Structure and (pseudo-)randomness in combinatorics

FOCS 2007 tutorial

October 20, 2007

Terence Tao (UCLA)

Large data

In combinatorics, one often deals with high-complexity objects, such as

- Functions $f : \mathbb{F}_2^n \to \mathbf{R}$ on a Hamming cube;
- Sets $A \subset \mathbb{F}_2^n$ in that Hamming cube $\mathbb{F}_2^n$; or
- Graphs $G = (V, E)$ on $|V| = N$ vertices.

One should think of $|\mathbb{F}_2^n| = 2^n$ and N as being very large, thus these objects have a large amount of informational entropy.

In this talk we will be primarily concerned with dense objects, e.g.

- Functions $f : \mathbb{F}_2^n \to \mathbf{R}$ with $\mathbf{E}_{x \in \mathbb{F}_2^n} f(x) := \frac{1}{2^n} \sum_{x \in \mathbb{F}_2^n} |f(x)|$ large;
- Sets $A \subset \mathbb{F}_2^n$ with $|A|/2^n$ large;
- Graphs $G = (V, E)$ with $|E|/|\binom{V}{2}|$ large.

In particular, we shall regard sparse objects (or sparse perturbations of dense objects) as "negligible".

All of the above objects can be modeled as elements of a (real) finite-dimensional Hilbert space H:

- The functions $f : \mathbb{F}_2^n \to \mathbf{R}$ form a Hilbert space H with inner product $\langle f, g \rangle_H := \mathbf{E}_{x \in \mathbb{F}_2^n} f(x) g(x)$.
- A set $A \subset \mathbb{F}_2^n$ can be identified with its indicator function $1_A : \mathbb{F}_2^n \to \{0, 1\}$, which lies in H.
- A graph $G = (V, E)$ can be identified with a symmetric function $1_E : V \times V \to \{0, 1\}$ in the Hilbert space of functions $f : V \times V \to \mathbf{R}$ with norm $\langle f, g \rangle_H := \mathbf{E}_{v,w \in V} f(v, w) g(v, w)$.

The dimension of these Hilbert spaces is finite, but extremely large. Thus these objects have many “degrees of freedom”.

In combinatorics one often has to deal with arbitrary objects in such a class - objects with no obvious usable structure.

Structure and pseudorandomness

While the space H of arbitrary objects under consideration has a huge number of degrees of freedom, the space of interesting or structured objects typically has a much smaller number of degrees of freedom. What "structured" means varies from context to context.

Examples of structure:

- Functions $f : \mathbb{F}_2^n \to \mathbf{R}$ which exhibit linear (Fourier) behaviour;
- Functions $f : \mathbb{F}_2^n \to \mathbf{R}$ which exhibit low-degree polynomial (Reed-Muller) behaviour;
- Sets $A \subset \mathbb{F}_2^n$ which only depend on a few of the coordinates of $\mathbb{F}_2^n$ (dictators, juntas);
- Graphs $G = (V, E)$ which are determined by a low-complexity vertex partition (e.g. complete bipartite graphs).

One might also consider computational complexity notions of structure.

Sometimes it is important to distinguish between several "quality levels" of structure:

- A "100%-structured" object might be one in which some statistic measuring structure is exactly equal to its theoretical maximum;
- A "99%-structured" object might be one in which some statistic measuring structure is very close to its theoretical maximum;
- A "1%-structured" object might be one in which some statistic measuring structure is within a multiplicative constant of its theoretical maximum.

Example: linearity

- A function $f : \mathbb{F}_2^n \to \{-1, +1\}$ is "100%-linear" if we have $f(x + y) = f(x)f(y)$ for all $x, y \in \mathbb{F}_2^n$;
- A function $f : \mathbb{F}_2^n \to \{-1, +1\}$ is "99%-linear" if we have $f(x + y) = f(x)f(y)$ for at least $1 - \varepsilon$ of all $x, y \in \mathbb{F}_2^n$;
- A function $f : \mathbb{F}_2^n \to \{-1, +1\}$ is "1%-linear" if we have $f(x + y) = f(x)f(y)$ for at least $\frac{1}{2} + \varepsilon$ of all $x, y \in \mathbb{F}_2^n$.

A 99%-linear function is always close to a 100%-linear one (Blum-Luby-Rubinfeld); a 1%-linear function always correlates with a 100%-linear one (Plancherel's theorem).

Given a concept of structure, one can often define a dual notion of pseudorandom objects - objects which are "almost orthogonal" or have "low correlation" with structured objects.

One can often show by standard probabilistic, counting, or entropy arguments that random objects tend to be almost orthogonal to all structured objects, thus justifying the terminology "pseudorandom".

Examples of pseudorandomness as duals of structure:

- Functions $f : \mathbb{F}_2^n \to \mathbf{R}$ which are Fourier-pseudorandom, i.e. have low Fourier coefficients (dual of Fourier structure);
- Functions $f : \mathbb{F}_2^n \to \mathbf{R}$ which are polynomially-pseudorandom, i.e. have low correlations with low-degree polynomials (dual of Reed-Muller structure);
- Sets $A \subset \mathbb{F}_2^n$ in which each coordinate has small low-height Fourier coefficients (dual of dictators and juntas);
- Graphs $G = (V, E)$ which are ε-regular (dual of

complete bipartite graphs).

In the previous examples, we began by defining structure and then created a dual notion of pseudorandomness. Thus pseudorandomness is defined “extrinsically”, by measuring its correlation with structured objects. In many cases we have an opposite situation: we begin with an “intrinsically defined” notion of pseudorandomness and wish to discover its dual notion of structure - the “obstructions” to that conception of pseudorandomness.

Computing such duals explicitly can sometimes be difficult, but is also very worthwhile; it provides a way to test whether a given object is structured or pseudorandom, or a combination of both.

Examples of "intrinsic" pseudorandomness:

- Functions $f : \mathbb{F}_2^n \to \mathbf{R}$ whose pair correlations $\mathbf{E}_{x \in \mathbb{F}_2^n} f(x) f(x+h)$ are small for most $h \in \mathbb{F}_2^n$;
- Functions $f : \mathbb{F}_2^n \to \mathbf{R}$ whose k-point correlations $\mathbf{E}_{x \in \mathbb{F}_2^n} f(x+h_1) \ldots f(x+h_k)$ are small for most $h_1, \ldots, h_k \in \mathbb{F}_2^n$;
- Functions $f : \mathbb{F}_2^n \to \mathbf{R}$ whose Gowers norms $\|f\|_{U^d(\mathbb{F}_2^n)} := (\mathbf{E}_{L:\mathbb{F}_2^d \to \mathbb{F}_2^n} \mathbf{E}_{x \in \mathbb{F}_2^n} \prod_{\omega \in \mathbb{F}_2^d} f(x + L\omega))^{1/2^d}$ are small;
- Graphs with a near-minimal (for a given edge density) number of 4-cycles.

Examples of structure as duals of pseudorandomness:

- A (bounded) function $f : \mathbb{F}_2^n \to \mathbf{R}$ has many large pair correlations if and only if has a large Fourier coefficient. (Plancherel's theorem)
- A (bounded) function $f : \mathbb{F}_2^n \to \mathbf{R}$ has large Gowers norm $\|f\|_{U^d(\mathbb{F}_2^n)}$ if and only if it has large correlation with a Reed-Muller codeword of degree at most $d-1$. (Gowers inverse conjecture; only completely proven for $d \leq 3$.)
- A graph has a large number of 4-cycles if and only if it is *not* ε-regular, i.e. it correlates with a complete bipartite graph. (Chung-Graham-Wilson)

General principles

0. Negligibility: pseudorandom objects tend to have negligible impact on statistics, averages, or correlations.

1. Dichotomy: Objects which are not pseudorandom tend to correlate with a structured object, and vice versa.

2. Structure theorem: Arbitrary objects can be decomposed into pseudorandom and structured components, possibly up to a small error.
3. Rigidity: Objects which are “almost”, “statistically”, or “locally” structured tend to be close to objects which actually *are* structured.
4. Classification: Structured objects can often be classified algebraically by using various bases.

These principles give a strategy to understand arbitrary objects, by splitting them into their pseudorandom and structured components.

Structure theorems in Hilbert spaces

Let us now focus on more rigorous formulations of the structure theorem principle. Specifically, given a (bounded) vector $f \in H$, we would like to decompose

$$f = f_{\rm str} + f_{\rm psd} + f_{\rm err}$$

where $f_{\rm str}$ is "structured", $f_{\rm psd}$ is "pseudorandom", and $f_{\rm err}$ is a small error. One can view $f_{\rm str}$ as an "effective" version of f, since $f_{\rm psd}$ and $f_{\rm err}$ are often negligible.

Sometimes we also want to enforce some orthogonality between $f_{\rm str}$, $f_{\rm psd}$, and $f_{\rm err}$.

Example: orthogonal projection

Theorem 1. Let V be a subspace of H (consisting of the "structured" vectors). Then every $f \in H$ can be uniquely decomposed as $f = f_{\mathrm{str}} + f_{\mathrm{psd}} + f_{\mathrm{err}}$, where

- f_{str} lies in V;
- f_{psd} is orthogonal to V; and
- $f_{\mathrm{err}} = 0$.

We recall that there are two standard proofs of this theorem: the first using the Gram-Schmidt orthogonalisation process, and the other by minimising $\|f - f_{\mathrm{str}}\|_H^2$ over all $f_{\mathrm{str}} \in V$. The latter proof is more relevant here; it relies on the dichotomy that if $f - f_{\mathrm{str}}$ is not orthogonal to V, then one can adjust f_{str} in V in order to decrease $\|f - f_{\mathrm{str}}\|_H^2$.

One can view this variational approach as a prototype of an "energy decrement argument" approach to structure theorems.

Example: thresholding

Theorem 2. Let $v_1, \ldots, v_n$ be an orthonormal basis of H (representing the fundamental "structured" vectors). Let $0 < \varepsilon \le 1$. Then every $f \in H$ with $\|f\|_H \le 1$ can be uniquely decomposed as $f = f_{\text{str}} + f_{\text{psd}} + f_{\text{err}}$, where

- $f_{\text{str}} = \sum_{i \in I} c_i v_i$ is such that $|I| \le 1/\varepsilon^2$ and $\varepsilon < |c_i| \le 1$;
- $f_{\text{psd}} = \sum_{i \notin I} c_i v_i$ is such that $|\langle f_{\text{psd}}, v_i \rangle| \le \varepsilon$ for all i; and
- $f_{\text{err}} = 0$.

Also, f_{str} and f_{psd} are orthogonal.

This theorem can be proven quickly from the Fourier inversion formula $f = \sum_i \langle f, v_i \rangle v_i$ and the Plancherel identity $\|f\|_H^2 = \sum_i |\langle f, v_i \rangle|^2$. But it is instructive to see a proof that relies less on these identities, and instead runs via the following algorithm:

- Step 0. Initialise $I = \emptyset$, $f_{\text{str}} = f_{\text{err}} = 0$, and $f_{\text{psd}} = f$.
- Step 1. If $|\langle f_{\text{psd}}, v_i \rangle| \leq \varepsilon$ for all i then STOP.
- Step 2. Otherwise, locate an i such that $|\langle f_{\text{psd}}, v_i \rangle| > \varepsilon$, and transfer i to I and $\langle f_{\text{psd}}, v_i \rangle v_i$ to f_{str}. Now return to Step 1.

Note that at each stage of this algorithm, the *energy* $\|f_{\mathrm{str}}\|_H^2$ of f_{str} increases by at least ε^2 (by Pythagoras' theorem); or equivalently, the energy of $\|f_{\mathrm{psd}}\|_H^2$ decreases by at least ε^2. Also by Pythagoras' theorem, we hve $0 \leq \|f_{\mathrm{str}}\|_H^2 \leq \|f\|_H^2 \leq 1$. So the algorithm must terminate after at most $1/\varepsilon^2$ steps.

One can view this algorithmic approach as a prototype of the "energy increment argument" approach to structure theorems.

Now we consider a common situation, in which we have a finite set $S \subset H$ of "fundamental structured vectors", which have magnitude at most 1, but which are not necessarily orthogonal. We would like to decompose an arbitrary $f \in H$ with $\|f\|_H \leq 1$ into components $f = f_{\mathrm{str}} + f_{\mathrm{psd}} + f_{\mathrm{err}}$, where

- f_{str} can be "efficiently represented" as a bounded linear combination of a few vectors from S;
- f_{psd} has low correlations with any vector from S; and
- f_{err} has a small norm $\|f_{\mathrm{err}}\|_H$.

Examples of the set S of fundamental structured vectors:

- S could be the set of linear functions $x \mapsto (-1)^{\xi \cdot x}$ on $\mathbb{F}_2^n$ (Fourier characters).
- S could be the set of polynomial functions of degree at most d on $\mathbb{F}_2^n$ (Reed-Muller codewords).
- S could be the set of indicator functions $1_{A \times B} : V \times V \to \{0, 1\}$, where $A, B \subset V$.

Our arguments here will not depend on the exact nature of S, other than the hypothesis that every vector in S has at most unit magnitude.

If we fix S, we can define structure and pseudorandomness more quantitatively:

Definition. A vector $f \in H$ is (M, K)-*structured* if one can write $f = \sum_{i=1}^{K} c_i v_i$ for some $v_i \in S$ and some real numbers c_i with $|c_i| \leq M$.

Definition. A vector $f \in H$ is ε-*pseudorandom* if we have $|\langle f, v \rangle| \leq \varepsilon$ for all $v \in S$.

The orthogonal projection theorem (Theorem 1), applied with V equal to the space spanned by S allows one to decompose $f = f_{\rm str} + f_{\rm psd} + f_{\rm err}$ where $f_{\rm psd}$ is 0-pseudorandom and $\|f_{\rm err}\|_H = 0$, but the only thing one gets to say about $f_{\rm str}$ is that it is (M, K)-structured for some $M, K < \infty$; no bound is provided.

The thresholding theorem (Theorem 2), in contrast, gives a decomposition $f = f_{\rm str} + f_{\rm psd} + f_{\rm err}$ where $f_{\rm psd}$ is ε-pseudorandom, $\|f_{\rm err}\|_H = 0$, and $f_{\rm str}$ is $(1, 1/\varepsilon^2)$-structured; but it requires the vectors in S to be orthonormal.

One can generalise Theorem 2 to non-orthonormal systems:

Weak structure theorem. Let $0 < \varepsilon \leq 1$. Then every $f \in H$ with $\|f\|_H \leq 1$ can be decomposed as $f = f_{\mathrm{str}} + f_{\mathrm{psd}} + f_{\mathrm{err}}$, where

- f_{str} is $(O_\varepsilon(1), 1/\varepsilon^2)$-structured;
- f_{psd} is ε-pseudorandom;
- $f_{\mathrm{err}} = 0$.

(The decomposition is no longer unique.)

The proof proceeds by a slight modification of the energy decrement argument:

- Step 0. Initialise $f_{\text{str}} = f_{\text{err}} = 0$, and $f_{\text{psd}} = f$.
- Step 1. If f_{psd} is ε-pseudorandom then STOP.
- Step 2. Otherwise, locate a $v \in S$ such that $|\langle f_{\text{psd}}, v\rangle| > \varepsilon$. Transfer a small multiple of v to f_{str}, ehough to decrease $\|f_{\text{psd}}\|_H^2$ by at least ε^2. Now return to Step 1.

It is not difficult to show that this algorithm establishes the theorem.

The weak structure theorem is often insufficient for many applications, because the pseudorandomness of f_{psd} is not particularly good compared with the complexity of f_{str}. However, it can be iterated to a better theorem:

Strong structure theorem. Let $0 < \varepsilon \leq 1$, and let $F : \mathbf{Z}^+ \to \mathbf{R}^+$ be an arbitrary function. Then every $f \in H$ with $\|f\|_H \leq 1$ can be decomposed as $f = f_{\mathrm{str}} + f_{\mathrm{psd}} + f_{\mathrm{err}}$, where

- f_{str} is (M, M)-structured for some $M = O_{F,\varepsilon}(1)$;
- f_{psd} is $1/F(M)$-pseudorandom;
- $\|f_{\mathrm{err}}\|_H \leq \varepsilon$.

Thus the pseudorandomness of f_{psd} can exceed the structure of f_{str} by an arbitrary amount. The catch is that the bound on M is poor, and that we must also allow the error f_{err} to be non-zero.

With a bit of additional effort one can make f_{str}, f_{psd}, and f_{err} orthogonal.

Sketch of proof:

- Set $M_0 = 1$ and $M_i = F(M_{i-1})$ for each $i = 1, 2, 3, \ldots$.
- For each i, we can decompose $f = f_{\mathrm{str},i} + f_{\mathrm{psd},i}$ where $f_{\mathrm{psd},i}$ is $1/M_i$-pseudorandom and $f_{\mathrm{str}i}$ is (essentially) (M_i, M_i)-structured.
- One can arrange matters so that all the $f_{\mathrm{str},i+1} - f_{\mathrm{str},i}$ are orthogonal to each other. In particular, $\|f_{\mathrm{str},i}\|_H^2$ is increasing. By the pigeonhole principle, we can thus find $i = O_\varepsilon(1)$ such that $\|f_{\mathrm{str},i}\|_H^2 - \|f_{\mathrm{str},i-1}\|_H^2 \leq \varepsilon$.
- Now set $f_{\mathrm{str}} := f_{\mathrm{str},i-1}$, $f_{\mathrm{psd}} := f - f_{\mathrm{str},i}$, $M = M_{i-1}$,

and $f_{\mathrm{err}} := f_{\mathrm{str},i} - f_{\mathrm{str},i-1}$.

As typical applications of the strong structure theorem, one can establish the graph regularity lemma of Szemerédi, and the arithmetic regularity lemma of Green. One can also obtain a hypergraph regularity lemma by a slightly more intricate application of the same ideas. These lemmas have a number of applications, for instance to establishing the testability of various graph-theoretic and arithmetic properties.

In these applications, the growth function F usually needs to be exponential growth. Since M is basically obtained by iterating F about $O(\varepsilon^{-O(1)})$ times, the bounds obtained by these methods is usually tower-exponential or worse in nature.

Structure theorems in measure spaces

In many cases, the Hilbert space H arises from a probability space $(X, \mathcal{X}, \mu)$ as the space $L^2(X, \mathcal{X}, \mu)$ of square-integrable, $\mathcal{X}$-measurable functions. For instance:

- For functions $f : \mathbb{F}_2^n \to \mathbf{R}$, $(X, \mathcal{X}, \mu)$ is the space $X = \mathbb{F}_2^n$ with uniform probability measure μ and the discrete σ-algebra $\mathcal{X}$.
- For graphs $G = (V, E)$, $(X, \mathcal{X}, \mu)$ is the space $X = V \times V$ with uniform probability measure μ and the discrete σ-algebra $\mathcal{B}$.

X is typically a finite set, so $\mathcal{X}$ is a partition of X.

In such contexts, one often wants the following properties:

- Positivity preservation: if f is non-negative, then f_{str} should also be non-negative.
- Comparison principle: if $|f| \leq g$, then one should have $|f_{\mathrm{str}}| \leq g_{\mathrm{str}}$. For instance, if f is bounded pointwise by 1, then f_{str} should be also.

The Hilbert space structure theorems do not provide such properties. However, this can be fixed by working with factors instead of vectors, and using conditional expectation instead of orthogonal projection.

A quick review of measure theory on finite sets:

Definition. A factor of $(X, \mathcal{X}, \mu)$ is a triplet $\mathcal{Y} = (Y, \mathcal{Y}, \pi)$, where Y is a set, $\mathcal{Y}$ is a σ-algebra (or partition) on Y, and $\pi : X \to Y$ is a measurable map, thus $\pi^{-1}(\mathcal{Y})$ is a coarsening of $\mathcal{X}$. The orthogonal projection $\mathbf{E}(f|\mathcal{Y})$ of $f \in L^2(X, \mathcal{X}, \mu)$ to $L^2(X, \pi^{-1}(\mathcal{Y}), \mu)$ is called the conditional expectation of f relative to Y.

Example 1: If X, Y are discrete, μ is uniform measure, $\pi : X \to Y$ is a colouring of X into distinct colour classes $\{\pi^{-1}(y) : y \in Y\}$, and $f : X \to \mathbf{R}$, then $\mathbf{E}(f|\mathcal{Y})(x) := \mathbf{E}_{\pi(x')=\pi(x)} f(x')$.

Example 2: Any function $f : X \to \mathbf{R}$ generates a factor $\mathcal{Y}_f = (\mathbf{R}, \mathcal{B}, f)$, where $\mathcal{B}$ is the Borel σ-algebra; this is the minimal factor with respect to which f is measurable, and is generated by the level sets $f^{-1}(\{x\})$ of f.

Example 3: In many applications, one needs a discretised version $\mathcal{Y}_{f,\varepsilon}$ of the above construction, in which $\mathcal{B}$ is now generated by the intervals $[n\varepsilon, (n+1)\varepsilon)$ for $n \in \mathbf{Z}$, thus f is "almost" measurable with respect to $\mathcal{Y}_{f,\varepsilon}$, which is generated by the level sets $f^{-1}([n\varepsilon, (n+1)\varepsilon))$.

(For technical reasons one sometimes has to shift the intervals $[n, \varepsilon, (n+1)\varepsilon)$ by a random translation.)

Conditional expectation is “better” than other orthogonal projections, because it preserves positivity,

$$f \geq 0 \implies \mathbf{E}(f|\mathcal{Y}) \geq 0$$

and also enjoys a comparison principle

$$|f| \leq g \implies |\mathbf{E}(f|\mathcal{Y})| \leq \mathbf{E}(g|\mathcal{Y}).$$

Definition. If $\mathcal{Y} = (Y, \mathcal{Y}, \pi)$ and $\mathcal{Y}' = (Y', \mathcal{Y}', \pi')$ are two factors of $(X, \mathcal{X}, \mu)$, we let $\mathcal{Y} \vee \mathcal{Y}' := (Y \times Y', \mathcal{Y} \times \mathcal{Y}', (\pi, \pi'))$ be the join of $\mathcal{Y}$ and $\mathcal{Y}'$.

Useful Pythagorean identities:

$$\|f\|_{L^2}^2 = \|\mathbf{E}(f|\mathcal{Y})\|_{L^2}^2 + \|f - \mathbf{E}(f|\mathcal{Y})\|_{L^2}^2$$

$$\|\mathbf{E}(f|\mathcal{Y} \vee \mathcal{Y}')\|_{L^2}^2 = \|\mathbf{E}(f|\mathcal{Y})\|_{L^2}^2 + \|\mathbf{E}(f|\mathcal{Y} \vee \mathcal{Y}') - \mathbf{E}(f|\mathcal{Y})\|_{L^2}^2$$

We now represent structure not by a collection S of vectors, but instead by a collection § of factors (e.g. factors generated by Reed-Muller codewords or by complete bipartite graphs). Fixing §, we can then define structure and pseudorandomness:

Definition. A function f is M-structured if it is measurable with respect to $\mathcal{Y}_1 \ldots \mathcal{Y}_m$ for some $m \leq M$, where each $\mathcal{Y}_i$ lies in §.

Definition. A function f is ε-pseudorandom if we have $\|\mathbf{E}(f|\mathcal{Y})\|_{L^2} \leq \varepsilon$.

By modifying the energy increment arguments discussed previously, one can obtain weak and strong structure theorems:

Weak structure theorem If $\|f\|_{L^2(X)} \leq 1$ and $\varepsilon > 0$, then we can decompose $f = f_{\mathrm{str}} + f_{\mathrm{psd}} + f_{\mathrm{err}}$ where

- f_{str} is $1/\varepsilon^2$-structured. In fact we have $f_{\mathrm{str}} = \mathbf{E}(f|\mathcal{Y})$ for some $1/\varepsilon^2$-structured factor $\mathcal{Y}$.
- f_{psd} is ε-pseudorandom.
- $f_{\mathrm{err}} = 0$.

Strong structure theorem If $\|f\|_{L^2(X)} \leq 1$, $\varepsilon > 0$, and $F : \mathbf{Z}^+ \to \mathbf{R}^+$, then we can decompose $f = f_{\rm str} + f_{\rm psd} + f_{\rm err}$ where

- $f_{\rm str}$ is M-structured for some $M = O_{F,\varepsilon}(1)$. In fact we have $f_{\rm str} = \mathbf{E}(f|\mathcal{Y})$ for some M-structured factor $\mathcal{Y}$.
- $f_{\rm psd}$ is $1/F(M)$-pseudorandom.
- $\|f_{\rm err}\|_{L^2} \leq \varepsilon$.

A weak structure theorem of this type (with the condition $\|f\|_{L^2(X)} \leq 1$ replaced by a weaker condition), together with the comparison principle, was decisive in establishing that the primes contained arbitrarily long arithmetic progressions.

Strong structure theorems of this type are related to structural theorems in ergodic theory, and can be used for instance to establish Szemerédi's theorem on arithmetic progressions.

Gowers uniformity

Now we specialise to a very specific notion of structure and pseudorandomness, given by the Gowers uniformity norm

$$\|f\|_{U^d(\mathbb{F}_2^n)} := (\mathbf{E}_{L:\mathbb{F}_2^d \to \mathbb{F}_2^n} \mathbf{E}_x \prod_{\omega \in \mathbb{F}_2^d} f(x + L\omega))^{1/2^d}$$

of a function $f : \mathbb{F}_2^n \to \mathbf{R}$ for $d \geq 1$. The d^{th} Gowers norm reflects the extent to which f behaves like a Reed-Muller codeword of order $d-1$ (i.e. $(-1)^P$, where P is a polynomial over $\mathbb{F}_2$ of degree at most d).

Examples:

$$\begin{aligned}
\|f\|_{U^1(\mathbb{F}_2^n)} &= |\mathbf{E}_{x\in\mathbb{F}_2^n} f(x)f(x+h)|^{1/2} \\
&= |\mathbf{E}_{x\in\mathbb{F}_2^n} f(x)| \\
\|f\|_{U^2(\mathbb{F}_2^n)} &= |\mathbf{E}_{x,h,k\in\mathbb{F}_2^n} f(x)f(x+h)f(x+k)f(x+h+k)|^{1/4} \\
\|f\|_{U^3(\mathbb{F}_2^n)} &= |\mathbf{E}_{x,h,k,l\in\mathbb{F}_2^n} f(x)f(x+h)f(x+k)\dots f(x+h+k+l)|^{1/8}
\end{aligned}$$

Functions with small U^d norm are called Gowers uniform of order $d-1$.

Some easy facts:

- Monotonicity:

$$\|f\|_{U^1} \leq \|f\|_{U^2} \leq \|f\|_{U^3} \leq \ldots \leq \|f\|_{L^\infty}.$$

- Cauchy-Schwarz-Gowers inequality:

$$|\mathbf{E}_{L:\mathbb{F}_2^d \to \mathbb{F}_2^n} \mathbf{E}_x \prod_{\omega \in \mathbb{F}_2^d} f_\omega(x + L\omega)| \leq \prod_{\omega \in \mathbb{F}_2^d} \|f_\omega\|_{U^d}.$$

- Norm properties:

$$\|f+g\|_{U^d} \leq \|f\|_{U^d} + \|g\|_{U^d}; \|cf\|_{U^d} = |c|\|f\|_{U^d}$$

$$\|f\|_{U^d} = 0 \iff f = 0 \text{ for } d \geq 2$$

If f takes values in $\{-1,+1\}$, then $\|f\|_{U^d}$ ranges between 0 and 1. If $\|f\|_{U^d}^{2^d} = 1-\varepsilon$, then we have the identity

$$f(x) = \prod_{\omega_1,\ldots,\omega_d=\{0,1\}:(\omega_1,\ldots,\omega_d)\neq 0} f(x+\omega_1 h_1 + \ldots + \omega_d h_d)$$

for randomly chosen $x, h_1, \ldots, h_d \in \mathbb{F}_2^n$ with probability $1-\varepsilon/2$. For instance, if $\|f\|_{U^2}^4 = 1-\varepsilon$, then

$$\mathbb{P}\left(f(x) = f(x+h)f(x+k)f(x+h+k)\right) = 1-\varepsilon/2.$$

From this, one can show

100% **inverse structure theorem** Let $f : \mathbb{F}_2^n \to \{-1, 1\}$ and $d \geq 1$. Then $\|f\|_{U^d} = 1$ if and only if f is a Reed-Muller codeword of order $d - 1$.

99% **inverse structure theorem** Let $f : \mathbb{F}_2^n \to \{-1, 1\}$, $d \geq 1$, and $\varepsilon > 0$. Then if $\|f\|_{U^d} \geq 1 - \delta$ for some sufficiently small $\delta = \delta(\varepsilon, d) > 0$, f is within ε in L^2 norm of a Reed-Muller codeword of order $d - 1$.

The first result is easy to prove by exploiting functional equations such as $f(x) = f(x+h)f(x+k)f(x+h+k)$.

The second result is due to Alon-Kaufman-Krivelevich-Litsyn-Ron, and implies that Reed-Muller codes are locally testable. The rough idea is to use expressions such as $f(x+h)f(x+k)f(x+h+k)$ as a "vote" as to what $f(x)$ should be, and then use majority vote to discover the Reed-Muller codeword.

Another approach is to proceed inductively, observing that if f has large U^d norm then $fT^h f$ will have large U^{d-1} norm for most h, where $T^h f(x) := f(x+h)$ is the shift of f by h.

The following result is conjectured:

1% inverse structure theorem? Let $f : \mathbb{F}_2^n \to \{-1, 1\}$, $d \geq 1$, and $\varepsilon > 0$. Then if $\|f\|_{U^d} \geq \varepsilon$, then there exists a Reed-Muller codeword g of order $d-1$ such that $|\langle f, g \rangle| \gg_{d,\varepsilon} 1$.

This is known for $d \leq 2$ by Plancherel's theorem, and also for $d = 3$ (Samorodnitsky). It remains open for $d > 3$, and is known as the Gowers inverse conjecture for $\mathbb{F}_2^n$. Very recently, Ben Green and I have been able to verify this conjecture in the case that f is a Reed-Muller codeword of much higher (but bounded) degree.

In the converse direction, one can easily show that

$\|f\|_{U^d} \geq |\langle f, g\rangle|$ for all Reed-Muller codewords g of order $d-1$.

The Gowers inverse conjecture, when combined with the general structured theorems discussed earlier, would have many useful applications. Basically, one would be able to split any function f into a bounded number of Reed-Muller codewords of order $d-1$, plus an error f_{psd} which is Gowers uniform of order $d-1$, and perhaps another small error f_{err}. This decomposition would allow us to understand local arithmetic patterns in functions in much the same way that the Szemerédi regularity lemma allows us to understand local patterns inside large graphs.

Besides the Gowers inverse conjecture, there are some related open problems in this area. One is to improve the quantitative bounds in the known results for that conjecture. Another is to establish an algorithmic version: the current arguments that produce a Reed-Muller codeword g correlating with a given function f of large norm are computationally expensive.

A related problem is to find a fast way to compute $\|f\|_{U^d(\mathbb{F}_2^n)}$ exactly. Clearly $\|f\|_{U^1(\mathbb{F}_2^n)}$ requires $O(2^n)$ computations. Using the fast Fourier transform, one can compute $\|f\|_{U^2(\mathbb{F}_2^n)}$ in $O(n2^n)$ computations. But even with the FFT, we only know how to compute $\|f\|_{U^3(\mathbb{F}_2^n)}$ in $O(n2^{2n})$ computations. Can we do better?