)f(x+2d)$ is approximately invariant under small perturbations of f in the U^2 norm.
- In particular, if f is close in the U^2 norm to the constant function $\delta\mathbf{1}$, then $\mathbb{E}_{x,d} f(x)f(x+d)f(x+2d) \approx \delta^3$.
- The definition of the U^2 norm, and basic facts about it (including the fact that it is a norm) can be given without the help of Fourier analysis.

A simple inverse theorem for the U^2 norm.

Theorem

If $f : \mathbb{Z}_N \to [-1, 1]$ and $\|f\|_{U^2} \geq c$, then there is some r such that $|\mathbb{E}_x f(x) e_N(-rx)| \geq c^2$.

Proof.

As we have seen, $\|\hat{f}\|_\infty \geq \|\hat{f}\|_4^2 = \|f\|_{U^2}^2$. □

Many proofs in arithmetic combinatorics rely on dichotomies, of which a typical one is the following.

Either $\|f\|_{U^2}$ is small (in which case use one argument) or f correlates significantly with a trignometric function (in which case use another).

In particular, this dichotomy underlies Roth's proof of Szemerédi's theorem for progressions of length 3.

Enter quadratic functions.

The earlier results hold also when $f : \mathbb{Z}_N \to \mathbb{C}$ with $\|f\|_\infty \leq 1$. Now observe that the identity

$$x^2 - 3(x+d)^2 + 3(x+2d)^2 - (x+3d)^2 = 0$$

implies that if $f(x) = e_N(x^2)$ and $g(x) = e_N(3x^2)$, then

$$\mathbb{E}_{x,d} f(x)\overline{g(x+d)}g(x+2d)\overline{f(x+3d)} = 1.$$

But

$$\begin{aligned} \|f\|_{U^2}^4 &= \mathbb{E}_{x,a,b} e_N(x^2 - (x+a)^2 - (x+b)^2 + (x+a+b)^2) \\ &= \mathbb{E}_{x,a,b} e_N(2ab) \\ &= O(N^{-1}) \end{aligned}$$

The moral of the previous example

In that example, $\|f\|_{U^2}$ was tiny, but if we replace f by 0 in the expression

$$\mathbb{E}_{x,d} f(x)\overline{g(x+d)}g(x+2d)\overline{f(x+3d)},$$

then we get a completely different answer.

Therefore, expressions of the form

$$\mathbb{E}_{x,d} f_1(x)f_2(x+d)f_3(x+2d)f_4(x+3d)$$

are not robust when subjected to U^2 perturbations. Moreover, the simplest examples that show this have a quadratic character.

The above fact is (one manifestation of) the reason that Szemerédi's theorem for progressions of length 3 is significantly easier than the general case.

Generalizing the U^2 norm.

It turns out that a simple generalization will do instead: the U^3 norm. This is defined by the formula

$$\begin{aligned}\|f\|_{U^3}^8 \quad = \quad & \mathbb{E}_{x,a,b,c} f(x)\overline{f(x+a)f(x+b)}f(x+a+b) \\ & \overline{f(x+c)}f(x+a+c)f(x+b+c)\overline{f(x+a+b+c)}\end{aligned}$$

A reasonably straightforward argument can be used to show that if f_1, f_2, f_3, f_4 are functions from $\mathbb{Z}_N$ to $\mathbb{C}$ and $\|f_i\|_\infty \leq 1$ for each i, then the expression

$$\mathbb{E}_{x,d} f_1(x) f_2(x+d) f_3(x+2d) f_4(x+3d)$$

is not sensitive to small U^3 perturbations. What's more, this generalizes in an obvious way to progressions of length k and the U^{k-1} norm.

One major difficulty: completing the square

The density of progressions of length 3 is robust when subjected to small U^2 perturbations.

The density of progressions of length 4 is robust when subjected to small U^3 perturbations

If f does not have small U^2 norm then f correlates significantly with a trigonometric function.

If f does not have small U^2 norm then f correlates significantly with ???

A natural conjecture

A trigonometric function on $\mathbb{Z}_N$ is a function f such that

$$f(x)\overline{f(x+a)f(x+b)}f(x+a+b)=1$$

for every x, a, b. Equivalently, it is a function such that

$$\frac{f(x+d)}{f(x)}=\frac{f(y+d)}{f(y)}$$

for every x, y, d. [Proof: First show that $|f(x)|=1$ for every x, then write $f(x)=e(g(x))$. We find that $g(x+1)-g(x)=g(y+1)-g(y)$ for every y, so g is linear.]

In other words, a trigonometric function is one that trivially maximizes the quantity

$$\mathbb{E}_{x,a,b}f(x)\overline{f(x+a)f(x+b)}f(x+a+b)$$

over all functions f with $\|f\|_\infty \leq 1$.

It is therefore natural to conjecture the following inverse theorem.

Let f be a function from $\mathbb{Z}_N$ such that $\|f\|_\infty \leq 1$ and $\|f\|_{U^3} \geq c$. Then there is a quadratic function $q : \mathbb{Z}_N \to \mathbb{Z}_N$ such that $\mathbb{E}_x f(x) e_N(-q(x)) \geq c'$, where c' depends on c only.

It is therefore natural to conjecture the following inverse theorem.

Let f be a function from $\mathbb{Z}_N$ such that $\|f\|_\infty \leq 1$ and $\|f\|_{U^3} \geq c$. Then there is a quadratic function $q : \mathbb{Z}_N \to \mathbb{Z}_N$ such that $\mathbb{E}_x f(x) e_N(-q(x)) \geq c'$, where c' depends on c only.

This has some plausibility because it is not hard to show that if f is a function that takes values of modulus 1 and if

$$f(x)\overline{f(x+a)f(x+b)}f(x+a+b) \overline{f(x+c)}f(x+a+c)f(x+b+c)\overline{f(x+a+b+c)} = 1$$

for every x, a, b, c, then $f(x)$ must be of the form $e_N(q(x))$.

It is therefore natural to conjecture the following inverse theorem.

Let f be a function from $\mathbb{Z}_N$ such that $\|f\|_\infty \leq 1$ and $\|f\|_{U^3} \geq c$. Then there is a quadratic function $q : \mathbb{Z}_N \to \mathbb{Z}_N$ such that $\mathbb{E}_x f(x) e_N(-q(x)) \geq c'$, where c' depends on c only.

This has some plausibility because it is not hard to show that if f is a function that takes values of modulus 1 and if

$$f(x)\overline{f(x+a)f(x+b)}f(x+a+b)$$
$$\overline{f(x+c)}f(x+a+c)f(x+b+c)\overline{f(x+a+b+c)} = 1$$

for every x, a, b, c, then $f(x)$ must be of the form $e_N(q(x))$.

However, the above statement is *false*.

Quadratic homomorphisms

Definition

Let $A \subset \mathbb{Z}_N$. A *quadratic homomorphism* from A to $\mathbb{Z}_N$ is a function q with the property that

$$\begin{aligned} q(x) - q(x+a) - q(x+b) + q(x+a+b) - q(x+c) \\ +q(x+a+c) + q(x+b+c) - q(x+a+b+c) = 0 \end{aligned}$$

for every x, a, b, c such that all the above linear combinations belong to A.

Equivalently,

$$q(x) - q(x+a) - q(x+b) + q(x+a+b)$$

is a function of a and b only.

The rough reason the naive conjecture is false

A quadratic homomorphism from $\mathbb{Z}_N$ to $\mathbb{Z}_N$ has to be a quadratic function, but quadratic homomorphisms on more general (dense) subsets $A \subset \mathbb{Z}_N$ can be genuinely different.

More precisely, if A is a dense subset of $\mathbb{Z}_N$ and q is a quadratic homomorphism from A to $\mathbb{Z}_N$, then let $f(x) = e_N(q(x))$ if $x \in A$ and let $f(x) = 0$ otherwise. Then $\|f\|_{U^3}$ is large but f does not have to correlate with a function of the form $e_N(ax^2 + bx + c)$.

Example: multidimensional arithmetic progressions

A set A of the form

$$\{x_0 + \sum_{i=1}^{d} s_i u_i : 0 \leq s_i < r_i\}$$

is a d-dimensional arithmetic progression with common differences $(u_1, \ldots, u_d)$ and side lengths $(r_1, \ldots, r_d)$.

Typically, a quadratic homomorphism on such a set will have a formula of the form

$$q(x_0 + \sum s_i u_i) = \sum_{i,j} a_{ij} s_i s_j + \sum_i b_i s_i + c$$

In other words, it is like a (not necessarily homogeneous) quadratic form on a d-dimensional vector space.

Definition

A *linear* (or *Freiman*) *homomorphism* on a set A is a function ϕ such that $\phi(x+d)-\phi(x)$ depends only on d.

A linear homomorphism on a d-dimensional progression has a formula of the form

$$\phi(x_0+\sum_i s_i u_i)=\sum_i b_i s_i + c.$$

Theorem

Let $\|f\|_\infty \leq 1$ and let $\|f\|_{U^3} \geq c$. Then there is a d-dimensional progression P and a constant c', with d and c' depending on c only, such that

$$\mathbb{E}_{x,a}\mathbb{E}_{b\in P} f(x)\overline{f(x+a)f(x+b)}f(x+a+b)e_N(-a\phi(b)) \geq c'$$

A key tool in proving this result is Freiman's theorem and in particular Ruzsa's proof of Freiman's theorem.

From bilinear to quadratic

We know that

$$2\lambda ab = \lambda(x^2 - (x+a)^2 - (x+b)^2 + (x+a+b)^2).$$

From this it follows that

$$\mathbb{E}_{x,a,b} f(x)\overline{f(x+a)f(x+b)}f(x+a+b)e_N(-2\lambda ab)$$

is equal to

$$\mathbb{E}_{x,a,b} g(x)\overline{g(x+a)g(x+b)}g(x+a+b) \;=\; \|g\|_{U^2}^4,$$

where $g(x) = f(x)e_N(-\lambda x^2)$.

But there is a rough equivalence between $\|g\|_{U^2}$ and $\|\hat{g}\|_\infty$, which tells us that if

$$\mathbb{E}_{x,a,b} f(x)\overline{f(x+a)f(x+b)}f(x+a+b)e_N(-2\lambda ab) \geq c$$

then there exists r such that $|\hat{g}(r)| = |\mathbb{E}_x f(x)e_N(-\lambda x^2 - rx)| \geq c^{1/2}$.

Rough principle: bilinear correlation implies quadratic correlation.

But there is a rough equivalence between $\|g\|_{U^2}$ and $\|\hat{g}\|_\infty$, which tells us that if

$$\mathbb{E}_{x,a,b} f(x)\overline{f(x+a)f(x+b)}f(x+a+b)e_N(-2\lambda ab) \geq c$$

then there exists r such that $|\hat{g}(r)| = |\mathbb{E}_x f(x)e_N(-\lambda x^2 - rx)| \geq c^{1/2}$.

Rough principle: bilinear correlation implies quadratic correlation.

Correction: symmetric bilinear correlation implies quadratic correlation.

A weak inverse theorem

Unfortunately, the “bilinear form” $(a,b)\mapsto a\phi(b)$ is not usually symmetric. One can nevertheless use the bilinear theorem to prove a “weak inverse theorem”.

Theorem

If $\|f\|_{U^3}\geq c$ then it is possible to partition $\mathbb{Z}_N$ into arithmetic progressions P_i of length N^{c_1} and find for each i a quadratic polynomial q_i such that the average correlation $|\mathbb{E}_{x\in P_i}f(x)e_N(-q_i(x))|$ is at least c_2.

This is enough for a proof of Szemerédi’s theorem for progressions of length 4. However, it is weak in the sense that the converse does not hold, even roughly.

The Green-Tao symmetry argument

Green and Tao found a way of replacing the bilinear form $a\phi(b)$ by a symmetric bilinear form. This allowed them to obtain the following strong inverse theorem (which has many equivalent formulations).

Theorem

If $\|f\|_{U^3} \geq c$ then there is a dense d-dimensional arithmetic progression $A \subset \mathbb{Z}_N$ and a quadratic homomorphism $q : A \to \mathbb{Z}_N$ such that $|\mathbb{E}_{x \in A} f(x) e_N(-q(x))| \geq c'$.

In other words, f must correlate with a *generalized quadratic phase function*.

The Green-Tao symmetry argument

Green and Tao found a way of replacing the bilinear form $a\phi(b)$ by a symmetric bilinear form. This allowed them to obtain the following strong inverse theorem (which has many equivalent formulations).

Theorem

If $\|f\|_{U^3} \geq c$ then there is a dense d-dimensional arithmetic progression $A \subset \mathbb{Z}_N$ and a quadratic homomorphism $q : A \to \mathbb{Z}_N$ such that $|\mathbb{E}_{x \in A} f(x) e_N(-q(x))| \geq c'$.

In other words, f must correlate with a *generalized quadratic phase function*.

- A crucial ingredient in the asymptotic estimate for the number of APs of length 4 in the primes.

The Green-Tao symmetry argument

Green and Tao found a way of replacing the bilinear form $a\phi(b)$ by a symmetric bilinear form. This allowed them to obtain the following strong inverse theorem (which has many equivalent formulations).

Theorem

If $\|f\|_{U^3} \geq c$ then there is a dense d-dimensional arithmetic progression $A \subset \mathbb{Z}_N$ and a quadratic homomorphism $q : A \to \mathbb{Z}_N$ such that $|\mathbb{E}_{x \in A} f(x) e_N(-q(x))| \geq c'$.

In other words, f must correlate with a *generalized quadratic phase function*.

- A crucial ingredient in the asymptotic estimate for the number of APs of length 4 in the primes.
- Other applications, to be discussed ...