Structure and randomness in combinatorics

Terence Tao
Department of Mathematics, UCLA
405 Hilgard Ave, Los Angeles CA 90095
tao@math.ucla.edu

Abstract

Combinatorics, like computer science, often has to deal with large objects of unspecified (or unusable) structure. One powerful way to deal with such an arbitrary object is to decompose it into more usable components. In particular, it has proven profitable to decompose such objects into a structured *component, a* pseudo-random *component, and a* small *component (i.e. an error term); in many cases it is the structured component which then dominates. We illustrate this philosophy in a number of model cases.*

1. Introduction

In many situations in combinatorics, one has to deal with an object of large complexity or entropy - such as a graph on N vertices, a function on N points, etc., with N large. We are often interested in the worst-case behaviour of such objects; equivalently, we are interested in obtaining results which apply to *all* objects in a certain class, as opposed to results for almost all objects (in particular, random or average case behaviour) or for very specially structured objects. The difficulty here is that the spectrum of behaviour of an arbitrary large object can be very broad. At one extreme, one has very *structured* objects, such as complete bipartite graphs, or functions with periodicity, linear or polynomial phases, or other algebraic structure. At the other extreme are *pseudorandom* objects, which mimic the behaviour of random objects in certain key statistics (e.g. their correlations with other objects, or with themselves, may be close to those expected of random objects).

Fortunately, there is a fundamental phenomenon that one often has a *dichotomy* between structure and pseudorandomness, in that given a reasonable notion of structure (or pseudorandomness), there often exists a dual notion of pseudorandomness (or structure) such that an arbitrary object can be decomposed into a structured component and a pseudorandom component (possibly with a small error). Here are two simple examples of such decompositions:

(i) An *orthogonal decomposition* $f = f_{\rm str} + f_{\rm psd}$ of a vector f in a Hilbert space into its orthogonal projection $f_{\rm str}$ onto a subspace V (which represents the "structured" objects), plus its orthogonal projection $f_{\rm psd}$ onto the orthogonal complement $V^\perp$ of V (which represents the "pseudorandom" objects).

(ii) A *thresholding* $f = f_{\rm str} + f_{\rm psd}$ of a vector f, where f is expressed in terms of some basis $v_1, \ldots, v_n$ (e.g. a Fourier basis) as $f = \sum_{1 \leq i \leq n} c_i v_i$, the "structured" component $f_{\rm str} := \sum_{i:|c_i| \geq \lambda} c_i v_i$ contains the contribution of the large coefficients, and the "pseudorandom" component $f_{\rm psd} := \sum_{i:|c_i| < \lambda} c_i v_i$ contains the contribution of the small coefficients. Here $\lambda > 0$ is a thresholding parameter which one is at liberty to choose.

Indeed, many of the decompositions we discuss here can be viewed as variants or perturbations of these two simple decompositions. More advanced examples of decompositions include the Szemerédi regularity lemma for graphs (and hypergraphs), as well as various *structure theorems* relating to the Gowers uniformity norms, used for instance in [16], [18]. Some decompositions from classical analysis, most notably the *spectral decomposition* of a self-adjoint operator into orthogonal subspaces associated with the pure point, singular continuous, and absolutely continuous spectrum, also have a similar spirit to the structure-randomness dichtomy.

The advantage of utilising such a decomposition is that one can use different techniques to handle the structured component and the pseudorandom component (as well as the error component, if it is present). Broadly speaking, the structured component is often handled by algebraic or geometric tools, or by reduction to a "lower complexity" problem than the original problem, whilst the contribution of the pseudorandom and error components is shown to be negligible by using inequalities from analysis (which can range from the humble Cauchy-Schwarz inequality to other, much more advanced, inequalities). A particularly notable use of this type of decomposition occurs in the many dif-

ferent proofs of Szemerédi's theorem [24]; see e.g. [30] for further discussion.

In order to make the above general strategy more concrete, one of course needs to specify more precisely what "structure" and "pseudorandomness" means. There is no single such definition of these concepts, of course; it depends on the application. In some cases, it is obvious what the definition of one of these concepts is, but then one has to do a non-trivial amount of work to describe the dual concept in some useful manner. We remark that *computational* notions of structure and randomness do seem to fall into this framework, but thus far all the applications of this dichotomy have focused on much simpler notions of structure and pseudorandomness, such as those associated to Reed-Muller codes.

In these notes we give some illustrative examples of this structure-randomness dichotomy. While these examples are somewhat abstract and general in nature, they should by no means be viewed as the definitive expressions of this dichotomy; in many applications one needs to modify the basic arguments given here in a number of ways. On the other hand, the core ideas in these arguments (such as a reliance on energy-increment or energy-decrement methods) appear to be fairly universal. The emphasis here will be on illustrating the "nuts-and-bolts" of structure theorems; we leave the discussion of the more advanced structure theorems and their applications to other papers.

One major topic we will not be discussing here (though it is lurking underneath the surface) is the role of ergodic theory in all of these decompositions; we refer the reader to [30] for further discussion. Similarly, the recent ergodic-theoretic approaches to hypergraph regularity, removal, and property testing in [31], [3] will not be discussed here, in order to prevent the exposition from becoming too unfocused. The lecture notes here also have some intersection with the author's earlier article [27].

2. Structure and randomness in a Hilbert space

Let us first begin with a simple case, in which the objects one is studying lies in some real finite-dimensional Hilbert space H, and the concept of structure is captured by some known set S of "basic structured objects". This setting is already strong enough to establish the Szemerédi regularity lemma, as well as variants such as Green's arithmetic regularity lemma. One should think of the dimension of H as being extremely large; in particular, we do not want any of our quantitative estimates to depend on this dimension.

More precisely, let us designate a finite collection $S \subset H$ of "basic structured" vectors of bounded length; we assume for concreteness that $\|v\|_H \leq 1$ for all $v \in S$. We would like to view elements of H which can be "efficiently represented" as linear combinations of vectors in S as *structured*, and vectors which have low correlation (or more precisely, small inner product) to all vectors in S as *pseudorandom*. More precisely, given $f \in H$, we say that f is (M,K)-*structured* for some $M, K > 0$ if one has a decomposition

$$f = \sum_{1 \leq i \leq M} c_i v_i$$

with $v_i \in S$ and $c_i \in [-K, K]$ for all $1 \leq i \leq M$. We also say that f is ε-*pseudorandom* for some $\varepsilon > 0$ if we have $|\langle f, v \rangle_H| \leq \varepsilon$ for all $v \in S$. It is helpful to keep some model examples in mind:

Example 2.1 (Fourier structure). Let $\mathbf{F}_2^n$ be a Hamming cube; we identify the finite field $\mathbf{F}_2$ with $\{0,1\}$ in the usual manner. We let H be the 2^n-dimensional space of functions $f : \mathbf{F}_2^n \to \mathbf{R}$, endowed with the inner product

$$\langle f, g \rangle_H := \frac{1}{2^n} \sum_{x \in \mathbf{F}_2^n} f(x) g(x),$$

and let S be the space of *characters*,

$$S := \{e_\xi : \xi \in \mathbf{F}_2^n\},$$

where for each $\xi \in \mathbf{F}_2^n$, e_ξ is the function $e_\xi(x) := (-1)^{x \cdot \xi}$. Informally, a structured function f is then one which can be expressed in terms of a small number (e.g. $O(1)$) characters, whereas a pseudorandom function f would be one whose Fourier coefficients

$$\hat{f}(\xi) := \langle f, e_\xi \rangle_H \tag{1}$$

are all small.

Example 2.2 (Reed-Muller structure). Let H be as in the previous example, and let $1 \leq k \leq n$. We now let $S = S_k(\mathbf{F}_2^n)$ be the space of Reed-Muller codes $(-1)^{P(x)}$, where $P : \mathbf{F}_2^n \to \mathbf{F}_2$ is any polynomial of n variables with coefficients and degree at most k. For $k = 1$, this gives the same notions of structure and pseudorandomness as the previous example, but as we increase k, we enlarge the class of structured functions and shrink the class of pseudorandom functions. For instance, the function $(x_1, \ldots, x_n) \mapsto (-1)^{\sum_{1 \leq i < j \leq n} x_i x_j}$ would be considered highly pseudorandom when $k = 1$ but highly structured for $k \geq 2$.

Example 2.3 (Product structure). Let V be a set of $|V| = n$ vertices, and let H be the n^2-dimensional space of functions $f : V \times V \to \mathbf{R}$, endowed with the inner product

$$\langle f, g \rangle_H := \frac{1}{n^2} \sum_{v,w \in V} f(v,w) g(v,w).$$

Note that any graph $G = (V, E)$ can be identified with an element of H, namely the indicator function $1_E : V \times V \to \{0,1\}$ of the set of edges. We let S be the collection of

tensor products $(v,w) \mapsto 1_A(v)1_B(w)$, where A, B are subsets of V. Observe that 1_E will be quite structured if G is a complete bipartite graph, or the union of a bounded number of such graphs. At the other extreme, if G is an ε-*regular* graph of some edge density $0 < \delta < 1$ for some $0 < \varepsilon < 1$, in the sense that the number of edges between A and B differs from $\delta|A||B|$ by at most $\varepsilon|A||B|$ whenever $A, B \subset V$ with $|A|, |B| \geq \varepsilon n$, then $1_E - \delta$ will be $O(\varepsilon)$-pseudorandom.

We are interested in obtaining quantative answers to the following general problem: given an arbitrary bounded element f of the Hilbert space H (let us say $\|f\|_H \leq 1$ for concreteness), can we obtain a decomposition

$$f = f_{\mathrm{str}} + f_{\mathrm{psd}} + f_{\mathrm{err}} \tag{2}$$

where f_{str} is a structured vector, f_{psd} is a pseudorandom vector, and f_{err} is some small error?

One obvious "qualitative" decomposition arises from using the vector space $\mathrm{span}(S)$ spanned by the basic structured vectors S. If we let f_{str} be the orthogonal projection from f to this vector space, and set $f_{\mathrm{psd}} := f - f_{\mathrm{str}}$ and $f_{\mathrm{err}} := 0$, then we have perfect control on the pseudorandom and error components: f_{psd} is 0-pseudorandom and f_{err} has norm 0. On the other hand, the only control on f_{str} we have is the qualitative bound that it is (K, M)-structured for some finite $K, M < \infty$. In the three examples given above, the vectors S in fact span all of H, and this decomposition is in fact trivial!

We would thus like to perform a tradeoff, increasing our control of the structured component at the expense of worsening our control on the pseudorandom and error components. We can see how to achieve this by recalling how the orthogonal projection of f to $\mathrm{span}(S)$ is actually constructed; it is the vector v in $\mathrm{span}(S)$ which minimises the "energy" $\|f - v\|_H^2$ of the residual $f - v$. The key point is that if $v \in \mathrm{span}(S)$ is such that $f - v$ has a non-zero inner product with a vector $w \in S$, then it is possible to move v in the direction w to decrease the energy $\|f - v\|_H^2$. We can make this latter point more quantitative:

Lemma 2.4 (Lack of pseudorandomness implies energy decrement)**.** *Let H, S be as above. Let $f \in H$ be a vector with $\|f\|_H^2 \leq 1$, such that f is* not *ε-pseudorandom for some $0 < \varepsilon \leq 1$. Then there exists $v \in S$ and $c \in [-1/\varepsilon, 1/\varepsilon]$ such that $|\langle f, v\rangle| \geq \varepsilon$ and $\|f - cv\|_H^2 \leq \|f\|_H^2 - \varepsilon^2$.*

Proof. By hypothesis, we can find $v \in S$ be such that $|\langle f, v\rangle| \geq \varepsilon$, thus by Cauchy-Schwarz and hypothesis on S

$$1 \geq \|v\|_H \geq |\langle f, v\rangle| \geq \varepsilon.$$

We then set $c := \langle f, v\rangle / \|v\|_H^2$ (i.e. cv is the orthogonal projection of f to the span of v). The claim then follows from Pythagoras' theorem. □

If we iterate this by a straightforward greedy algorithm argument we now obtain

Corollary 2.5 (Non-orthogonal weak structure theorem)**.** *Let H, S be as above. Let $f \in H$ be such that $\|f\|_H \leq 1$, and let $0 < \varepsilon \leq 1$. Then there exists a decomposition* (2) *such that f_{str} is $(1/\varepsilon^2, 1/\varepsilon)$-structured, f_{psd} is ε-pseudorandom, and f_{err} is zero.*

Proof. We perform the following algorithm.

- Step 0. Initialise $f_{\mathrm{str}} := 0$, $f_{\mathrm{err}} := 0$, and $f_{\mathrm{psd}} := f$. Observe that $\|f_{\mathrm{psd}}\|_H^2 \leq 1$.
- Step 1. If f_{psd} is ε-pseudorandom then **S**TOP. Otherwise, by Lemma 2.4, we can find $v \in S$ and $c \in [-1/\varepsilon, 1/\varepsilon]$ such that $\|f_{\mathrm{psd}} - cv\|_H^2 \leq \|f_{\mathrm{psd}}\|_H^2 - \varepsilon^2$.
- Step 2. Replace f_{psd} by $f_{\mathrm{psd}} - cv$ and replace f_{str} by $f_{\mathrm{str}} + cv$. Now return to Step 1.

It is clear that the "energy" $\|f_{\mathrm{psd}}\|_H^2$ decreases by at least ε^2 with each iteration of this algorithm, and thus this algorithm terminates after at most $1/\varepsilon^2$ such iterations. The claim then follows. □

Corollary 2.5 is not very useful in applications, because the control on the structure of f_{str} are relatively poor compared to the pseudorandomness of f_{psd} (or vice versa). One can do substantially better here, by allowing the error term f_{err} to be non-zero. More precisely, we have

Theorem 2.6 (Strong structure theorem)**.** *Let H, S be as above, let $\varepsilon > 0$, and let $F : \mathbf{Z}^+ \to \mathbf{R}^+$ be an arbitrary function. Let $f \in H$ be such that $\|f\|_H \leq 1$. Then we can find an integer $M = O_{F,\varepsilon}(1)$ and a decomposition* (2) *where f_{str} is (M, M)-structured, f_{psd} is $1/F(M)$-pseudorandom, and f_{err} has norm at most ε.*

Here and in the sequel, we use subscripts in the $O()$ asymptotic notation to denote that the implied constant depends on the subscripts. For instance, $O_{F,\varepsilon}(1)$ denotes a quantity bounded by $C_{F,\varepsilon}$, for some quantity $C_{F,\varepsilon}$ depending only on F and ε. Note that the pseudorandomness of f_{psd} can be of arbitrarily high quality compared to the complexity of f_{str}, since we can choose F to be whatever we please; the cost of doing so, of course, is that the upper bound on M becomes worse when F is more rapidly growing.

To prove Theorem 2.6, we first need a variant of Corollary 2.5 which gives some orthogonality between f_{str} and f_{psd}, at the cost of worsening the complexity bound on f_{str}.

Lemma 2.7 (Orthogonal weak structure theorem)**.** *Let H, S be as above. Let $f \in H$ be such that $\|f\|_H \leq 1$, and let $0 < \varepsilon \leq 1$. Then there exists a decomposition* (2) *such that f_{str} is $(1/\varepsilon^2, O_\varepsilon(1))$-structured, f_{psd} is ε-pseudorandom, f_{err} is zero, and $\langle f_{\mathrm{str}}, f_{\mathrm{psd}}\rangle_H = 0$.*

Proof. We perform a slightly different iteration to that in Corollary 2.5, where we insert an additional orthogonalisation step within the iteration to a subspace V:

- Step 0. Initialise $V := \{0\}$ and $f_{\mathrm{err}} := 0$.
- Step 1. Set f_{str} to be the orthogonal projection of f to V, and $f_{\mathrm{psd}} := f - f_{\mathrm{str}}$.
- Step 2. If f_{psd} is ε-pseudorandom then **STOP**. Otherwise, by Lemma 2.4, we can find $v \in S$ and $c \in [-1/\varepsilon, 1/\varepsilon]$ such that $|\langle f_{\mathrm{psd}}, v\rangle_H| \geq \varepsilon$ and $\|f_{\mathrm{psd}} - cv\|_H^2 \leq \|f_{\mathrm{psd}}\|_H^2 - \varepsilon^2$.
- Step 3. Replace V by $\mathrm{span}(V \cup \{v\})$, and return to Step 1.

Note that at each stage, $\|f_{\mathrm{psd}}\|_H$ is the minimum distance from f to V. Because of this, we see that $\|f_{\mathrm{psd}}\|_H^2$ decreases by at least ε^2 with each iteration, and so this algorithm terminates in at most $1/\varepsilon^2$ steps.

Suppose the algorithm terminates in M steps for some $M \leq 1/\varepsilon^2$. Then we have constructed a nested flag

$$\{0\} = V_0 \subset V_1 \subset \ldots \subset V_M$$

of subspaces, where each V_i is formed from V_{i-1} by adjoining a vector v_i in S. Furthermore, by construction we have $|\langle f_i, v_i\rangle| \geq \varepsilon$ for some vector f_i of norm at most 1 which is orthogonal to V_{i-1}. Because of this, we see that v_i makes an angle of $\Theta_\varepsilon(1)$ with V_{i-1}. As a consequence of this and the Gram-Schmidt orthogonalisation process, we see that $v_1, \ldots, v_i$ is a well-conditioned basis of V_i, in the sense that any vector $w \in W_i$ can be expressed as a linear combination of $v_1, \ldots, v_i$ with coefficients of size $O_{\varepsilon,i}(\|w\|_H)$. In particular, since f_{str} has norm at most 1 (by Pythagoras' theorem) and lies in V_M, we see that f_{str} is a linear combination of $v_1, \ldots, v_M$ with coefficients of size $O_{M,\varepsilon}(1) = O_\varepsilon(1)$, and the claim follows. □

We can now iterate the above lemma and use a pigeonholing argument to obtain the strong structure theorem.

Proof of Theorem 2.6. We first observe that it suffices to prove a weakened version of Theorem 2.6 in which f_{str} is $(O_{M,\varepsilon}(1), O_{M,\varepsilon}(1))$-structured rather than (M, M) structured. This is because one can then recover the original version of Theorem 2.6 by making F more rapidly growing, and redefining M; we leave the details to the reader. Also, by increasing F if necessary we may assume that F is integer-valued and $F(M) > M$ for all M.

We now recursively define $M_0 := 1$ and $M_i := F(M_{i-1})$ for all $i \geq 1$. We then recursively define $f_0, f_1, \ldots$ by setting $f_0 := f$, and then for each $i \geq 1$ using Lemma 2.7 to decompose $f_{i-1} = f_{\mathrm{str},i} + f_i$ where $f_{\mathrm{str},i}$ is $(O_{M_i}(1), O_{M_i}(1))$-structured, and f_i is $1/M_i$-pseudorandom and orthogonal to $f_{\mathrm{str},i}$. From Pythagoras' theorem we see that the quantity $\|f_i\|_H^2$ is decreasing, and varies between 0 and 1. By the pigeonhole principle, we can thus find $1 \leq i \leq 1/\varepsilon^2+1$ such that $\|f_{i-1}\|_H^2 - \|f_i\|_H^2 \leq \varepsilon^2$; by Pythagoras' theorem, this implies that $\|f_{\mathrm{str},i}\|_H \leq \varepsilon$. If we then set $f_{\mathrm{str}} := f_{\mathrm{str},0} + \ldots + f_{\mathrm{str},i-1}$, $f_{\mathrm{psd}} := f_i$, $f_{\mathrm{err}} := f_{\mathrm{str},i}$, and $M := M_{i-1}$, we obtain the claim. □

Remark 2.8. By tweaking the above argument a little bit, one can also ensure that the quantities $f_{\mathrm{str}}, f_{\mathrm{psd}}, f_{\mathrm{err}}$ in Theorem 2.6 are orthogonal to each other. We leave the details to the reader.

Remark 2.9. The bound $O_{F,\varepsilon}(1)$ on M in Theorem 2.6 is quite poor in practice; roughly speaking, it is obtained by iterating F about $O(1/\varepsilon^2)$ times. Thus for instance if F is of exponential growth (which is typical in applications), M can be tower-exponential size in ε. These excessively large values of M unfortunately seem to be necessary in many cases, see e.g. [8] for a discussion in the case of the Szemerédi regularity lemma, which can be deduced as a consequence of Theorem 2.6.

To illustrate how the strong regularity lemma works in practice, we use it to deduce the arithmetic regularity lemma of Green [13] (applied in the model case of the Hamming cube $\mathbf{F}_2^n$). Let A be a subset of $\mathbf{F}_2^n$, and let $1_A : \mathbf{F}_2^n \to \{0,1\}$ be the indicator function. If V is an affine subspace (over $\mathbf{F}_2$) of $\mathbf{F}_2^n$, we say that A is *ε-regular* in V for some $0 < \varepsilon < 1$ if we have

$$|\mathbf{E}_{x \in V}(1_A(x) - \delta_V)e_\xi(x)| \leq \varepsilon$$

for all characters e_ξ, where $\mathbf{E}_{x\in V} f(x) := \frac{1}{|V|}\sum_{x\in V} f(x)$ denotes the average value of f on V, and $\delta_V := \mathbf{E}_{x\in V} 1_A(x) = |A \cap V|/|V|$ denotes the density of A in V. The following result is analogous to the celebrated Szemerédi regularity lemma:

Lemma 2.10 (Arithmetic regularity lemma)**.** *[13] Let $A \subset \mathbf{F}_2^n$ and $0 < \varepsilon \leq 1$. Then there exists a subspace V of codimension $d = O_\varepsilon(1)$ such that A is ε-regular on all but $\varepsilon 2^d$ of the translates of V.*

Proof. It will suffice to establish the claim with the weaker claim that A is $O(\varepsilon^{1/4})$-regular on all but $O(\sqrt{\varepsilon}2^d)$ of the translates of V, since one can simply shrink ε to obtain the original version of Lemma 2.10.

We apply Theorem 2.6 to the setting in Example 2.1, with $f := 1_A$, and F to be chosen later. This gives us an integer $M = O_{F,\varepsilon}(1)$ and a decomposition

$$1_A = f_{\mathrm{str}} + f_{\mathrm{psd}} + f_{\mathrm{err}} \tag{3}$$

where f_{str} is (M, M)-structured, f_{psd} is $1/F(M)$-pseudorandom, and $\|f_{\mathrm{err}}\|_H \leq \varepsilon$. The function f_{str} is a combination of at most M characters, and thus there exists

a subspace $V \subset \mathbf{F}_2^n$ of codimension $d \leq M$ such that f_{str} is constant on all translates of V.

We have

$$\mathbf{E}_{x \in \mathbf{F}_2^n} |f_{\mathrm{err}}(x)|^2 \leq \varepsilon = \varepsilon 2^d |V|/|\mathbf{F}_2^n|.$$

Dividing $\mathbf{F}_2^n$ into 2^d translates $y+V$ of V, we thus conclude that we must have

$$\mathbf{E}_{x \in y+V} |f_{\mathrm{err}}(x)|^2 \leq \sqrt{\varepsilon} \tag{4}$$

on all but at most $\sqrt{\varepsilon} 2^d$ of the translates $y+V$.

Let $y+V$ be such that (4) holds, and let δ_{y+V} be the average of A on $y+V$. The function f_{str} equals a constant value on $y+V$, call it c_{y+V}. Averaging (3) on $y+V$ we obtain

$$\delta_{y+V} = c_{y+V} + \mathbf{E}_{x \in y+V} f_{\mathrm{psd}}(x) + \mathbf{E}_{x \in y+V} f_{\mathrm{err}}(x).$$

Since $f_{\mathrm{psd}}(x)$ is $1/F(M)$-pseudorandom, some simple Fourier analysis (expressing 1_{y+V} as an average of characters) shows that

$$|\mathbf{E}_{x \in y+V} f_{\mathrm{psd}}(x)| \leq \frac{2^n}{|V|F(M)} \leq \frac{2^M}{F(M)}$$

while from (4) and Cauchy-Schwarz we have

$$|\mathbf{E}_{x \in y+V} f_{\mathrm{err}}(x)| \leq \varepsilon^{1/4}$$

and thus

$$\delta_{y+V} = c_{y+V} + O\left(\frac{2^M}{F(M)}\right) + O(\varepsilon^{1/4}).$$

By (3) we therefore have

$$1_A(x) - \delta_{y+V} = f_{\mathrm{psd}}(x) + f_{\mathrm{err}}(x) + O\left(\frac{2^M}{F(M)}\right) + O(\varepsilon^{1/4}).$$

Now let e_ξ be an arbitrary character. By arguing as before we have

$$|\mathbf{E}_{x \in y+V} f_{\mathrm{psd}}(x) e_\xi(x)| \leq \frac{2^M}{F(M)}$$

and

$$|\mathbf{E}_{x \in y+V} f_{\mathrm{err}}(x) e_\xi(x)| \leq \varepsilon^{1/4}$$

and thus

$$\mathbf{E}_{x \in y+V} (1_A(x) - \delta_{y+V}) e_\xi(x) = O\left(\frac{2^M}{F(M)}\right) + O(\varepsilon^{1/4}).$$

If we now set $F(M) := \varepsilon^{-1/4} 2^M$ we obtain the claim. □

For some applications of this lemma, see [13]. A decomposition in a similar spirit can also be found in [5], [15]. The weak structure theorem for Reed-Muller codes was also employed in [18], [14] (under the name of a *Koopman-von Neumann type theorem*).

Now we obtain the Szemerédi regularity lemma itself. Recall that if $G = (V, E)$ is a graph and A, B are nonempty disjoint subsets of V, we say that the pair (A, B) is *ε-regular* if for any $A' \subset A, B' \subset B$ with $|A'| \geq \varepsilon |A|$ and $|B'| \geq \varepsilon |B|$, the number of edges between A' and B' differs from $\delta_{A,B}|A'||B'|$ by at most $\varepsilon |A'||B'|$, where $\delta_{A,B} = |E \cap (A \times B)|/|A||B|$ is the edge density between A and B.

Lemma 2.11 (Szemerédi regularity lemma)**.** *[24] Let $0 < \varepsilon < 1$ and $m \geq 1$. Then if $G = (V, E)$ is a graph with $|V| = n$ sufficiently large depending on ε and m, then there exists a partition $V = V_0 \cup V_1 \cup \ldots \cup V_{m'}$ with $m \leq m' \leq O_{\varepsilon,m}(1)$ such that $|V_0| \leq \varepsilon n$, $|V_1| = \ldots = |V_{m'}|$, and such that all but at most $\varepsilon (m')^2$ of the pairs (V_i, V_j) for $1 \leq i < j \leq m'$ are ε-regular.*

Proof. It will suffice to establish the weaker claim that $|V_0| = O(\varepsilon n)$, and all but at most $O(\sqrt{\varepsilon}(m')^2)$ of the pairs (V_i, V_j) are $O(\varepsilon^{1/12})$-regular. We can also assume without loss of generality that ε is small.

We apply Theorem 2.6 to the setting in Example 2.3 with $f := 1_E$ and F to be chosen later. This gives us an integer $M = O_{F,\varepsilon}(1)$ and a decomposition

$$1_E = f_{\mathrm{str}} + f_{\mathrm{psd}} + f_{\mathrm{err}} \tag{5}$$

where f_{str} is (M, M)-structured, f_{psd} is $1/F(M)$-pseudorandom, and $\|f_{\mathrm{err}}\|_H \leq \varepsilon$. The function f_{str} is a combination of at most M tensor products of indicator functions $1_{A_i \times B_i}$. The sets A_i and B_i partition V into at most 2^{2M} sets, which we shall refer to as *atoms*. If $|V|$ is sufficiently large depending on M, m and ε, we can then partition $V = V_0 \cup \ldots \cup V_{m'}$ with $m \leq m' \leq (m + 2^{2M})/\varepsilon$, $|V_0| = O(\varepsilon n)$, $|V_1| = \ldots = |V_{m'}|$, and such that each V_i for $1 \leq i \leq m'$ is entirely contained within an atom. In particular f_{str} is constant on $V_i \times V_j$ for all $1 \leq i < j \leq m'$. Since ε is small, we also have $|V_i| = \Theta(n/m')$ for $1 \leq i \leq m$.

We have

$$\mathbf{E}_{(v,w) \in V \times V} |f_{\mathrm{err}}(v, w)|^2 \leq \varepsilon$$

and in particular

$$\mathbf{E}_{1 \leq i < j \leq m'} \mathbf{E}_{(v,w) \in V_i \times V_j} |f_{\mathrm{err}}(v, w)|^2 = O(\varepsilon).$$

Then we have

$$\mathbf{E}_{(v,w) \in V_i \times V_j} |f_{\mathrm{err}}(v, w)|^2 \leq \sqrt{\varepsilon} \tag{6}$$

for all but $O(\sqrt{\varepsilon}(m')^2)$ pairs (i, j).

Let (i,j) be such that (6) holds. On $V_i \times V_j$, $f_{\rm str}$ is equal to a constant value c_{ij}. Also, from the pseudorandomness of $f_{\rm psd}$ we have

$$|\sum_{(v,w)\in A'\times B'} f_{\rm psd}(v,w)| \leq \frac{n^2}{F(M)} = O_{m,\varepsilon,M}\left(\frac{|V_i||V_j|}{F(M)}\right)$$

for all $A' \subset V_i$ and $B' \subset V_j$. By arguing very similarly to the proof of Lemma 2.10, we can conclude that the edge density δ_{ij} of E on $V_i \times V_j$ is

$$\delta_{ij} = c_{ij} + O(\varepsilon^{1/4}) + O_{m,\varepsilon,M}\left(\frac{1}{F(M)}\right)$$

and that

$$|\sum_{(v,w)\in A'\times B'} (1_E(v,w) - \delta_{ij})| = \big(O(\varepsilon^{1/4}) + O_{m,\varepsilon,M}\left(\frac{1}{F(M)}\right)\big)|V_i||V_j|$$

for all $A' \subset V_i$ and $B' \subset V_j$. This implies that the pair (V_i, V_j) is $O(\varepsilon^{1/12}) + O_{m,\varepsilon,M}(1/F(M)^{1/3})$-regular. The claim now follows by choosing F to be a sufficiently rapidly growing function of M, which depends also on m and ε. □

Similar methods can yield an alternate proof of the regularity lemma for hypergraphs [11], [12], [21], [22]; see [29]. To oversimplify enormously, one works on higher product spaces such as $V \times V \times V$, and uses partial tensor products such as $(v_1, v_2, v_3) \mapsto 1_A(v_1)1_E(v_2, v_3)$ as the structured objects. The lower-order functions such as $1_E(v_2, v_3)$ which appear in the structured component are then decomposed again by another application of structure theorems (e.g. for $1_E(v_2, v_3)$, one would use the ordinary Szemerédi regularity lemma). The ability to arbitrarily select the various functions F appearing in these structure theorems becomes crucial in order to obtain a satisfactory hypergraph regularity lemma.

See also [1] for another graph regularity lemma involving an arbitrary function F which is very similar in spirit to Theorem 2.6. In the opposite direction, if one applies the weak structure theorem (Corollary 2.5) to the product setting (Example 2.3) one obtains a "weak regularity lemma" very close to that in [6].

3. Structure and randomness in a measure space

We have seen that the Hilbert space model for separating structure from randomness is satisfactory for many applications. However, there are times when the "L^2" type of control given by this model is insufficient. A typical example arises when one wants to decompose a function $f : X \to \mathbf{R}$ on a probability space $(X, \mathbf{X}, \mu)$ into structured and pseudorandom pieces, plus a small error. Using the Hilbert space model (with $H = L^2(X)$), one can control the L^2 norm of (say) the structured component $f_{\rm str}$ by that of the original function f, indeed the construction in Theorem 2.6 ensures that $f_{\rm str}$ is an orthogonal projection of f onto a subspace generated by some vectors in S. However, in many applications one also wants to control the L^∞ norm of the structured part by that of f, and if f is non-negative one often also wishes $f_{\rm str}$ to be non-negative also. More generally, one would like a *comparison principle*: if f, g are two functions such that f dominates g pointwise (i.e. $|g(x)| \leq f(x)$), and $f_{\rm str}$ and $g_{\rm str}$ are the corresponding structured components, we would like $f_{\rm str}$ to dominate $g_{\rm str}$. One cannot deduce these facts purely from the knowledge that $f_{\rm str}$ is an orthogonal projection of f. If however we have the stronger property that $f_{\rm str}$ is a *conditional expectation* of f, then we can achieve the above objectives. This turns out to be important when establishing structure theorems for *sparse* objects, for which purely L^2 methods are inadequate; this was in particular a key point in the recent proof [16] that the primes contained arbitrarily long arithmetic progressions.

In this section we fix the probability space $(X, \mathbf{X}, \mu)$, thus $\mathbf{X}$ is a σ-algebra on the set X, and $\mu : \mathbf{X} \to [0,1]$ is a probability measure, i.e. a countably additive non-negative measure. In many applications one can assume that the σ-algebra $\mathbf{X}$ is finite, in which case it can be identified with a finite partition $X = A_1 \cup \ldots \cup A_k$ of X into *atoms* (so that $\mathbf{X}$ consists of all sets which can be expressed as the union of atoms).

Example 3.1 (Uniform distribution). If X is a finite set, $\mathbf{X} = 2^X$ is the power set of X, and $\mu(E) := |E|/|X|$ for all $E \subset X$ (i.e. μ is uniform probability measure on X), then $(X, \mathbf{X}, \mu)$ is a probability space, and the atoms are just singleton sets.

We recall the concepts of a *factor* and of *conditional expectation*, which will be fundamental to our analysis.

Definition 3.2 (Factor)**.** A *factor* of $(X, \mathbf{X}, \mu)$ is a triplet $\mathbf{Y} = (Y, \mathbf{Y}, \pi)$, where Y is a set, $\mathbf{Y}$ is a σ-algebra, and $\pi : X \to Y$ is a measurable map. If $\mathbf{Y}$ is a factor, we let $\mathcal{B}_{\mathbf{Y}} := \{\pi^{-1}(E) : E \in \mathbf{Y}\}$ be the sub-σ-algebra of $\mathbf{X}$ formed by pulling back $\mathbf{Y}$ by π. A function $f : X \to \mathbf{R}$ is said to be $\mathbf{Y}$*-measurable* if it is measurable with respect to $\mathcal{B}_{\mathbf{Y}}$. If $f \in L^2(X, \mathbf{X}, \mu)$, we let $\mathbf{E}(f|Y) = \mathbf{E}(f|\mathcal{B}_Y)$ be the orthogonal projection of f to the closed subspace $L^2(X, \mathcal{B}_Y, \mu)$ of $L^2(X, \mathbf{X}, \mu)$ consisting of $\mathbf{Y}$-measurable functions. If $\mathbf{Y} = (Y, \mathbf{Y}, \pi)$ and $\mathbf{Y}' = (Y', \mathbf{Y}', \pi')$ are two factors, we let $\mathbf{Y} \vee \mathbf{Y}'$ denote the factor $(Y \times Y', \mathbf{Y} \otimes \mathbf{Y}', \pi \oplus \pi')$.

Example 3.3 (Colourings). Let X be a finite set, which we give the uniform distribution as in Example 3.1. Suppose we *colour* this set using some finite *palette* Y by introducing a map $\pi : X \to Y$. If we endow Y with the discrete σ-algebra $\mathbf{Y} = 2^Y$, then $(Y, \mathbf{Y}, \pi)$ is a factor of $(X, \mathbf{X}, \mu)$. The σ-algebra $\mathcal{B}_{\mathbf{Y}}$ is then generated by the *colour classes* $\pi^{-1}(y)$ of the colouring π. The expectation $\mathbf{E}(f|Y)$ of a function $f : X \to \mathbf{R}$ is then given by the formula $\mathbf{E}(f|Y)(x) := \mathbf{E}_{x' \in \pi^{-1}(\pi(x))} f(x')$ for all $x \in X$, where $\pi^{-1}(\pi(x))$ is the colour class that x lies in.

In the previous section, the concept of structure was represented by a set S of vectors. In this section, we shall instead represent structure by a collection $\mathcal{S}$ of *factors*. We say that a factor $\mathbf{Y}$ has *complexity* at most M if it is the join $\mathbf{Y} = \mathbf{Y}_1 \vee \ldots \vee \mathbf{Y}_m$ of m factors from $\mathcal{S}$ for some $0 \leq m \leq M$. We also say that a function $f \in L^2(X)$ is *ε-pseudorandom* if we have $\|\mathbf{E}(f|\mathbf{Y})\|_{L^2(X)} \leq \varepsilon$ for all $\mathbf{Y} \in \mathcal{S}$. We have an analogue of Lemma 2.4:

Lemma 3.4 (Lack of pseudorandomness implies energy increment)**.** *Let $(X, \mathbf{X}, \mu)$ and $\mathcal{S}$ be as above. Let $f \in L^2(X)$ be such that $f - \mathbf{E}(f|\mathbf{Y})$ is* not *ε-pseudorandom for some $0 < \varepsilon \leq 1$ and some factor $\mathbf{Y}$. Then there exists $\mathbf{Y}' \in \mathcal{S}$ such that $\|\mathbf{E}(f|\mathbf{Y} \vee \mathbf{Y}')\|^2_{L^2(X)} \geq \|\mathbf{E}(f|\mathbf{Y})\|^2_{L^2(X)} + \varepsilon^2$.*

Proof. By hypothesis we have

$$\|\mathbf{E}(f - \mathbf{E}(f|\mathbf{Y})|\mathbf{Y}')\|^2_{L^2(X)} \geq \varepsilon^2$$

for some $\mathbf{Y}' \in \mathcal{S}$. By Pythagoras' theorem, this implies that

$$\|\mathbf{E}(f - \mathbf{E}(f|\mathbf{Y})|\mathbf{Y} \vee \mathbf{Y}')\|^2_{L^2(X)} \geq \varepsilon^2.$$

By Pythagoras' theorem again, the left-hand side is $\|\mathbf{E}(f|\mathbf{Y} \vee \mathbf{Y}')\|^2_{L^2(X)} - \|\mathbf{E}(f|\mathbf{Y})\|^2_{L^2(X)}$, and the claim follows. □

We then obtain an analogue of Lemma 2.7:

Lemma 3.5 (Weak structure theorem)**.** *Let $(X, \mathbf{X}, \mu)$ and $\mathcal{S}$ be as above. Let $f \in L^2(X)$ be such that $\|f\|_{L^2(X)} \leq 1$, let $\mathbf{Y}$ be a factor, and let $0 < \varepsilon \leq 1$. Then there exists a decomposition $f = f_{\mathrm{str}} + f_{\mathrm{psd}}$, where $f_{\mathrm{str}} = \mathbf{E}(f|\mathbf{Y} \vee \mathbf{Y}')$ for some factor $\mathbf{Y}'$ of complexity at most $1/\varepsilon^2$, and f_{psd} is ε-pseudorandom.*

Proof. We construct factors $\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_m \in \mathcal{S}$ by the following algorithm:

- Step 0: Initialise $m = 0$.
- Step 1: Write $\mathbf{Y}' := \mathbf{Y}_1 \vee \ldots \vee \mathbf{Y}_m$, $f_{\mathrm{str}} := \mathbf{E}(f|\mathbf{Y} \vee \mathbf{Y}')$, and $f_{\mathrm{psd}} := f - f_{\mathrm{str}}$.
- Step 2: If f_{psd} is ε-pseudorandom then **STOP**. Otherwise, by Lemma 3.4 we can find $\mathbf{Y}_{m+1} \in \mathcal{S}$ such that $\|\mathbf{E}(f|\mathbf{Y} \vee \mathbf{Y}' \vee \mathbf{Y}_{m+1})\|^2_{L^2(X)} \geq \|\mathbf{E}(f|\mathbf{Y} \vee \mathbf{Y}')\|^2_{L^2(X)} + \varepsilon^2$.
- Step 3: Increment m to $m + 1$ and return to Step 1.

Since the "energy" $\|f_{\mathrm{str}}\|^2_{L^2(X)}$ ranges between 0 and 1 (by the hypothesis $\|f\|_{L^2(X)} \leq 1$) and increments by ε^2 at each stage, we see that this algorithm terminates in at most $1/\varepsilon^2$ steps. The claim follows. □

Iterating this we obtain an analogue of Theorem 2.6:

Theorem 3.6 (Strong structure theorem)**.** *Let $(X, \mathbf{X}, \mu)$ and $\mathcal{S}$ be as above. Let $f \in L^2(X)$ be such that $\|f\|_{L^2(X)} \leq 1$, let $\varepsilon > 0$, and let $F : \mathbf{Z}^+ \to \mathbf{R}^+$ be an arbitrary function. Then we can find an integer $M = O_{F,\varepsilon}(1)$ and a decomposition* (2) *where $f_{\mathrm{str}} = \mathbf{E}(f|\mathbf{Y})$ for some factor $\mathbf{Y}$ of complexity at most M, f_{psd} is $1/F(M)$-pseudorandom, and f_{err} has norm at most ε.*

Proof. Without loss of generality we may assume $F(M) \geq 2M$. Also, it will suffice to allow $\mathbf{Y}$ to have complexity $O(M)$ rather than M.

We recursively define $M_0 := 1$ and $M_i := F(M_{i-1})^2$ for all $i \geq 1$. We then recursively define factors $\mathbf{Y}_0, \mathbf{Y}_1, \mathbf{Y}_2, \ldots$ by setting $\mathbf{Y}_0$ to be the trivial factor, and then for each $i \geq 1$ using Lemma 2.7 to find a factor $\mathbf{Y}'_i$ of complexity at most M_i such that $f - \mathbf{E}(f|\mathbf{Y}_{i-1} \vee \mathbf{Y}'_i)$ is $1/F(M_{i-1})$-pseudorandom, and then setting $\mathbf{Y}_i := \mathbf{Y}_{i-1} \vee \mathbf{Y}'_i$. By Pythagoras' theorem and the hypothesis $\|f\|_{L^2(X)} \leq 1$, the energy $\|\mathbf{E}(f|\mathbf{Y}_i)\|^2_{L^2(X)}$ is increasing in i, and is bounded between 0 and 1. By the pigeonhole principle, we can thus find $1 \leq i \leq 1/\varepsilon^2 + 1$ such that $\|\mathbf{E}(f|\mathbf{Y}_i)\|^2_{L^2(X)} - \|\mathbf{E}(f|\mathbf{Y}_{i-1})\|^2_{L^2(X)} \leq \varepsilon^2$; by Pythagoras' theorem, this implies that $\|\mathbf{E}(f|\mathbf{Y}_i) - \mathbf{E}(f|\mathbf{Y}_{i-1})\|_{L^2(X)} \leq \varepsilon$. If we then set $f_{\mathrm{str}} := \mathbf{E}(f|\mathbf{Y}_{i-1})$, $f_{\mathrm{psd}} := f - \mathbf{E}(f|\mathbf{Y}_i)$, $f_{\mathrm{err}} := \mathbf{E}(f|\mathbf{Y}_i) - \mathbf{E}(f|\mathbf{Y}_{i-1})$, and $M := M_{i-1}$, we obtain the claim. □

This theorem can be used to give alternate proofs of Lemma 2.10 and Lemma 2.11; we leave this as an exercise to the reader (but see [25] for a proof of Lemma 2.11 essentially relying on Theorem 3.6).

As mentioned earlier, the key advantage of these types of structure theorems is that the structured component f_{str} is now obtained as a conditional expectation of the original function f rather than merely an orthogonal projection, and so one has good "L^1" and "L^∞" control on f_{str} rather than just L^2 control. In particular, these structure theorems are good for controlling *sparsely supported functions* f (such as the normalised indicator function of a sparse set), by obtaining a densely supported function f_{str} which models the behaviour of f in some key respects. Let us give a simplified "sparse structure theorem" which is too restrictive for real applications, but which serves to illustrate the main concept.

Theorem 3.7 (Sparse structure theorem, toy version)**.** *Let $0 < \varepsilon < 1$, let $F : \mathbf{Z}^+ \to \mathbf{R}^+$ be a function, and let N*

be an integer parameter. Let $(X, \mathbf{X}, \mu)$ and $\mathcal{S}$ be as above, and depending on N. Let $\nu \in L^1(X)$ be a non-negative function (also depending on N) with the property that for every $M \geq 0$, we have the "pseudorandomness" property

$$\|\mathbf{E}(\nu|\mathbf{Y})\|_{L^\infty(X)} \leq 1 + o_M(1) \tag{7}$$

for all factors $\mathbf{Y}$ of complexity at most M, where $o_M(1)$ is a quantity which goes to zero as N goes to infinity for any fixed M. Let $f : X \to \mathbf{R}$ (which also depends on N) obey the pointwise estimate $0 \leq f(x) \leq \nu(x)$ for all $x \in X$. Then, if N is sufficiently large depending on F and ε, we can find an integer $M = O_{F,\varepsilon}(1)$ and a decomposition (2) *where $f_{\rm str} = \mathbf{E}(f|\mathbf{Y})$ for some factor $\mathbf{Y}$ of complexity at most M, $f_{\rm psd}$ is $1/F(M)$-pseudorandom, and $f_{\rm err}$ has norm at most ε. Furthermore, we have*

$$0 \leq f_{\rm str}(x) \leq 1 + o_{F,\varepsilon}(1) \tag{8}$$

and

$$\int_X f_{\rm str}\, d\mu = \int_X f\, d\mu. \tag{9}$$

An example to keep in mind is where $X = \{1, \ldots, N\}$ with the uniform probability measure μ, $\mathcal{S}$ consists of the σ-algebras generated by a single discrete interval $\{n \in \mathbf{Z} : a \leq n \leq b\}$ for $1 \leq a \leq b \leq N$, and ν being the function $\nu(x) = \log N 1_A(x)$, where A is a randomly chosen subset of $\{1, \ldots, N\}$ with $\P(x \in A) = \frac{1}{\log N}$ for all $1 \leq x \leq N$; one can then verify (7) with high probability using tools such as Chernoff's inequality. Observe that ν is bounded in $L^1(X)$ uniformly in N, but is unbounded in $L^2(X)$. Very roughly speaking, the above theorem states that any dense subset B of A can be effectively "modelled" in some sense by a dense subset of $\{1, \ldots, N\}$, normalised by a factor of $\frac{1}{\log N}$; this can be seen by applying the above theorem to the function $f := \log N 1_B(x)$.

Proof. We run the proof of Lemma 3.5 and Theorem 3.6 again. Observe that we no longer have the bound $\|f\|_{L^2(X)} \leq 1$. However, from (7) and the pointwise bound $0 \leq f \leq \nu$ we know that

$$\begin{aligned}\|\mathbf{E}(f|\mathbf{Y})\|_{L^2(X)} &\leq \|\mathbf{E}(\nu|\mathbf{Y})\|_{L^2(X)} \\ &\leq \|\mathbf{E}(\nu|\mathbf{Y})\|_{L^\infty(X)} \\ &\leq 1 + o_M(1)\end{aligned}$$

for all $\mathbf{Y}$ of complexity at most M. In particular, for N large enough depending on M we have

$$\|\mathbf{E}(f|\mathbf{Y})\|^2_{L^2(X)} \leq 2 \tag{10}$$

(say). This allows us to obtain an analogue of Lemma 3.5 as before (with slightly worse constants), assuming that N is sufficiently large depending on ε, by repeating the proof more or less verbatim. One can then repeat the proof of Theorem 3.6, again using (10), to obtain the desired decomposition. The claim (8) follows immediately from (7), and (9) follows since $\int_X \mathbf{E}(f|\mathbf{Y})\, d\mu = \int_X f\, d\mu$ for any factor $\mathbf{Y}$. □

Remark 3.8. In applications, one does not quite have the property (7); instead, one can bound $\mathbf{E}(\nu|\mathbf{Y})$ by $1 + o_M(1)$ outside of a small exceptional set, which has measure $o(1)$ with respect to μ and ν. In such cases it is still possible to obtain a structure theorem similar to Theorem 3.7; see [16, Theorem 8.1], [26, Theorem 3.9], or [34, Theorem 4.7]. These structure theorems have played an indispensable role in establishing the existence of patterns (such as arithmetic progressions) inside sparse sets such as the prime numbers, by viewing them as dense subsets of sparse pseudorandom sets (such as the *almost prime* numbers), and then appealing to a sparse structure theorem to model the original set by a much denser set, to which one can apply deep theorems (such as Szemerédi's theorem [24]) to detect the desired pattern.

The reader may observe one slight difference between the concept of pseudorandomness discussed here, and the concept in the previous section. Here, a function $f_{\rm psd}$ is considered pseudorandom if its conditional expectations $\mathbf{E}(f_{\rm psd}|\mathbf{Y})$ are small for various structured $\mathbf{Y}$. In the previous section, a function $f_{\rm psd}$ is considered pseudorandom if its correlations $\langle f_{\rm psd}, g\rangle_H$ were small for various structured g. However, it is possible to relate the two notions of pseudorandomness by the simple device of using a structured function g to generate a structured factor $\mathbf{Y}_g$. In measure theory, this is usually done by taking the level sets $g^{-1}([a,b])$ of g and seeing what σ-algebra they generate. In many quantitative applications, though, it is too expensive to take *all* of these the level sets, and so instead one only takes a finite number of these level sets to create the relevant factor. The following lemma illustrates this construction:

Lemma 3.9 (Correlation with a function implies non-trivial projection)**.** *Let $(X, \mathbf{X}, \mu)$ be a probability space. Let $f \in L^1(X)$ and $g \in L^2(X)$ be such that $\|f\|_{L^1(X)} \leq 1$ and $\|g\|_{L^2(X)} \leq 1$. Let $\varepsilon > 0$ and $0 \leq \alpha < 1$, and let $\mathbf{Y}$ be the factor $\mathbf{Y} = (\mathbf{R}, \mathbf{Y}, g)$, where $\mathbf{Y}$ is the σ-algebra generated by the intervals $[(n+\alpha)\varepsilon, (n+1+\alpha)\varepsilon)$ for $n \in \mathbf{Z}$. Then we have*

$$\|\mathbf{E}(f|\mathbf{Y})\|_{L^2(X)} \geq |\langle f, g\rangle_{L^2(X)}| - \varepsilon.$$

Proof. Observe that the atoms of $\mathcal{B}_{\mathbf{Y}}$ are generated by level sets $g^{-1}([(n+\alpha)\varepsilon, (n+1+\alpha)\varepsilon))$, and on these level sets g fluctuates by at most ε. Thus

$$\|g - \mathbf{E}(g|\mathbf{Y})\|_{L^\infty(X)} \leq \varepsilon.$$

Since $\|f\|_{L^1(X)} \leq 1$, we conclude

$$\big|\langle f, g\rangle_{L^2(X)} - \langle f, \mathbf{E}(g|\mathbf{Y})\rangle_{L^2(X)}\big| \leq \varepsilon.$$

On the other hand, by Cauchy-Schwarz and the hypothesis $\|g\|_{L^2(X)} \leq 1$ we have

$$\begin{aligned}|\langle f, \mathbf{E}(g|\mathbf{Y})\rangle_{L^2(X)}| &= |\langle \mathbf{E}(f|\mathbf{Y}), g\rangle_{L^2(X)}| \\ &\leq \|\mathbf{E}(f|\mathbf{Y})\|_{L^2(X)}.\end{aligned}$$

The claim follows. □

This type of lemma is relied upon in the above-mentioned papers [16], [26], [34] to convert pseudorandomness in the conditional expectation sense to pseudorandomness in the correlation sense. In applications it is also convenient to randomise the shift parameter α in order to average away all boundary effects; see e.g. [32, Lemma 3.6].

4. Structure and randomness via uniformity norms

In the preceding sections, we specified the notion of structure (either via a set S of vectors, or a collection $\mathcal{S}$ of factors), which then created a dual notion of pseudorandomness for which one had a structure theorem. Such decompositions give excellent control on the structured component f_{str} of the function, but the control on the pseudorandom part f_{psd} can be rather weak. There is an opposing approach, in which one first specifies the notion of pseudorandomness one would like to have for f_{psd}, and then works as hard as one can to obtain a useful corresponding notion of structure. In this approach, the pseudorandom component f_{psd} is easy to dispose of, but then all the difficulty gets shifted to getting an adequate control on the structured component.

A particularly useful family of notions of pseudorandomness arises from the *Gowers uniformity norms* $\|f\|_{U^d(G)}$. These norms can be defined on any finite additive group G, and for complex-valued functions $f : G \to \mathbf{C}$, but for simplicity let us restrict attention to a Hamming cube $G = \mathbf{F}_2^n$ and to real-valued functions $f : \mathbf{F}_2^n \to \mathbf{R}$. (For more general groups and complex-valued functions, see [33]. For applications to graphs and hypergraphs, one can use the closely related *Gowers box norms*; see [11], [12], [20], [26], [30], [33].) In that case, the uniformity norm $\|f\|_{U^d(\mathbf{F}_2^n)}$ can be defined for $d \geq 1$ by the formula

$$\|f\|_{U^d(\mathbf{F}_2^n)}^{2^d} := \mathbf{E}_{L:\mathbf{F}_2^d \to \mathbf{F}_2^n} \prod_{a \in \mathbf{F}_2^d} f(L(a))$$

where L ranges over all affine-linear maps from $\mathbf{F}_2^d$ to $\mathbf{F}_2^n$ (not necessarily injective). For instance, we have

$$\begin{aligned}\|f\|_{U^1(\mathbf{F}_2^n)} &= |\mathbf{E}_{x,h \in \mathbf{F}_2^n} f(x) f(x+h)|^{1/2} \\ &= |\mathbf{E}_{x \in \mathbf{F}_2^n} f(x)| \\ \|f\|_{U^2(\mathbf{F}_2^n)} &= |\mathbf{E}_{x,h,k \in \mathbf{F}_2^n} f(x) f(x+h) f(x+k) \\ &\quad \times f(x+h+k)|^{1/4} \\ &= |\mathbf{E}_{h \in \mathbf{F}_2^n} |\mathbf{E}_{x \in \mathbf{F}_2^n} f(x) f(x+h)|^2|^{1/4} \\ \|f\|_{U^3(\mathbf{F}_2^n)} &= |\mathbf{E}_{x,h_1,h_2,h_3 \in \mathbf{F}_2^n} f(x) f(x+h_1) f(x+h_2) \\ &\quad \times f(x+h_3) f(x+h_1+h_2) f(x+h_1+h_3) \\ &\quad \times f(x+h_2+h_3) f(x+h_1+h_2+h_3)|^{1/8}.\end{aligned}$$

It is possible to show that the norms $\|\|_{U^d(\mathbf{F}_2^n)}$ are indeed a norm for $d \geq 2$, and a semi-norm for $d = 1$; see e.g. [33]. These norms are also monotone in d:

$$0 \leq \|f\|_{U^1(\mathbf{F}_2^n)} \leq \|f\|_{U^2(\mathbf{F}_2^n)} \leq \|f\|_{U^3(\mathbf{F}_2^n)} \leq \ldots \leq \|f\|_{L^\infty(\mathbf{F}_2^n)}. \tag{11}$$

The $d = 2$ norm is related to the Fourier coefficients $\hat{f}(\xi)$ defined in (1) by the important (and easily verified) identity

$$\|f\|_{U^2(\mathbf{F}_2^n)} = (\sum_{\xi \in \mathbf{F}_2^n} |\hat{f}(\xi)|^4)^{1/4}. \tag{12}$$

More generally, the uniformity norms $\|f\|_{U^d(\mathbf{F}_2^n)}$ for $d \geq 1$ are related to Reed-Muller codes of order $d-1$ (although this is partly conjectural for $d \geq 4$), but the relationship cannot be encapsulated in an identity as elegant as (12) once $d \geq 3$. We will return to this point shortly.

Let us informally call a function $f : \mathbf{F}_2^n \to \mathbf{R}$ *pseudorandom of order* $d-1$ if $\|f\|_{U^d(\mathbf{F}_2^n)}$ is small; thus for instance functions with small U^2 norm are *linearly pseudorandom* (or *Fourier-pseudorandom*, functions with small U^3 norm are *quadratically pseudorandom*, and so forth. It turns out that functions which are pseudorandom to a suitable order become negligible for the purpose of various multilinear correlations (and the higher the order of pseudorandomness, the more complex the multilinear correlations that become negligible). This can be demonstrated by repeated application of the Cauchy-Schwarz inequality. We give a simple instance of this:

Lemma 4.1 (Generalised von Neumann theorem)**.** *Let $T_1, T_2 : \mathbf{F}_n^2 \to \mathbf{F}_2^n$ be invertible linear transformations such that $T_1 - T_2$ is also invertible. Then for any $f, g, h : \mathbf{F}_n^2 \to [-1, 1]$ we have*

$$|\mathbf{E}_{x,r \in \mathbf{F}_2^n} f(x) g(x + T_1 r) h(x + T_2 r)| \leq \|f\|_{U^2(\mathbf{F}_2^n)}.$$

Proof. By changing variables $r' := T_2 r$ if necessary we may assume that T_2 is the identity map I. We rewrite the left-hand side as

$$|\mathbf{E}_{x \in \mathbf{F}_2^n} h(x) \mathbf{E}_{r \in \mathbf{F}_2^n} f(x - r) g(x + (T_1 - I) r)|$$

and then use Cauchy-Schwarz to bound this from above by

$$(\mathbf{E}_{x\in\mathbf{F}_2^n}|\mathbf{E}_{r\in\mathbf{F}_2^n} f(x-r)g(x+(T_1-I)r)|^2)^{1/2}$$

which one can rewrite as

$$|\mathbf{E}_{x,r,r'\in\mathbf{F}_2^n} f(x-r)f(x-r')g(x+(T_1-I)r)g(x+(T_1-I)r')|^{1/2};$$

applying the change of variables $(y,s,h) := (x+(T_1-I)r, T_1 r, r-r')$, this can be rewritten as

$$|\mathbf{E}_{y,h\in\mathbf{F}_2^n} g(y)g(y+(T_1-I)h)\mathbf{E}_{s\in\mathbf{F}_2^n} f(y+s)f(y+s+h)|^{1/2};$$

applying Cauchy-Schwarz, again, one can bound this by

$$\left|\mathbf{E}_{y,h\in\mathbf{F}_2^n}|\mathbf{E}_{s\in\mathbf{F}_2^n} f(y+s)f(y+s+h)|^2\right|^{1/4}.$$

But this is equal to $\|f\|_{U^2(\mathbf{F}_2^n)}$, and the claim follows. □

For a more systematic study of such "generalised von Neumann theorems", including some weighted versions, see Appendices B and C of [19].

In view of these generalised von Neumann theorems, it is of interest to locate conditions which would force a Gowers uniformity norm $\|f\|_{U^d(\mathbf{F}_2^n)}$ to be small. We first give a "soft" characterisation of this smallness, which at first glance seems too trivial to be of any use, but is in fact powerful enough to establish Szemerédi's theorem (see [28]) as well as the Green-Tao theorem [16]. It relies on the obvious identity

$$\|f\|_{U^d(\mathbf{F}_2^n)}^{2^d} = \langle f, \mathcal{D}f\rangle_{L^2(\mathbf{F}_2^n)}$$

where the *dual function* $\mathcal{D}f$ of f is defined as

$$\mathcal{D}f(x) := \mathbf{E}_{L:\mathbf{F}_2^d\to\mathbf{F}_2^n; L(0)=x} \prod_{a\in\mathbf{F}_2^d\backslash\{0\}} f(L(a)). \quad (13)$$

As a consequence, we have

Lemma 4.2 (Dual characterisation of pseudorandomness)**.** *Let S denote the set of all dual functions $\mathcal{D}F$ with $\|F\|_{L^\infty(\mathbf{F}_2^n)} \leq 1$. Then if $f : \mathbf{F}_2^n \to [-1,1]$ is such that $\|f\|_{U^d(\mathbf{F}_2^n)} \geq \varepsilon$ for some $0 < \varepsilon \leq 1$, then we have $\langle f, g\rangle \geq \varepsilon^{2^d}$ for some $g \in S$.*

In the converse direction, one can use the *Cauchy-Schwarz-Gowers inequality* (see e.g. [10], [16], [19], [33]) to show that if $\langle f, g\rangle \geq \varepsilon$ for some $g \in S$, then $\|f\|_{U^d(\mathbf{F}_2^n)} \geq \varepsilon$.

The above lemma gives a "soft" way to detect pseudorandomness, but is somewhat unsatisfying due to the rather non-explicit description of the "structured" set S. To investigate pseudorandomness further, observe that we have the recursive identity

$$\|f\|_{U^d(\mathbf{F}_2^n)}^{2^d} = \mathbf{E}_{h\in\mathbf{F}_2^n}\|ff_h\|_{U^{d-1}(\mathbf{F}_2^n)}^{2^{d-1}} \quad (14)$$

(which, incidentally, can be used to quickly deduce the monotonicity (11)). From this identity and induction we quickly deduce the modulation symmetry

$$\|fg\|_{U^d(\mathbf{F}_2^n)} = \|f\|_{U^d(\mathbf{F}_2^n)} \quad (15)$$

whenever $g \in S_{d-1}(\mathbf{F}_2^n)$ is a Reed-Muller code of order at most $d-1$. In particular, we see that $\|g\|_{U^d(\mathbf{F}_2^n)} = 1$ for such codes; thus a code of order $d-1$ or less is definitely *not* pseudorandom of order d. A bit more generally, by combining (15) with (11) we see that

$$|\langle f, g\rangle_{L^2(\mathbf{F}_2^n)}| = \|fg\|_{U^1(\mathbf{F}_2^n)} \leq \|fg\|_{U^d(\mathbf{F}_2^n)} = \|f\|_{U^d(\mathbf{F}_2^n)}.$$

In particular, any function which has a large correlation with a Reed-Muller code $g \in S_{d-1}(\mathbf{F}_2^n)$ is not pseudorandom of order d. It is conjectured that the converse is also true:

Conjecture 4.3 (Gowers inverse conjecture for $\mathbf{F}_2^n$)**.** *If $d \geq 1$ and $\varepsilon > 0$ then there exists $\delta > 0$ with the following property: given any $n \geq 1$ and any $f : \mathbf{F}_2^n \to [-1,1]$ with $\|f\|_{U^d(\mathbf{F}_2^n)} \geq \varepsilon$, there exists a Reed-Muller code $g \in S_{d-1}(\mathbf{F}_2^n)$ of order at most $d-1$ such that $|\langle f, g\rangle_{L^2(\mathbf{F}_2^n)}| \geq \delta$.*

This conjecture, if true, would allow one to apply the machinery of previous sections and then decompose a bounded function $f : \mathbf{F}_2^n \to [-1,1]$ (or a function dominated by a suitably pseudorandom function ν) into a function $f_{\rm str}$ which was built out of a controlled number of Reed-Muller codes of order at most $d-1$, a function $f_{\rm psd}$ which was pseudorandom of order d, and a small error. See for instance [14] for further discussion.

The Gowers inverse conjecture is trivial to verify for $d = 1$. For $d = 2$ the claim follows quickly from the identity (12) and the Plancherel identity

$$\|f\|_{L^2(\mathbf{F}_2^n)}^2 = \sum_{\xi\in\mathbf{F}_2^n} |\hat{f}(\xi)|^2.$$

The conjecture for $d = 3$ was first established by Samorodnitsky [23], using ideas from [9] (see also [17], [33] for related results). The conjecture for $d > 3$ remains open; a key difficulty here is that there are a huge number of Reed-Muller codes (about $2^{\Omega(n^{d-1})}$ or so, compared to the dimension 2^n of $L^2(\mathbf{F}_2^n)$) and so we definitely do not have the type of orthogonality that one enjoys in the Fourier case $d = 2$. For related reasons, we do not expect any identity of the form (12) for $d > 3$ which would allow the very few Reed-Muller codes which correlate with f to dominate the enormous number of Reed-Muller codes which do not in the right-hand side.

However, we can present some evidence for it here in the "99%-structured" case when ε is very close to 1. Let us first handle the case when $\varepsilon = 1$:

Proposition 4.4 (100%-structured inverse theorem)**.** *Suppose $d \geq 1$ and $f : \mathbf{F}_2^n \to [-1,1]$ is such that $\|f\|_{U^d(\mathbf{F}_2^n)} = 1$. Then f is a Reed-Muller code of order at most $d-1$.*

Proof. We induct on d. The case $d = 1$ is obvious. Now suppose that $d \geq 2$ and that the claim has already been proven for $d-1$. If $\|f\|_{U^d(\mathbf{F}_2^n)} = 1$, then from (14) we have

$$\mathbf{E}_{h \in \mathbf{F}_2^n} \|ff_h\|_{U^{d-1}(\mathbf{F}_2^n)}^{2^{d-1}} = 1.$$

On the other hand, from (11) we have $\|ff_h\|_{U^{d-1}(\mathbf{F}_2^n)} \leq 1$ for all h. This forces $\|ff_h\|_{U^{d-1}(\mathbf{F}_2^n)} = 1$ for all h. By induction hypothesis, ff_h must therefore be a Reed-Muller code of order at most $d-2$ for all h. Thus for every h there exists a polynomial $P_h : \mathbf{F}_2^n \to \mathbf{F}_2$ of degree at most $d-2$ such that

$$f(x+h) = f(x)(-1)^{P_h(x)}$$

for all $x, h \in \mathbf{F}_2^n$. From this one can quickly establish by induction that for every $0 \leq m \leq n$, the function f is a Reed-Muller code of degree at most $d-1$ on $\mathbf{F}_2^m$ (viewed as a subspace of $\mathbf{F}_2^n$), and the claim follows. □

To handle the case when ε is very close to 1 is trickier (we can no longer afford an induction on dimension, as was done in the above proof). We first need a rigidity result.

Proposition 4.5 (Rigidity of Reed-Muller codes)**.** *For every $d \geq 1$ there exists $\varepsilon > 0$ with the following property: if $n \geq 1$ and $f \in S_{d-1}(\mathbf{F}_2^n)$ is a Reed-Muller code of order at most $d-1$ such that $\mathbf{E}_{x \in \mathbf{F}_2^n} f(x) \geq 1 - \varepsilon$, then $f \equiv 1$.*

Proof. We again induct on d. The case $d = 1$ is obvious, so suppose $d \geq 2$ and that the claim has already been proven for $d-1$. If $\mathbf{E}_{x \in \mathbf{F}_2^n} f(x) \geq 1-\varepsilon$, then $\mathbf{E}_{x \in \mathbf{F}_2^n} |1 - f(x)| \leq \varepsilon$. Using the crude bound $|1 - ff_h| = O(|1-f| + |1-f_h|)$ we conclude that $\mathbf{E}_{x \in \mathbf{F}_2^n} |1 - ff_h(x)| \leq O(\varepsilon)$, and thus

$$\mathbf{E}_{x \in \mathbf{F}_2^n} ff_h(x) \geq 1 - O(\varepsilon)$$

for every $h \in \mathbf{F}_2^n$. But ff_h is a Reed-Muller code of order $d-2$, thus by induction hypothesis we have $ff_h \equiv 1$ for all h if ε is small enough. This forces f to be constant; but since f takes values in $\{-1, +1\}$ and has average at least $1 - \varepsilon$, we have $f \equiv 1$ as desired for ε small enough. □

Proposition 4.6 (99%-structured inverse theorem)**.** *[2] For every $d \geq 1$ and $0 < \varepsilon < 1$ there exists $0 < \delta < 1$ with the following property: if $n \geq 1$ and $f : \mathbf{F}_2^n \to [-1,1]$ is such that $\|f\|_{U^d(\mathbf{F}_2^n)} \geq 1 - \delta$, then there exists a Reed-Muller code $g \in S_{d-1}(\mathbf{F}_2^n)$ such that $\langle f, g \rangle_{L^2(\mathbf{F}_2^n)} \geq 1 - \varepsilon$.*

Proof. We again induct on d. The case $d = 1$ is obvious, so suppose $d \geq 2$ and that the claim has already been proven for $d-1$. Fix ε, let δ be a small number (depending on d and ε) to be chosen later, and suppose $f : \mathbf{F}_2^n \to [-1,1]$ is such that $\|f\|_{U^d(\mathbf{F}_2^n)} \geq 1 - \delta$. We will use $o(1)$ to denote any quantity which goes to zero as $\delta \to 0$, thus $\|f\|_{U^d(\mathbf{F}_2^n)} \geq 1 - o(1)$. We shall say that a statement is true for *most* $x \in \mathbf{F}_2^n$ if it is true for a proportion $1 - o(1)$ of values $x \in \mathbf{F}_2^n$.

Applying (14) we have

$$\mathbf{E}_{h \in \mathbf{F}_2^n} \|ff_h\|_{U^d(\mathbf{F}_2^n)} \geq 1 - o(1)$$

while from (11) we have $\|ff_h\|_{U^d(\mathbf{F}_2^n)} \leq 1$. Thus we have $\|ff_h\|_{U^d(\mathbf{F}_2^n)} = 1 - o(1)$ for all h in a subset H of $\mathbf{F}_2^n$ of density $1 - o(1)$. Applying the inductive hypothesis, we conclude that for all $h \in H$ there exists a polynomial $P_h : \mathbf{F}_2^n \to \mathbf{F}_2$ of degree at most $d-2$ such that

$$\mathbf{E}_{x \in \mathbf{F}_2^n} f(x) f(x+h) (-1)^{P_h(x)} \geq 1 - o(1).$$

Since f is bounded in magnitude by 1, this implies for each $h \in H$ that

$$f(x+h) = f(x)(-1)^{P_h(x)} + o(1) \tag{16}$$

for most x. For similar reasons it also implies that $|f(x)| = 1 + o(1)$ for most x.

Now suppose that $h_1, h_2, h_3, h_4 \in H$ form an *additive quadruple* in the sense that $h_1 + h_2 = h_3 + h_4$. Then from (16) we see that

$$f(x+h_1+h_2) = f(x)(-1)^{P_{h_1}(x) + P_{h_2}(x+h_1)} + o(1) \tag{17}$$

for most x, and similarly

$$f(x+h_3+h_4) = f(x)(-1)^{P_{h_3}(x) + P_{h_4}(x+h_3)} + o(1)$$

for most x. Since $|f(x)| = 1 + o(1)$ for most x, we conclude that

$$(-1)^{P_{h_1}(x) + P_{h_2}(x+h_1) - P_{h_3}(x) - P_{h_4}(x+h_3)} = 1 + o(1)$$

for most x. In particular, the average of the left-hand side in x is $1 - o(1)$. Applying Lemma 4.5 (and assuming δ small enough), we conclude that the left-hand side is *identically* 1, thus

$$P_{h_1}(x) + P_{h_2}(x+h_1) = P_{h_3}(x) + P_{h_4}(x+h_3) \tag{18}$$

for all additive quadruples $h_1 + h_2 = h_3 + h_4$ in H and all x.

Now for any $k \in \mathbf{F}_2^n$, define the quantity $Q(k) \in \mathbf{F}_2$ by the formula

$$Q(k) := P_{h_1}(0) + P_{h_2}(h_1) \tag{19}$$

whenever $h_1, h_2 \in H$ are such that $h_1 + h_2 \in H$. Note that the existence of such an h_1, h_2 is guaranteed since most h lie in H, and (18) ensures that the right-hand side of (19) does not depend on the exact choice of h_1, h_2 and so Q is well-defined.

Now let $x \in \mathbf{F}_2^n$ and $h \in H$. Then, since most elements of $\mathbf{F}_2^n$ lie in H, we can find $r_1, r_2, s_1, s_2 \in H$ such that $r_1 + r_2 = x$ and $s_1 + s_2 = x + h$. From (17) we see that

$$f(y+x) = f(y+r_1+r_2) = f(y)(-1)^{P_{r_1}(y)+P_{r_2}(y+r_1)}+o(1)$$

and

$$f(y+x+h) = f(y+s_1+s_2) = f(y)(-1)^{P_{s_1}(y)+P_{s_2}(y+s_1)}+o(1)$$

for most y. Also from (16)

$$f(y+x+h) = f(y+x)(-1)^{P_h(y+x)} + o(1)$$

for most y. Combining these (and the fact that $|f(y)| = 1 + o(1)$ for most y) we see that

$$(-1)^{P_{s_1}(y)+P_{s_2}(y+s_1)-P_{r_1}(y)-P_{r_2}(y+r_1)-P_h(y+x)} = 1+o(1)$$

for most y. Taking expectations and applying Lemma 4.5 as before, we conclude that

$$P_{s_1}(y)+P_{s_2}(y+s_1)-P_{r_1}(y)-P_{r_2}(y+r_1)-P_h(y+x) = 0$$

for *all* y. Specialising to $y = 0$ and applying (19) we conclude that

$$P_h(x) = Q(x+h) - Q(x) = Q_h(x) - Q(x) \qquad (20)$$

for all $x \in \mathbf{F}_2^n$ and $h \in H$; thus we have succesfully "integrated" $P_h(x)$. We can then extend $P_h(x)$ to all $h \in \mathbf{F}_2^n$ (not just $h \in H$) by viewing (20) as a *definition*. Observe that if $h \in \mathbf{F}_2^n$, then $h = h_1 + h_2$ for some $h_1, h_2 \in H$, and from (20) we have

$$P_h(x) = P_{h_1}(x) + P_{h_2}(x + h_1).$$

In particular, since the right-hand side is a polynomial of degree at most $d - 2$, the left-hand side is also. Thus we see that $Q_h - Q$ is a polynomial of degree at most $d-2$ for all h, which easily implies that Q itself is a polynomial of degree at most $d-1$. If we then set $g(x) := f(x)(-1)^{Q(x)}$, then from (16), (20) we see that for every $h \in H$ we have

$$g(x + h) = g(x) + o(1)$$

for most x. From Fubini's theorem, we thus conclude that there exists an x such that $g(x+h) = g(x)+o(1)$ for most h, thus g is almost constant. Since $|g(x)| = 1 + o(1)$ for most x, we thus conclude the existence of a sign $\epsilon \in \{-1, +1\}$ such that $g(x) = \epsilon + o(1)$ for most x. We conclude that

$$f(x) = \epsilon(-1)^{Q(x)} + o(1)$$

for most x, and the claim then follows (assuming δ is small enough). □

Remark 4.7. The above argument requires $\|f\|_{U^d(\mathbf{F}_2^n)}$ to be very close to 1 for two reasons. Firstly, one wishes to exploit the rigidity property; and secondly, we implicitly used at many occasions the fact that if two properties each hold $1 - o(1)$ of the time, then they jointly hold $1 - o(1)$ of the time as well. These two facts break down once we leave the "99%-structured" world and instead work in a "1%-structured" world in which various statements are only true for a proportion at least ε for some small ε. Nevertheless, the proof of the Gowers inverse conjecture for $d = 2$ in [23] has some features in common with the above argument, giving one hope that the full conjecture could be settled by some extension of these methods.

Remark 4.8. The above result was essentially proven in [2] (extending an argument in [4] for the linear case $d = 2$), using a "majority vote" version of the dual function (13).

5. Concluding remarks

Despite the above results, we still do not have a systematic theory of structure and randomness which covers all possible applications (particularly for "sparse" objects). For instance, there seem to be analogous structure theorems for random variables, in which one uses Shannon entropy instead of L^2-based energies in order to measure complexity; see [25]. In analogy with the ergodic theory literature (e.g. [7]), there may also be some advantage in pursuing *relative* structure theorems, in which the notions of structure and randomness are all relative to some existing "known structure", such as a reference factor $\mathbf{Y}_0$ of a probability space $(X, \mathbf{X}, \mu)$. Finally, in the iterative algorithms used above to prove the structure theorems, the additional structures used at each stage of the iteration were drawn from a fixed stock of structures (S in the Hilbert space case, $\mathcal{S}$ in the measure space case). In some applications it may be more effective to adopt a more *adaptive* approach, in which the stock of structures one is using varies after each iteration. A simple example of this approach is in [32], in which the structures used at each stage of the iteration are adapted to a certain spatial scale which decreases rapidly with the iteration. I expect to see several more permutations and refinements of these sorts of structure theorems developed for future applications.

6. Acknowledgements

The author is supported by a grant from the MacArthur Foundation, and by NSF grant CCF-0649473. The author is also indebted to Ben Green for helpful comments and references.

References

[1] N. Alon, E. Fischer, M. Krivelevich, B. Szegedy, *Efficient testing of large graphs*, Proc. of 40^{th} FOCS, New York, NY, IEEE (1999), 656–666. Also: Combinatorica 20 (2000), 451–476.

[2] N. Alon, T. Kaufman, M. Krivelevich, S. Litsyn and D. Ron, *Testing low-degree polynomials over GF(2)*, RANDOM-APPROX 2003, 188–199. Also: Testing Reed-Muller codes, IEEE Transactions on Information Theory 51 (2005), 4032–4039.

[3] T. Austin, *On the structure of certain infinite random hypergraphs*, preprint.

[4] M. Blum, M. Luby, R. Rubinfeld, *Self-testing/correcting with applications to numerical problems*, J. Computer and System Sciences **47** (1993), 549–595.

[5] J. Bourgain, *A Szemerédi type theorem for sets of positive density in* $\mathbf{R}^k$, Israel J. Math. **54** (1986), no. 3, 307–316.

[6] A. Frieze, R. Kannan, *Quick approximation to matrices and applications*, Combinatorica **19** (1999), no. 2, 175–220.

[7] H. Furstenberg, Recurrence in Ergodic theory and Combinatorial Number Theory, Princeton University Press, Princeton NJ 1981.

[8] T. Gowers, *Lower bounds of tower type for Szemerédi's uniformity lemma*, Geom. Func. Anal., **7** (1997), 322–337.

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

[10] T. Gowers, *A new proof of Szemeredi's theorem*, Geom. Func. Anal., **11** (2001), 465-588.

[11] T. Gowers, *Quasirandomness, counting, and regularity for* 3*-uniform hypergraphs*, Comb. Probab. Comput. **15**, No. 1-2. (2006), pp. 143–184.

[12] T. Gowers, *Hypergraph regularity and the multidimensional Szemerédi theorem*, preprint.

[13] B. Green, *A Szemerédi-type regularity lemma in abelian groups*, Geom. Func. Anal. **15** (2005), no. 2, 340–376.

[14] B. Green, *Montréal lecture notes on quadratic Fourier analysis*, preprint.

[15] B. Green, S. Konyagin, *On the Littlewood problem modulo a prime*, preprint.

[16] B. Green, T. Tao, *The primes contain arbitrarily long arithmetic progressions*, Annals of Math., to appear.

[17] B. Green, T. Tao, *An inverse theorem for the Gowers* $U^3(G)$ *norm*, preprint.

[18] B. Green, T. Tao, *New bounds for Szemerédi's theorem, I: Progressions of length 4 in finite field geometries*, preprint.

[19] B. Green, T. Tao, *Linear equations in primes*, preprint.

[20] L. Lovász, B. Szegedy, *Szemerédi's regularity lemma for the analyst*, preprint.

[21] V. Rödl, M. Schacht, *Regular partitions of hypergraphs*, preprint.

[22] V. Rödl, J. Skokan, *Regularity lemma for* k*-uniform hypergraphs*, *Random Structures Algorithms* **25** (2004), no. 1, 1–42.

[23] A. Samorodnitsky, *Hypergraph linearity and quadraticity tests for boolean functions*, preprint.

[24] E. Szemerédi, *On sets of integers containing no* k *elements in arithmetic progression*, Acta Arith. **27** (1975), 299–345.

[25] T. Tao, *Szemerèdis regularity lemma revisited*, Contrib. Discrete Math. **1** (2006), 8–28.

[26] T. Tao, *The Gaussian primes contain arbitrarily shaped constellations*, J. dAnalyse Mathematique **99** (2006), 109–176.

[27] T. Tao, *The dichotomy between structure and randomness, arithmetic progressions, and the primes*, 2006 ICM proceedings, Vol. I., 581–608.

[28] T. Tao, *A quantitative ergodic theory proof of Szemerédi's theorem*, preprint.

[29] T. Tao, *A variant of the hypergraph removal lemma*, preprint.

[30] T. Tao, *The ergodic and combinatorial approaches to Szemerédi's theorem*, preprint.

[31] T. Tao, *A correspondence principle between (hyper)graph theory and probability theory, and the (hyper)graph removal lemma*, preprint.

[32] T. Tao, *Norm convergence of multiple ergodic averages for commuting transformations*, preprint.

[33] T. Tao and V. Vu, Additive Combinatorics, Cambridge Univ. Press, 2006.

[34] T. Tao, T. Ziegler, *The primes contain arbitrarily long polynomial progressions*, preprint.