Hypergraph Regularity and the multidimensional Szemerédi Theorem.

W. T. Gowers

Abstract. *We prove analogues for hypergraphs of Szemerédi's regularity lemma and the associated counting lemma for graphs. As an application, we give the first combinatorial proof of the multidimensional Szemerédi theorem of Furstenberg and Katznelson, and the first proof that provides an explicit bound. Similar results with the same consequences have been obtained independently by Nagle, Rödl and Schacht.*

§1. Introduction.

Szemerédi's theorem states that, for every real number $\delta > 0$ and every positive integer k, there exists a positive integer N such that every subset A of the set $\{1, 2, \ldots, N\}$ of size at least δN contains an arithmetic progression of length k. There are now three substantially different proofs of the theorem, Szemerédi's original combinatorial argument [Sz1], an ergodic-theory proof due to Furstenberg (see for example [FKO]) and a proof by the author using Fourier analysis [G1]. Interestingly, there has for some years been a highly promising programme for yet another proof of the theorem, pioneered by Vojta Rödl (see for example [R]), developing an argument of Ruzsa and Szemerédi [RS] that proves the result for progressions of length three. Let us briefly sketch their argument.

The first step is the famous regularity lemma of Szemerédi [Sz2]. If G is a graph and A and B are sets of vertices in V, then let $e(A, B)$ stand for the number of pairs $(x, y) \in A \times B$ such that xy is an edge of G. Then the *density* $d(A, B)$ of the pair (A, B) is $e(A, B)/|A||B|$. The pair is *ϵ-regular* if $|d(A', B') - d(A, B)| \leqslant \epsilon$ for all subsets $A' \subset A$ and $B' \subset B$ such that $|A'| \geqslant \epsilon|A|$ and $|B'| \geqslant \epsilon|B|$. The basic idea is that a pair is regular with density d if it resembles a random graph with edge-probability d. Very roughly, the regularity lemma asserts that every graph can be decomposed into a few pieces, almost all of which are random-like. The precise statement is as follows.

Theorem 1.1. *Let $\epsilon > 0$. Then there exists a positive integer K such that, given any graph G, the vertices can be partitioned into at most K sets V_i, with sizes differing by at most 1, such that all but at most ϵK^2 of the pairs (V_i, V_j) are ϵ-regular.*

A partition is called *ϵ-regular* if it satisfies the conclusion of Theorem 1.1. The regularity lemma is particularly useful in conjunction with the following further result, known as the counting lemma.

Theorem 1.2. *For every $\alpha > 0$ and every k there exists $\epsilon > 0$ with the following property. Let $V_1, \ldots, V_k$ be sets of vertices in a graph G, and suppose that for each pair (i, j) the pair (V_i, V_j) is ϵ-regular with density d_{ij}. Let H be a graph with vertex set $(x_1, \ldots, x_k)$ and let $v_i \in V_i$ be chosen uniformly at random, the choices being independent. Then the probability that v_iv_j is an edge of G if and only if x_ix_j is an edge of H differs from $\prod_{x_ix_j \in H} d_{ij} \prod_{x_ix_j \notin H}(1 - d_{ij})$ by at most α.*

Roughly, this result tells us that the k-partite graph induced by the sets $V_1, \ldots, V_k$ contains the right number of labelled copies of the graph H. Let us briefly see why this result is true when H is a triangle. Suppose that U, V, W are three sets of vertices and the pairs (U, V), (V, W) and (W, U) are ϵ-regular with densities ζ, η and θ respectively. Then a typical vertex of U has about $\zeta|V|$ neighbours in V and $\theta|W|$ neighbours in W. By the regularity of the pair (V, W), these two neighbourhoods span about $\eta(\zeta|V|)(\theta|W|)$ edges in G, creating that many triangles. Summing over all vertices of U we obtain the result.

The next step in the chain of reasoning is the following innocent-looking statement about graphs with few triangles. Some of the details of the proof will be sketched rather than given in full.

Theorem 1.3. *For every constant $a > 0$ there exists a constant $c > 0$ with the following property. If G is any graph with n vertices that contains at most cn^3 triangles, then it is possible to remove at most an^2 edges from G to make it triangle-free.*

Proof. This theorem is a simple consequence of the regularity lemma. Indeed, let $\epsilon = \epsilon(a) > 0$ be sufficiently small and let $V_1, \ldots, V_K$ be an ϵ-regular partition of the vertices of G. If there are fewer than $a|V_i||V_j|/100$ edges between V_i and V_j, then remove all those edges, and also remove all edges from V_i to V_j if (V_i, V_j) is not an ϵ-regular pair. Since the partition is ϵ-regular, we have removed fewer than an^2 edges, and the resulting graph must either be triangle-free or contain several triangles. To see why this is, suppose that (x, y, z) is a triangle in G (after the edges have been removed), and suppose that $(x, y, z) \in V_i \times V_j \times V_k$. Then by our construction the pair (V_i, V_j) must be regular and must span many edges (because we did not remove the edge (x, y)) and similarly for the pairs (V_j, V_k) and (V_i, V_k). But then, by the counting lemma for triangles, the sets V_i, V_j and V_k span at least $a^3|V_i||V_j||V_k|/10^6$ triangles. Each V_i has cardinality at least $n/2K$, where K depends on ϵ only (which itself depends on a only). This proves that the result is true provided that $c \leqslant a^3/10^6K^3$. □

Ruzsa and Szemerédi [RS] observed that Theorem 1.3 implies Szemerédi's theorem for progressions of length 3. More recently, Solymosi noticed [S1,2] that it also implied the following two-dimensional generalization. (Actually, neither of these statements is quite accurate. There are several closely related graph-theoretic results that have these consequences and can be proved using the regularity lemma, of which Theorem 1.3 is one. Ruzsa and Szemerédi and Solymosi did not use Theorem 1.3 itself but their arguments are not importantly different.)

Corollary 1.4. *For every $\delta > 0$ there exists N such that every subset $A \subset [N]^2$ of size at least δN^2 contains a triple of the form (x, y), $(x + d, y)$, $(x, y + d)$ with $d > 0$.*

Proof. First, note that an easy argument allows us to replace A by a set B that is symmetric about some point. Briefly, if the point (x, y) is chosen at random then the intersection of A with $(x, y) - A$ has expected size $c\delta^2 N^2$ for some absolute constant $c > 0$, lives inside the grid $[-N, N]^2$, and has the property that $B = (x, y) - B$. So B is still reasonably dense, and if it contains a subset K then it also contains a translate of $-K$. So we shall not worry about the condition $d > 0$. (I am grateful to Ben Green for bringing this trick to my attention. As it happens, the resulting improvement to the theorem is something of a side issue, since the positivity of d does not tend to be used in applications. See for instance Corollary 1.5 below. See also the remark at the beginning of the proof of Theorem 10.3.)

Without loss of generality, the original set A is symmetric in this sense. Let X be the set of all vertical lines through $[N]^2$, that is, subsets of the form $\{(x, y) : x = u\}$ for some $u \in [N]$. Similarly, let Y be the set of all horizontal lines. Define a third set, Z, of diagonal lines, that is, lines of constant $x + y$. These sets form the vertex sets of a tripartite graph, where a line in one set is joined to a line in another if and only if their intersection belongs to A. For example, the line $x = u$ is joined to the line $y = v$ if and only if $(u, v) \in A$ and the line $x = u$ is joined to the line $x + y = w$ if and only if $(u, w - u) \in A$.

Suppose that the resulting graph G contains a triangle of lines $x = u$, $y = v$, $x+y = w$. Then the points (u, v), $(u, w - u)$ and $(w - v, v)$ all lie in A. Setting $d = w - u - v$, we can rewrite them as (u, v), $(u, v + d)$, $(u + d, v)$, which shows that we are done unless $d = 0$. When $d = 0$, we have $u + v = w$, which corresponds to the degenerate case when the vertices of the triangle in G are three lines that intersect in a single point. Clearly, this can happen in at most $|A| = o(N^3)$ ways.

Therefore, if A contains no configuration of the desired kind, then the hypothesis of Theorem 1.3 holds, and we can remove $o(N^2)$ edges from G to make it triangle-free. But this is a contradiction, because there are at least δN^2 degenerate triangles and they are edge-disjoint. □

Corollary 1.5. *For every $\delta > 0$ there exists N such that every subset A of $\{1, 2, \ldots, N\}$ of size at least δN contains an arithmetic progression of length 3.*

Proof. Define $B \subset [N]^2$ to be the set of all (x, y) such that $x+2y \in A$. It is straightforward to show that B has density at least $\eta > 0$ for some η that depends on δ only. Applying Corollary 1.2 to B we obtain inside it three points (x, y), $(x + d, y)$ and $(x, y + d)$. Then the three numbers $x + 2y$, $x + d + 2y$ and $x + 2(y + d)$ belong to A and form an arithmetic progression. □

And now the programme for proving Szemerédi's theorem in general starts to become clear. Suppose, for example, that one would like to prove it for progressions of length 4. After a little thought, one sees that the direction in which one should generalize Theorem 1.3 is the one that takes graphs to 3-*uniform hypergraphs*, that is, set-systems consisting of subsets of size 3 of a set X. If H is a 3-uniform hypergraph, then a *simplex* in H is a set of 4 vertices x, y, z and w of H (that is, elements of the set X) such that the four triples xyz, xyw, xzw and yzw all belong to H. The following theorem of Frankl and Rödl is a direct generalization of Theorem 1.3, but its proof is much harder.

Theorem 1.6. *For every constant $a > 0$ there exists a constant $c > 0$ with the following property. If H is any 3-uniform hypergraph with n vertices that contains at most cn^4 simplices, then it is possible to remove at most an^3 edges from H to make it simplex-free.*

As observed by Solymosi, it is straightforward to generalize the proof of Theorem 1.4 and show that Theorem 1.6 has the following consequence.

Theorem 1.7. *For every $\delta > 0$ there exists N such that every subset $A \subset [N]^3$ of size at least δN^3 contains a quadruple of points of the form*

$$\{(x, y, z), (x + d, y, z), (x, y + d, z), (x, y, z + d)\}$$

with $d > 0$.

Similarly, Szemerédi's theorem for progressions of length four is an easy consequence of Theorem 1.7 (and once again one does not need the positivity of d).

It may look as though this section contains enough hints to enable any sufficiently diligent mathematician to complete a proof of the entire theorem. Indeed, here is a sketch for the 3-uniform case. First, one proves the appropriate 3-graph analogue of Szemerédi's regularity lemma. Then, given a hypergraph H, one applies this lemma. Next, one removes all sparse triples and all triples that fail to be regular. If the resulting hypergraph contains a simplex, then any three of the four sets in which its vertices lie must form a dense regular triple, and therefore (by regularity) the hypergraph contains many simplices, contradicting the original assumption.

The trouble with the above paragraph is that it leaves unspecified what it means for a triple to be regular. It turns out to be surprisingly hard to come up with an appropriate definition, where "appropriate" means that it must satisfy two conditions. First, it should be weak enough that for a regularity lemma to hold: that is, one should always be able to divide a hypergraph up into regular pieces. Second, it should be strong enough to yield the conclusion that four sets of vertices, any three of which form a dense regular triple, should span many simplices. The definition that Frankl and Rödl used for this purpose is complicated and it proved very hard to generalize. In [G2] we gave a different proof which is in some ways more natural. The purpose of this paper is to generalize the results of [G2] from 3-uniform hypergraphs to k-uniform hypergraphs for arbitrary k, thereby proving the full multidimensional version of Szemerédi's theorem (Theorem 10.3 below), which was first proved by Furstenberg and Katznelson [FK]. This is the first proof of the multidimensional Szemerédi theorem that is not based on Furstenberg's ergodic-theoretic approach, and also the first proof that gives an explicit bound. The bound, however, is very weak - giving an Ackermann-type dependence on the initial parameters.

Although this paper is self-contained, we recommend reading [G2] first. The case $k = 3$ contains nearly all the essential ideas, and they are easier to understand when definitions and proofs can be given directly. Here, because we are dealing with a general k, many of the definitions have to be presented inductively. The resulting proofs can be neater, but they may appear less motivated if one has not examined smaller special cases. The bulk of [G2] consists of background material and general discussion (such as, for example, a complete proof of the regularity lemma for graphs and a detailed explanation of how the ideas relate to those of the analytic approach to Szemerédi's theorem in [G1]). Rather than repeat all that motivating material, we refer the reader to that paper for it.

The main results of this paper have been obtained independently by Nagle, Rödl and Schacht. They too prove hypergraph generalizations of the regularity and counting lemmas that imply Theorem 10.3 and Szemerédi's theorem. However, they formulate their generalizations differently and there are substantial differences between their proof and ours. Broadly speaking, they take the proof of Frankl and Rödl as their starting point, whereas we start with the arguments of [G2]. At the end of [G2] we explain in more detail what the main difference is between the two approaches.

§2. Some basic definitions.

2.1. *Hypergraphs and chains.*

An *r-partite hypergraph* is a sequence $X_1, \ldots, X_r$ of disjoint sets, together with a collection $\mathcal{H}$ of subsets A of $X_1 \cup \ldots \cup X_r$ with the property that $|A \cap X_i| \leqslant 1$ for every i. The sets X_i are called *vertex sets* and their elements are *vertices*. The elements of $\mathcal{H}$ are called *edges*, or sometimes *hyperedges* if there is a danger of confusing them with edges in the graph-theoretic sense. A hypergraph is *k-uniform* if all its edges have size k. (Thus, a 2-uniform hypergraph is a graph.)

An r-partite hypergraph $\mathcal{H}$ is called an *r-partite chain* if it has the additional property that B is an edge of $\mathcal{H}$ whenever A is an edge of $\mathcal{H}$ and $B \subset A$. Thus, an r-partite chain is a particular kind of combinatorial simplicial complex, or down-set. Our use of the word "chain" is non-standard (in particular, it should not be confused with the notion of a chain complex in algebraic topology). We use it because it is quicker to write than "simplicial complex".

If the largest size of any edge of $\mathcal{H}$ is k, then we shall sometimes say that $\mathcal{H}$ is a *k-chain.*

2.2. *Homomorphisms and r-partite functions.*

Let $E_1, \ldots, E_r$ and $X_1, \ldots, X_r$ be two sequences of disjoint finite sets. If ϕ is a map from $E_1 \cup \ldots \cup E_r$ to $X_1 \cup \ldots \cup X_r$ such that $\phi(E_i) \subset X_i$ for every i, we shall say that ϕ is an *r-partite function.*

Let $\mathcal{J}$ be an r-partite chain with vertex sets $E_1, \ldots, E_r$ and let $\mathcal{H}$ be an r-partite chain with vertex sets $X_1, \ldots, X_r$. Let ϕ be an r-partite function from the vertices of $\mathcal{J}$ to the vertices of $\mathcal{H}$. We shall say that ϕ is a *homomorphism* from $\mathcal{J}$ to $\mathcal{H}$ if $\phi(A) \in \mathcal{H}$ whenever $A \in \mathcal{J}$. We shall write $\mathrm{Hom}(\mathcal{J}, \mathcal{H})$ for the set of all homomorphisms from $\mathcal{J}$ to $\mathcal{H}$.

2.3. *A-functions and $\mathcal{J}$-functions.*

Let Φ be the set of all r-partite maps from $E_1 \cup \ldots \cup E_r$ to $X_1 \cup \ldots \cup X_r$. We shall also consider some special classes of functions defined on Φ. If A is a subset of $E_1 \cup \ldots \cup E_r$ such that $|A \cap E_i| \leqslant 1$ for every i, then a function $f : \Phi \to [-1, 1]$ will be called an *A-function* if the value of $f(\phi)$ depends only on the image $\phi(A)$. If $\mathcal{J}$ is an r-partite chain with vertex sets $E_1, \dots, E_r$, then a *$\mathcal{J}$-function* is a function $f : \Phi \to [-1,1]$ that can be written as a product $f = \prod_{A \in \mathcal{J}} f^A$, where each f^A is an A-function.

The definition of A-functions and $\mathcal{J}$-functions is introduced in order to deal with situations where we have a function of several variables that can be written as a product of other functions each of which depends on only some of those variables. Let us clarify this with a small example. Suppose that we have three sets X_1, X_2 and X_3 and a function $f : X_1^2 \times X_2 \times X_3 \to [-1,1]$ of the form

$$f(x_1, x_1', x_2, x_3) = f_1(x_1, x_2) f_2(x_1, x_3) f_3(x_1', x_2) f_4(x_1', x_3) \ .$$

Let $E_1 = \{1, 1'\}$, $E_2 = \{2\}$ and $E_3 = \{3\}$. There is an obvious one-to-one correspondence between quadruples (x_1, x_1', x_2, x_3) and tripartite maps from $E_1 \cup E_2 \cup E_3$: given such a sequence one associates with it the map ϕ that takes 1 to x_1, $1'$ to x_1', 2 to x_2 and 3 to x_3. Therefore, we can if we wish change to a more opaque notation and write

$$f(\phi) = f_1(\phi) f_2(\phi) f_3(\phi) f_4(\phi) \ .$$

Now $f_2(\phi) = f_2(\phi(1), \phi(3)) = f_2\big(\phi(\{1,3\})\big)$, so f_2 is a $\{1,3\}$-function. Similar remarks can be made about f_1, f_3 and f_4. It follows that f is a $\mathcal{J}$-function if we take $\mathcal{J}$ to be the chain consisting of the sets $\{1,2\}$, $\{1,3\}$, $\{1',2\}$ and $\{1',3\}$ and all their subsets. The fact that the subsets are not mentioned in the formula does not matter, since if C is one of these subsets we can take the function that is identically 1 as our C-function.

An important and more general example is the following. As above, let $\mathcal{J}$ be an r-partite chain with vertex sets $E_1, \dots, E_r$ and let $\mathcal{H}$ be an r-partite chain with vertex sets $X_1, \dots, X_r$. For each ϕ in Φ and each $A \in \mathcal{J}$ let $H^A(\phi)$ equal 1 if $\phi(A) \in \mathcal{H}$ and 0 otherwise. Let $\mathcal{H}(\phi) = \prod_{A \in \mathcal{J}} H^A(\phi)$. Then $\mathcal{H}(\phi)$ equals 1 if $\phi \in \mathrm{Hom}(\mathcal{J}, \mathcal{H})$ and 0 otherwise. In other words, the characteristic function of $\mathrm{Hom}(\mathcal{J}, \mathcal{H})$ is a $\mathcal{J}$-function. We stress that $\mathcal{H}(\phi)$ depends on $\mathcal{J}$; however, it is convenient to suppress this dependence in the notation. Our counting lemma will count homomorphisms from small chains $\mathcal{J}$ to large quasirandom chains $\mathcal{H}$, so we can regard our main aim as being to estimate the sum of $H(\phi)$ over all $\phi \in \Phi$. However, in order to do so we need to consider more general $\mathcal{J}$-functions.

The $\mathcal{J}$-functions we consider will be supported in a chain $\mathcal{H}$ in the following sense. Let us say that an A-function f^A is *supported in* $\mathcal{H}$ if $f^A(\phi)$ is zero whenever $\phi(A)$ fails to be an edge of $\mathcal{H}$. Equivalently, f^A is supported in $\mathcal{H}$ if $f^A = f^A H^A$, where H^A is as defined above. Then we shall say that f is a *$\mathcal{J}$-function on* $\mathcal{H}$ if it can be written as a product $\prod_{A \in \mathcal{J}} f^A$, where each f^A is an A-function supported in $\mathcal{H}$. If f is a $\mathcal{J}$-function on $\mathcal{H}$, then $f(\phi) = 0$ whenever ϕ does not belong to $\mathrm{Hom}(\mathcal{J}, \mathcal{H})$. That is, $f(\phi) = f(\phi)\mathcal{H}(\phi)$. Notice that the product of any $\mathcal{J}$ function with the function $\mathcal{H}$ will be a $\mathcal{J}$-function on $\mathcal{H}$.

2.4. *The index of a set, and relative density in a chain.*

Let $\mathcal{H}$ be an r-partite chain with vertex sets $X_1, \ldots, X_r$. Given a set $E \in \mathcal{H}$, define its *index* $i(E)$ to be the set of all i such that $E \cap X_i$ is non-empty. (Recall that $E \cap X_i$ is a singleton for each such i.) For any set A in any r-partite chain, let $H(A)$ be the collection of all sets $E \in \mathcal{H}$ of index equal to that of A. If A has cardinality k, then let $H_*(A)$ be the collection of all sets D of index $i(A)$ such that $C \in \mathcal{H}$ whenever $C \subset D$ and C has cardinality $k-1$. (Since $\mathcal{H}$ is a chain, it follows from this that all proper subsets of D belong to $\mathcal{H}$.) Clearly $H(A) \subset H_*(A)$. The *relative density of* $H(A)$ *in* $\mathcal{H}$ is defined to be $|H(A)|/|H_*(A)|$. We will denote it by δ_A.

The relative density δ_A can be interpreted probabilistically: it is the conditional probability that a randomly chosen set $D \subset X_1 \cup \ldots \cup X_r$ of index $i(A)$ belongs to $\mathcal{H}$ (and hence to $H(A)$), given that all its proper subsets belong to $\mathcal{H}$.

Notational remark. It may help the reader to remember the definitions in this section if we explicitly point out that most of the time we are adopting the following conventions. The symbols $\mathcal{J}$ and $\mathcal{K}$ are used for chains of fixed size that are embedded into a chain $\mathcal{H}$ of size tending to infinity. From these we sometimes form other chains: for instance, $\mathcal{J}_1$ will be a chain of fixed size derived from a chain $\mathcal{J}$, and $\mathcal{H}_{x,x'}$ will be a chain of size tending to infinity derived from a chain $\mathcal{H}$. The letter H will tend to be reserved for set systems connected with $\mathcal{H}$ where the sets all have the same index. The same goes for functions derived from $\mathcal{H}$. For example, we write $\mathcal{H}(\phi)$ because we use the full chain $\mathcal{H}$ to define the function, whereas we write $H^A(\phi)$ because for that we just use sets of index $i(A)$, and hence of size $|A|$. Similarly, we write $H_*(A)$ because all sets in $H_*(A)$ have index $i(A)$.

2.5. $\mathrm{Oct}(f)$ *for an A-function f.*

We are building up to a definition of quasirandomness for $\mathcal{H}(A)$. An important ingredient of the definition is a weighted count of combinatorial octahedra, which generalizes the following idea. If f is a function of three variables x, y and z that range over sets X, Y and Z, respectively, then we define a quantity $\mathrm{Oct}(f)$ to be the sum over all $x, x' \in X$, $y, y' \in Y$ and $z, z' \in Z$ of

$$f(x,y,z)f(x,y,z')f(x,y',z)f(x,y',z')f(x',y,z)f(x',y,z')f(x',y',z)f(x',y',z') .$$

Similarly, if f is a function of k variables $x_1, \dots, x_k$, with each x_i taken from a set X_i, then

$$\mathrm{Oct}(f) = \sum_{x_1^0, x_1^1 \in X_1} \cdots \sum_{x_k^0, x_k^1 \in X_k} \prod_{\epsilon \in \{0,1\}^k} f(x_1^{\epsilon_1}, \dots, x_k^{\epsilon_k}) \ .$$

Now let us express this definition in the language of A-functions. As before, let $\mathcal{J}$ and $\mathcal{H}$ be r-partite chains with vertex sets $E_1, \dots, E_r$ and $X_1, \dots, X_r$, let Φ be the set of all r-partite maps from $E_1 \cup \ldots \cup E_r$ to $X_1 \cup \ldots \cup X_r$ and let $A \in \mathcal{J}$. We can think of an A-function as a function defined on the product of those X_i for which $i \in i(A)$. However, we can also think of it as a function f defined on Φ such that $f(\phi)$ depends only on $\phi(A)$. To define $\mathrm{Oct}(f)$ in these terms, we construct a set system $\mathcal{B}$ as follows. For each $i \in i(A)$ let U_i be a set of cardinality 2, let U be the union of the U_i and let $\mathcal{B}$ consist of the 2^k sets $B \subset U$ such that $|B \cap U_i| = 1$ for every i. Let Ω be the set of all maps ω from U to $\bigcup X_i$ such that $\omega(U_i) \subset X_i$ for every $i \in i(A)$. For each $\omega \in \Omega$ define $f^B(\omega)$ to be the value of $f(\phi)$ for any map $\phi \in \Phi$ such that $\phi(A) = \omega(B)$. (This is well-defined, since f is an A-function.) Then define

$$\mathrm{Oct}(f) = \sum_{\omega \in \Omega} \prod_{B \in \mathcal{B}} f^B(\omega) \ .$$

A moment's thought will show that this definition agrees with the earlier one (which, for simplicity, dealt with the case $A = \{1, 2, \dots, k\}$).

2.6. *Octahedral quasirandomness.*

We come now to the most important definition of this section. Let f be an A-function, and suppose that it is supported in $H_*(A)$, in the sense that $f(\phi) = 0$ whenever $\phi(A) \notin H_*(A)$. Equivalently, suppose that whenever $f(\phi) \neq 0$ we have $f(C) \in \mathcal{H}$ for every proper subset $C \subset A$. Loosely speaking, we shall say that f is *octahedrally quasirandom relative to* $\mathcal{H}$ if $\mathrm{Oct}(f)$ is significantly smaller than one might expect.

To turn this idea into a precise definition, we need to decide what we expect. Let $\mathcal{B}$ be the set system defined in the previous subsection. If $B \in \mathcal{B}$, then $f^B(\omega)$ is defined to be the value of $f(\phi)$ for any ϕ with $\phi(A) = \omega(B)$. If $f^B(\omega) \neq 0$, then $f(\phi) \neq 0$ so $\phi(A) \in H_*(A)$, by assumption, and hence $\omega(B) \in H_*(A)$. Therefore, a necessary condition for $\prod_{B \in \mathcal{B}} f^B(\omega)$ to be non-zero is that $\omega(D) \in \mathcal{H}$ for every D that is a proper subset of some $B \in \mathcal{B}$. Let $\mathcal{K}$ be the chain consisting of all such sets. Then, since $|f^B(\omega)| \leqslant 1$ for every B and every ω, a trivial upper bound for $\mathrm{Oct}(f)$ is

$$\sum_{\omega \in \Omega} \prod_{D \in \mathcal{K}} H^D(\omega) \ ,$$

which we shall call $\mathrm{Oct}(H_*(A))$, since it counts the number of (labelled, possibly degenerate) combinatorial k-dimensional octahedra in $H_*(A)$.

We could if we wanted declare $\mathrm{Oct}(f)$ to be small if it is small compared with $\mathrm{Oct}(H_*(A))$. Instead, however, since we shall be working exclusively with quasirandom chains, it is more convenient to work out how many octahedra we expect $H_*(A)$ to have, given the various relative densities, and use that quantity for comparison.

This is easily done if we recall the interpretation of relative densities as conditional probabilities. Suppose that we choose ω randomly from Ω, and also that $\mathcal{H}$ behaves in a random way. Then the probability that $H^D(\omega) = 1$ given that $H^C(\omega) = 1$ for every $C \subsetneq D$ is the probability that $\omega(D) \in \mathcal{H}$ given that $\omega(C) \in \mathcal{H}$ for every $C \subsetneq D$, which is δ_D. Because $\mathcal{H}$ behaves randomly, we expect all these conditional probabilities to be independent, so we expect that the $\sum_{\omega\in\Omega}\prod_{D\in\mathcal{K}} H^D(\omega)$ will be approximately $|\Omega|\prod_{D\in\mathcal{K}}\delta_D$. Accordingly, we shall say that f is *η-octahedrally quasirandom* if

$$\mathrm{Oct}(f) \leqslant \eta|\Omega| \prod_{D\in\mathcal{K}} \delta_D \ .$$

Since octahedral quasirandomness is the only form of quasirandomness that we use in this paper, we shall often omit the word "octahedrally" from this definition.

It is not necessary to do so, but one can rewrite the right-hand side more explicitly. For each proper subset $C \subsetneq A$, there are $2^{|C|}$ sets $D \in \mathcal{K}$ with the same index as C. (We can think of these as $|C|$-dimensional faces of the octahedron with index $i(C)$.) Moreove, writing N_i for the cardinality of X_i, we have that $|\Omega| = \prod_{i\in A} N_i^2$. Therefore,

$$\eta|\Omega| \prod_{D\in\mathcal{K}} \delta_D = \eta \prod_{C\subsetneq A} \delta_C^{2^{|C|}} \prod_{i\in A} N_i^2 \ .$$

The main use of the definition of quasirandomness for A-functions is to give us a precise way of saying what it means for a k-partite k-uniform hypergraph to "sit quasirandomly inside a k-partite $(k-1)$-chain". Let A and $\mathcal{H}$ be as above. The k-uniform hypergraph we would like to discuss is $H(A)$. Associated with this hypergraph is its "characteristic function" H^A and its relative density δ_A. The $(k-1)$-chain is the set of all edges of $\mathcal{H}$ with index some proper subset of A. Define an A-function f^A by setting $f^A(\phi)$ to equal $H^A(\phi) - \delta_A$ if $\phi(A) \in H_*(A)$ and zero otherwise. An important fact about f^A is that its average is zero. To see this, note that

$$|\Phi|^{-1} \sum_{\phi(A)\in H_*(A)} H^A(\phi) = |\Phi|^{-1} \sum_{\phi\in\Phi} H^A(\phi) \ ,$$

which is the probability that $\phi(A) \in \mathcal{H}$ for a randomly chosen r-partite map ϕ. Meanwhile,

$$|\Phi|^{-1} \sum_{\phi(A) \in H_*(A)} \delta_A$$

is the probability that $\phi(A) \in H_*(A)$ multiplied by the conditional probability that $\phi(A) \in H(A)$ given that $\phi(A) \in H_*(A)$, which is the same thing.

We shall say that $H(A)$ is *η-octahedrally quasirandom*, or just *η-quasirandom*, relative to $\mathcal{H}$, if the function f^A is η-quasirandom according to the definition given earlier. The counting lemma, which we shall prove in §5, will show that if $\mathcal{H}$ is an r-partite chain and all its different parts of the form $H(A)$ are quasirandom in this sense, then $\mathcal{H}$ behaves like a random chain with the same relative densities.

§3. A technical lemma for r-partite chains.

The main task of this section is to prove the following lemma. We have used the word "technical" to describe it, but it is the engine that drives all our later arguments. The statement is somewhat complicated because it has had to be artificially strengthened so that it is amenable to an inductive proof.

Lemma 3.1. *Let $\mathcal{J}_0$ be an r-partite chain with vertex sets $E_1, \ldots, E_r$ and let $A_0 \in \mathcal{J}_0$. Let $\mathcal{H}$ be an r-partite chain with vertex sets $X_1, \ldots, X_r$, and let Φ be the set of r-partite maps ϕ from $E_1 \cup \ldots \cup E_r$ to $X_1 \cup \ldots \cup X_r$. Let $G : \Phi \to [-1, 1]$ be a $\mathcal{J}_0$-function on $\mathcal{H}$, and suppose that G can be written $\prod_{A \in \mathcal{J}_0} g^A$, where each g^A is an A-function supported in $\mathcal{H}$ and $g^A = H^A$ whenever $A_0 \subset A$. Let f be an A_0-function supported in $\mathcal{H}$ and let $F = fG$. Let $\mathcal{J}$ be any subchain of $\mathcal{J}_0$ that contains A_0. Then there exists an r-partite k-chain $\mathcal{K}$ with vertex sets $F_1, \ldots, F_r$, satisfying the following properties.*

(i) $\mathcal{K}$ depends on $\mathcal{J}$ only.

(ii) There is a homomorphism γ from $\mathcal{K}$ to $\mathcal{J}$ such that for every set $A \in \mathcal{J}$ there are exactly 2^k sets $C \in \mathcal{K}$ with $\gamma(C) = A$.

(iii) For each $i \in i(A_0)$ there is a subset U_i of F_i of cardinality 2, and every set $C \subset \bigcup_{i \in A_0} U_i$ such that $|C \cap U_i| \leqslant 1$ for each $i \in A_0$ belongs to $\mathcal{K}$.

(iv) Let $\mathcal{B}$ be the set of $B \subset \bigcup_{i \in A_0} U_i$ such that $|B \cap U_i| = 1$ for every $i \in A_0$ and let Ω be the set of all r-partite maps from $F_1 \cup \ldots \cup F_r$ to $X_1 \cup \ldots \cup X_r$. Then we have the inequality

$$\Big(\sum_{\phi \in \Phi} F(\phi)\Big)^{2^k} \leqslant \sum_{\omega \in \Omega} \prod_{C \in \mathcal{K}} H^C(\omega) \prod_{B \in \mathcal{B}} f^B(\omega) \ ,$$

where $f^B(\omega)$ is defined to be the value of $f(\phi)$ for any map $\phi \in \Phi$ such that $\phi(A_0) = \omega(B)$.

Before we give the proof of Lemma 3.1, here are a few remarks about it. The most important of the properties of $\mathcal{K}$ is the inequality in (iv). The left-hand side of this inequality can be interpreted as follows. First, rewrite it as

$$\sum_{\phi_1,\dots,\phi_{2^k}\in\Phi} F(\phi_1)\dots F(\phi_{2^k})\ ,$$

and write $F(\phi_1,\dots,\phi_{2^k})$ for $F(\phi_1)\dots F(\phi_{2^k})$. Now a sequence of 2^k functions $\phi_1,\dots,\phi_{2^k}$ in Φ can be thought of as a single r-partite function with range $X_1 \cup \dots \cup X_r$ and domain a suitable set built out of 2^k copies of $E_1 \cup \dots \cup E_r$. Moreover, if in each of these copies we put a copy of $\mathcal{J}_0$ and call the resulting chain $\mathcal{L}$, then $F(\phi_1,\dots,\phi_{2^k})$ is the product of an $\mathcal{L}$-function with the function $f(\phi_1(A_0))\dots f(\phi_{2^k}(A_0))$. Thus, if $\mathcal{J} = \mathcal{J}_0$ then the main thing we have done in getting from the left-hand side to the right-hand side is to replace $\mathcal{L}$ by a chain $\mathcal{K}$ which also has 2^k sets for each set in $\mathcal{J}_0$, but which allows us to put together all the copies of A_0 so that they form the faces of a k-octahedron. If $\mathcal{J}$ is a proper subset of $\mathcal{J}_0$ then we have also passed to a subchain.

The reason for this rather curious passage to a subset $\mathcal{J}$ of the original chain $\mathcal{J}_0$ is that we shall apply the lemma in the case where $\mathcal{J}$ is the set of all edges of $\mathcal{J}_0$ of size at most $k-1$. However, the proof of the lemma is by induction on k and we cannot afford to restrict to sets of size at most $k-2$ in the inductive hypothesis.

A third remark is that Lemma 3.1 follows from a certain special case, which is easier (for notational reasons) to prove. We state this implication as a separate lemma.

Lemma 3.2. *Lemma 3.1 is true in general if it is true whenever the vertex sets of $\mathcal{J}$ are singletons.*

Proof. Let all the notation be as in the statement of Lemma 3.1. Our task is to exhibit a chain $\mathcal{K}$ with the required properties. Let $m = |E_1| + \dots + |E_r|$ and define new chains $\mathcal{J}_0^*$ and $\mathcal{J}^*$ to consist of the same vertices and edges as $\mathcal{J}_0$ and $\mathcal{J}$, respectively, but with the vertex set $E_1 \cup \dots \cup E_r$ partitioned into m singletons instead of into the r sets $E_1,\dots,E_r$. Next, build an m-partite chain $\mathcal{H}^*$ out of $\mathcal{H}$ as follows. The vertex sets are all sets of the form $\{a\} \times X_i$, where a is an element of E_i. Let π be the projection that takes a vertex (a,x) of $\mathcal{H}^*$ to the vertex x of $\mathcal{H}$. Then Y is an edge of $\mathcal{H}^*$ if πY is an edge of $\mathcal{H}$ and the restriction of π to Y is an injection. (This second condition says that if $x \neq y$ then (a,x) and (a,y) cannot both belong to the same edge Y of $\mathcal{H}^*$.)

To each r-partite map $\phi \in \Phi$ we can associate an m-partite map $\phi^* : E_1 \cup \ldots \cup E_r \to \bigcup_{i=1}^r E_i \times X_i$ by setting $\phi^*(a) = (a, \phi(a))$. Let Φ^* be the set of such m-partite maps. Obviously, the map $\phi \mapsto \phi^*$ is a one-to-one correspondence between Φ and Φ^*: indeed, the map $\phi^* \mapsto \pi \circ \phi^*$ is its inverse. More importantly, $\phi(A) \in \mathcal{H}$ if and only if $\phi^*(A) \in \mathcal{H}^*$. This follows from the fact that $\phi(A) = \pi\phi^*(A)$, since $\pi\phi^*(A) \in \mathcal{H}$ if and only if $\phi^*(A) \in \mathcal{H}^*$ by the definition of $\mathcal{H}^*$ (and the fact that π restricted to $\phi^*(A)$ is clearly an injection).

Another useful fact is that if $f : \Phi \to [-1, 1]$ then $f \circ \gamma : \Phi^* \to [-1, 1]$, and f is an A-function supported in $\mathcal{H}$ if and only if $f \circ \pi$ is an A-function supported in $\mathcal{H}^*$. Let us quickly check this. Suppose that $f(\phi)$ depends only on $\phi(A)$ and let $\phi^*(A) = \psi^*(A)$. Then $\phi(A) = \psi(A)$ (applying γ to both sides), so $f(\phi) = f(\psi)$ and hence $f \circ \pi(\phi^*) = f \circ \pi(\psi^*)$). This shows that $f \circ \pi$ depends only on $\phi^*(A)$. Conversely, if $f \circ \pi$ depends only on $\phi^*(A)$ and $\phi(A) = \psi(A)$ then $\phi^*(A) = \psi^*(A)$ (as can easily be seen from the definition of ϕ^* and ψ^*), from which it follows that $f \circ \pi(\phi^*) = f \circ \pi(\psi^*)$ and hence that $f(\phi) = f(\psi)$. Now suppose that f is an A-function supported in $\mathcal{H}$ and that $f \circ \pi(\phi^*) \neq 0$. Then $\pi\phi^*(A) \in \mathcal{H}$, which implies that $\phi^*(A) \in \mathcal{H}^*$. Hence, $f \circ \pi$ is supported in $\mathcal{H}^*$. Conversely, if $f \circ \pi$ is an A-function supported in $\mathcal{H}^*$ and $f(\phi) \neq 0$ then $f \circ \pi(\phi^*) \neq 0$, so $\phi^*(A) \in \mathcal{H}^*$, and hence $\phi(A) = \pi\phi^*(A) \in \mathcal{H}$. This implies that $\phi(A) \in \mathcal{H}$, and hence that ϕ is supported in $\mathcal{H}$.

Having established these simple facts, let us apply Lemma 3.1 to the chains $\mathcal{J}_0^*$, $\mathcal{J}^*$ and $\mathcal{H}^*$ and to the function $G^* = \prod_{A \in \mathcal{J}^*} g^A \circ \pi$. The remarks above show that the hypotheses of the lemma are satisfied, so it provides us with an m-partite chain $\mathcal{K}^*$ with vertex sets $V_1, \ldots, V_m$, all of which have size 2^k. Let γ be the 2^k-to-one map given to us by property (ii). We can create from $\mathcal{K}^*$ an r-partite chain $\mathcal{K}$ by simply combining these vertex sets into larger ones: for each $i \leqslant r$ we let F_i be the union of all V_j such that $\gamma(V_j) \subset E_i$. (To put this less formally: each V_j consists of 2^k copies of a vertex of $\mathcal{J}$, and we combine the V_j according to the sets E_i in which the corresponding vertices of $\mathcal{J}$ lie.)

It is clear that $\mathcal{K}$ depends only on $\mathcal{J}$. To establish property (ii), all we have to do is check that γ takes elements of F_i to elements of E_i, which it does since that is how we defined the F_i. The rest of (ii) is obvious since the only difference between $\mathcal{K}$ and $\mathcal{K}^*$ is in the partition we have put on the vertex set, and similarly for $\mathcal{J}$ and $\mathcal{J}^*$.

Next, let $i \in i(A_0)$. Let a be the unique element of $A_0 \cap E_i$. Then $\{a\}$ is a vertex set of $\mathcal{J}^*$, and (iii) gives us a subset U_a of V_a of cardinality 2, which we can rename U_i, since it is also a subset of F_i of cardinality 2. Property (iii) now follows easily for $\mathcal{K}$ - again since we are talking about precisely the same sets as we were for $\mathcal{K}^*$.

Finally, let $\mathcal{B}$ be defined as in property (iv), and note that it is the same collection of sets, whether we are talking about $\mathcal{K}$ or $\mathcal{K}^*$. Lemma 3.1 for $\mathcal{J}_0^*$, $\mathcal{J}^*$ and $\mathcal{H}^*$ gives us the inequality

$$\Big(\sum_{\phi^*\in\Phi^*} F(\phi^*)\Big)^{2^k} \leqslant \sum_{\omega^*\in\Omega^*}\prod_{C\in\mathcal{K}^*}(H^*)^C(\omega^*)\prod_{B\in\mathcal{B}} f^B(\omega^*)\ ,$$

where Φ^* and Ω^* have obvious meanings and $F(\phi^*)$ and $f^B(\omega^*)$ are defined to be $F(\phi)$ and $f^B(\omega)$ for the corresponding functions $\phi \in \Phi$ and $\omega \in \Omega$. We also have $(H^*)^C(\omega^*) = H(\gamma(\omega^*(C))) = H(\omega(C)) = H^C(\omega)$. Therefore, this inequality is precisely what we want, and Lemma 3.1 is established for $\mathcal{J}_0$, $\mathcal{J}$ and $\mathcal{H}$. □

The proof of Lemma 3.1 is a good example of an argument that is simple to carry out in small cases (some examples are given in [G2]) and routine to generalize, but not easy to present in its general form without to some extent hiding its simplicity. Essentially, it consists of a large number of applications of the Cauchy-Schwarz inequality together with a few rearrangments of sums. However, to prove it in general, one must use induction, and as we have seen the inductive hypothesis turns out to be quite complicated.

Quite a lot of the main argument consists in little more than translating statements from one notation to another. To make the proof more comprehesible, we shall start with a preparatory lemma that will allow us to do this translation quickly and efficiently.

Suppose that $\mathcal{J}_0$ and $\mathcal{H}$ are r-partite chains with vertex sets $E_1,\dots,E_r$ and $X_1,\dots,X_r$, respectively, and suppose also that E_1 has only one element, e. Let $\mathcal{J}_1$ be the $(r-1)$-partite chain consisting of all sets $A \subset E_2\cup\ldots\cup E_r$ that belong to $\mathcal{J}_0$, and suppose that $\mathcal{J}_0$ satisfies the following condition.

(*) If A and B are any two sets in $\mathcal{J}_1$ with the same index then $\{e\}\cup A \in \mathcal{J}_0$ if and only if $\{e\}\cup B\in\mathcal{J}_0$.

Given any element $x \in X_1$ and any $(r-1)$-partite map ϕ' from $E_2\cup\ldots\cup E_r$ to $X_2\cup\ldots\cup X_r$, let (x,ϕ') denote the r-partite map from $E_1\cup\ldots\cup E_r$ to $X_1\cup\ldots\cup X_r$ that takes e to x and a to $\phi'(a)$ if $a\neq e$. Given two elements $x,x'\in X_1$, we shall be interested to know under what circumstances both (x,ϕ') and (x',ϕ') are homomorphisms. To be precise, we shall construct an $(r-1)$-partite chain $\mathcal{H}_{x,x'}$ with vertex sets $X_2,\dots,X_r$ such that (x,ϕ') and (x',ϕ') are both homomorphisms from $\mathcal{J}_0$ to $\mathcal{H}$ if and only if ϕ' is a homomorphism from $\mathcal{J}_1$ to $\mathcal{H}_{x,x'}$.

Let Φ be the set of all r-partite maps from $E_1\cup\ldots\cup E_r$ to $X_1\cup\ldots\cup X_r$ and let Φ' be the set of all $(r-1)$-partite maps from $E_2\cup\ldots\cup E_r$ to $X_2\cup\ldots\cup X_r$. Given a function $f:\Phi\to[-1,1]$ and two elements $x,x'\in X_1$, define a function $f_{x,x'}:\Phi'\to[-1,1]$ by

$f_{x,x'}(\phi') = f(x,\phi')f(x',\phi')$. Another useful property of the chain $\mathcal{H}_{x,x'}$ shows that, under suitable conditions, $f_{x,x'}$ relates to $\mathcal{H}_{x,x'}$ much as f relates to $\mathcal{H}$.

For convenience, we shall write $e \cup A$ instead of $\{e\} \cup A$, and similarly with other unions involving singletons. If $A \in \mathcal{J}_1$ then $(H_{x,x'})^A$ will have the obvious meaning: if $\phi' \in \Phi'$ then $(H_{x,x'})^A(\phi') = 1$ if $\phi'(A) \in \mathcal{H}_{x,x'}$ and 0 otherwise.

Lemma 3.3. *Let $\mathcal{J}_0$ and $\mathcal{H}$ be r-partite chains with vertex sets $E_1 \cup \ldots \cup E_r$ and $X_1 \cup \ldots \cup X_r$. Suppose that E_1 is a singleton $\{e\}$ and that $\mathcal{J}_0$ satisfies condition $(*)$ above. Then for each pair x, x' of elements of X_1 there exists an $(r-1)$-partite chain $\mathcal{H}_{x,x'}$ with vertex sets $X_2, \ldots, X_r$, with the following properties.*

- *(i) A function $\phi' \in \Phi'$ is a homomorphism from $\mathcal{J}_1$ to $\mathcal{H}_{x,x'}$ if and only if both (x,ϕ') and (x',ϕ') are homomorphisms from $\mathcal{J}_0$ to $\mathcal{H}$.*
- *(ii) If $A \in \mathcal{J}_1$ and $e \cup A \in \mathcal{J}$, then $(H^{e\cup A})_{x,x'} = (H_{x,x'})^A$.*
- *(iii) If $A \in \mathcal{J}_1$ and f is an $(e \cup A)$-function supported in $\mathcal{H}$, then $f_{x,x'}$ is an A-function supported in $\mathcal{H}_{x,x'}$.*

Proof. Define $\mathcal{H}_{x,x'}$ to be the $(r-1)$-partite chain with vertex sets $X_2, \ldots, X_r$ consisting of all sets Y that satisfy the following conditions. First, $Y \in \mathcal{H}$. Secondly, for every $Z \subset Y$, if $e \cup A \in \mathcal{J}_0$ for some $A \in \mathcal{J}_1$ with $i(A) = i(Z)$, then $x \cup Z \in \mathcal{H}$ and $x' \cup Z \in \mathcal{H}$. (We have to talk about subsets of Y in order to ensure that $\mathcal{H}_{x,x'}$ is a chain and not just a set system.) Let us prove the three properties.

(i) Suppose that $\phi' \in \Phi'$ and that (x,ϕ') and (x',ϕ') both belong to $\mathrm{Hom}(\mathcal{J}_0,\mathcal{H})$. Let $A \in \mathcal{J}_1$. Then $A \in \mathcal{J}_0$ and $(x,\phi')(A) = \phi'(A)$, so $\phi'(A) \in \mathcal{H}$. Let $Z \subset \phi'(A)$ and suppose that the index of Z is C. If $e \cup C \in \mathcal{J}_0$, then $(x,\phi')(e \cup C)$ and $(x',\phi')(e \cup C)$ belong to $\mathcal{H}$. That is, $x \cup \phi'(C)$ and $x' \cup \phi'(C)$ belong to $\mathcal{H}$. But $\phi'(C) = Z$, so $x \cup Z$ and $x' \cup Z$ belong to $\mathcal{H}$. We have proved the conditions necessary for $\phi'(A)$ to belong to $\mathcal{H}_{x,x'}$. This shows that $\phi' \in \mathrm{Hom}(\mathcal{J}_1,\mathcal{H}_{x,x'})$.

Conversely, suppose that $\phi' \in \mathrm{Hom}(\mathcal{J}_1,\mathcal{H}_{x,x'})$ and let $A \in \mathcal{J}$. If $e \notin A$, then $(x,\phi')(A) = (x',\phi')(A) = \phi'(A)$, which belongs to $\mathcal{H}_{x,x'}$ by our hypothesis on ϕ', and hence to $\mathcal{H}$, since $\mathcal{H}_{x,x'} \subset \mathcal{H}$. If $A = e \cup B$ then $(x,\phi')(A) = x \cup \phi'(B)$. But $B \in \mathcal{J}_1$ so $\phi'(B) \in \mathcal{H}_{x,x'}$, which implies, since $e \cup B \in \mathcal{J}_0$, that $x \cup \phi'(B) \in \mathcal{H}$. Similarly, $x' \cup \phi'(B) \in \mathcal{H}$. This shows that (x,ϕ') and (x',ϕ') both belong to $\mathrm{Hom}(\mathcal{J}_0,\mathcal{H})$.

(ii) Because $e \cup A \in \mathcal{J}_0$, we have that $\phi'(A) \in \mathcal{H}_{x,x'}$ if and only if $x \cup \phi'(A)$ and $x' \cup \phi'(A)$ belong to $\mathcal{H}$. But $x \cup \phi'(A) \in \mathcal{H}$ if and only if $H^{e\cup A}(x,\phi') = 1$ and $x' \cup \phi'(A) \in \mathcal{H}$ if and only if $H^{e\cup A}(x',\phi') = 1$. It follows that

$$(H_{x,x'})^A(\phi') = H^{e\cup A}(x,\phi')H^{e\cup A}(x',\phi') = (H^{e\cup A})_{x,x'}(\phi') \ .$$

(iii) If f is an $(e \cup A)$-function supported in $\mathcal{H}$, then $f = fH^{e\cup A}$, as we remarked in subsection 2.3. But then $f_{x,x'} = f_{x,x'}(H^{e\cup A})_{x,x'}$, which equals $f_{x,x'}(H_{x,x'})^A$, by part (ii) of this lemma. This proves (by the same remark but the reverse implication) that $f_{x,x'}$ is an A-function supported in $\mathcal{H}_{x,x'}$. $\square$

Notation. We shall use the customary notation $[k]$ for the set $\{1, 2, \ldots, k\}$ and $[2, k]$ for the set $\{2, 3, \ldots, k\}$.

Proof of Lemma 3.1. We prove this result by induction on k. By Lemma 3.2 it is enough to prove it in the case where the vertex sets E_i of $\mathcal{J}_0$ are singletons. However, we may, and will, assume the general case as our inductive hypothesis. It will also simplify our notation slightly and involve no loss of generality if we assume that each E_i is the singleton $\{i\}$ and that $A_0 = [k]$. In order to make the structure of the proof a bit clearer, we shall divide it into steps.

Step 1. *Splitting up the sum and applying the Cauchy-Schwarz inequality to obtain an expression to which the inductive hypothesis can be applied.*

As earlier, let $\mathcal{J}_1$ be the set of all $A \in \mathcal{J}_0$ such that $1 \notin A$. We shall also let $\mathcal{J}_2$ be the subset of $\mathcal{J}_1$ consisting of those $A \in \mathcal{J}_1$ for which $1 \cup A \in \mathcal{J}_0$. Then $\mathcal{J}_0 = \mathcal{J}_1 \cup \{1 \cup A : A \in \mathcal{J}_2\}$. Recall that Φ' is the set of $(r-1)$-partite maps from $[2, r]$ to $X_2 \cup \ldots \cup X_r$ and note that there is a one-to-one correspondence between Φ and $X_1 \times \Phi'$, as is implicit in the notation (x, ϕ') introduced before Lemma 3.3, which we shall use again here. Therefore,

$$\begin{aligned}\Big(\sum_{\phi\in\Phi} F(\phi)\Big)^{2^k} &= \Big(\sum_{\phi\in\Phi} f(\phi) \prod_{A\in\mathcal{J}_0} g^A(\phi)\Big)^{2^k}\\ &= \Big(\sum_{\phi'\in\Phi'} \sum_{x\in X_1} f(x,\phi') \prod_{A\in\mathcal{J}_1} g^A(\phi') \prod_{A\in\mathcal{J}_2} g^{1\cup A}(x,\phi')\Big)^{2^k}\ .\end{aligned}$$

Strictly speaking we should write $g^A(x, \phi')$ in the first product, but since this quantity does not depend on x when $1 \notin A$, we use the abbreviation $g^A(\phi')$.

Now, the inner sum is zero unless $\phi' \in \mathrm{Hom}(\mathcal{J}_1, \mathcal{H})$, since if $A \in \mathcal{J}_1$ and $\phi'(A) \notin \mathcal{H}$ then $g^A(\phi') = 0$. Therefore, setting $M = |\mathrm{Hom}(\mathcal{J}_1, \mathcal{H})|$, the Cauchy-Schwarz inequality implies that the last line is at most

$$\Big(M \sum_{\phi'\in\Phi'} \Big(\sum_{x\in X_1} f(x,\phi') \prod_{A\in\mathcal{J}_1} g^A(\phi') \prod_{A\in\mathcal{J}_2} g^{1\cup A}(x,\phi')\Big)^2\Big)^{2^{k-1}}\ .$$

Since the functions g^A with $A \in \mathcal{J}_1$ have no dependence on x, we will increase this last expression if we remove any such g^A from the product, or if we replace it by the larger function H^A. Therefore, we can set $\mathcal{J}_3 = \mathcal{J}_1 \cap \mathcal{J}$ and say that this last expression is at most

$$\Big(M \sum_{\phi'} \Big(\sum_{x} f(x,\phi') \prod_{A\in\mathcal{J}_3} H^A(\phi') \prod_{A\in\mathcal{J}_2} g^{1\cup A}(x,\phi')\Big)^2\Big)^{2^{k-1}}$$

$$= M^{2^{k-1}} \Big(\sum_{x,x'} \sum_{\phi'} f_{x,x'}(\phi') \prod_{A\in\mathcal{J}_3} H^A(\phi') \prod_{A\in\mathcal{J}_2} (g^{1\cup A})_{x,x'}(\phi')\Big)^{2^{k-1}} .$$

By Hölder's inequality, this is at most

$$M^{2^{k-1}} |X_1|^{2^k-2} \sum_{x,x'} \Big(\sum_{\phi'} f_{x,x'}(\phi') \prod_{A\in\mathcal{J}_3} H^A(\phi') \prod_{A\in\mathcal{J}_2} (g^{1\cup A})_{x,x'}(\phi')\Big)^{2^{k-1}} .$$

Step 2. *Rewriting the bracketed part of the last expression in terms of the chain $\mathcal{H}_{x,x'}$ in order to show that it satisfies the conditions for the inductive hypothesis to be applicable.*

The function f is a $[k]$-function supported in $\mathcal{H}$ and $[2,k] \in \mathcal{J}_1$. Hence, by Lemma 3.3, $f_{x,x'}$ is a $[2,k]$-function supported in $\mathcal{H}_{x,x'}$. Lemma 3.3 also tells us, for each $A \in \mathcal{J}_1$, that $g^{1\cup A}$ is an A-function supported in $\mathcal{H}_{x,x'}$. Here we are using the fact that $\mathcal{J}_0$ and $\mathcal{J}$ satisfy the condition $(*)$, for the trivial reason that there is at most one set of any given index. (It is for this purpose that we bothered with Lemma 3.2.)

The functions H^A must be treated slightly differently. If $A \in \mathcal{J}_3$, then H^A is certainly an A-function supported in $\mathcal{H}$, but it need not be supported in $\mathcal{H}_{x,x'}$. However, if $A \in \mathcal{J}_3$ and $H^A(\phi') \neq 0$, then $\phi'(A) \in \mathcal{H} \setminus \mathcal{H}_{x,x'}$, so there exists $C \subset A$ such that $1 \cup C \in \mathcal{J}_0$ and either $x \cup \phi'(C) \notin \mathcal{H}$ or $x' \cup \phi'(C) \notin \mathcal{H}$, which implies that $C \in \mathcal{J}_2$ and $(g^{1\cup C})_{x,x'}(\phi') = 0$. Therefore, we make no difference to the product if we replace each $H^A(\phi')$ in the first product by $(H_{x,x'})^A(\phi')$.

Note also that if $A \in \mathcal{J}_2$ and $[2,k] \subset A$, then $[k] \subset 1 \cup A$, so by hypothesis $g^{1\cup A} = H^{1\cup A}$. But then, by Lemma 3.3 again, $(g^{1\cup A})_{x,x'} = (H^{1\cup A})_{x,x'} = (H_{x,x'})^A$.

These observations show that we can rewrite the last line above as

$$M^{2^{k-1}} |X_1|^{2^k-2} \sum_{x,x'} \Big(\sum_{\phi'} f_{x,x'}(\phi') \prod_{A\in\mathcal{J}_2\cup\mathcal{J}_3} u^A_{x,x'}(\phi')\Big)^{2^{k-1}} ,$$

where each function $u^A_{x,x'}$ is an A-function (that depends on x and x') supported in $\mathcal{H}_{x,x'}$, and it equals $(H_{x,x'})^A$ when $[2,k] \subset A$. Moreover, $f_{x,x'}$ is a $[2,k]$-function supported in $\mathcal{H}_{x,x'}$ and $\prod_{A\in\mathcal{J}_2\cup\mathcal{J}_3} u^A_{x,x'}$ is a $(\mathcal{J}_2 \cup \mathcal{J}_3)$-function on $\mathcal{H}_{x,x'}$.

Step 3. *Applying the inductive hypothesis.*

We have just shown that the conditions for Lemma 3.1 are satisfied, with r, $\mathcal{J}_0$, A_0, $\mathcal{H}$ and g^A replaced by $r-1$, $\mathcal{J}_2 \cup \mathcal{J}_3$, $[2,k]$, $\mathcal{H}_{x,x'}$ and $u^A_{x,x'}$, respectively. Now let $\mathcal{J}_4 = \mathcal{J}_2 \cap \mathcal{J}$. Then $\mathcal{J}_4 \cup \mathcal{J}_3$ is a subchain of $\mathcal{J}_2 \cup \mathcal{J}_3$ and $[2,k] \in \mathcal{J}_4 \cup \mathcal{J}_3$, so we can apply the inductive hypothesis with $\mathcal{J}_4 \cup \mathcal{J}_3$ playing the part of $\mathcal{J}$. It provides us with an $(r-1)$-partite chain $\mathcal{K}_1$ and a homomorphism γ' from $\mathcal{K}_1$ to $\mathcal{J}_4 \cup \mathcal{J}_3$ such that for each $A \in \mathcal{J}_4 \cup \mathcal{J}_3$ there are exactly 2^{k-1} sets $C \in \mathcal{K}_1$ with $\gamma'(C) = A$. Let $V_2, \ldots, V_r$ be the vertex sets of $\mathcal{K}_1$. Then we are also provided with sets $U_2, \ldots, U_k$ such that $U_i \subset V_i$ and $|U_i| = 2$ for each i, and such that every subset of $U_2 \cup \ldots \cup U_k$ that intersects each U_i at most once belongs to $\mathcal{K}_1$. Letting $\mathcal{B}_1$ be the set of all such sets that intersect each U_i exactly once and letting Ω' be the set of all $(r-1)$-partite maps from $V_2 \cup \ldots \cup V_r$ to $X_2 \cup \ldots \cup X_r$, we are also given the inequality

$$\Big(\sum_{\phi'} f_{x,x'}(\phi') \prod_{A \in \mathcal{J}_2 \cup \mathcal{J}_3} u^A_{x,x'}(\phi')\Big)^{2^{k-1}} \leqslant \sum_{\omega' \in \Omega'} \prod_{C' \in \mathcal{K}'} (H_{x,x'})^{C'}(\omega') \prod_{B' \in \mathcal{B}_1} (f_{x,x'})^{B'}(\omega') \,.$$

And now, putting together all the inequalities we have shown so far, we have that

$$\Big(\sum_{\phi \in \Phi} F(\phi)\Big)^{2^k} \leqslant M^{2^{k-1}} |X_1|^{2^k - 2} \sum_{x,x'} \sum_{\omega' \in \Omega'} \prod_{C' \in \mathcal{K}_1} (H_{x,x'})^{C'}(\omega') \prod_{B' \in \mathcal{B}_1} (f_{x,x'})^{B'}(\omega') \,.$$

Step 4. *Rewriting the right-hand side of the above inequality in terms of $\mathcal{H}$ rather than $\mathcal{H}_{x,x'}$.*

The product $\prod_{C' \in \mathcal{K}_1} (H_{x,x'})^{C'}(\omega')$ is the characteristic function of the event that ω' is a homomorphism from $\mathcal{K}_1$ to $\mathcal{H}_{x,x'}$. Let $\mathcal{K}_0$ be the r-partite chain with vertex sets $\{1\}, V_2, V_3, \ldots, V_r$ consisting of all sets $A \in \mathcal{K}_1$ together with all sets of the form $1 \cup A$ such that $A \in \mathcal{K}_1$ and $\gamma'(A) \in \mathcal{J}_4$. (Recall that this is the case if $1 \cup \gamma'(A) \in \mathcal{J}$.) Then $\mathcal{K}_1$ is the collection of all sets $A \in \mathcal{K}_0$ that are subsets of $V_2 \cup \ldots \cup V_r$, so it is built out of $\mathcal{K}_0$ just as $\mathcal{J}_1$ is built out of $\mathcal{J}_0$. Furthermore, $\mathcal{K}_0$ satisfies condition $(*)$: if $A, B \in \mathcal{K}_1$, $i(A) = i(B)$ and $1 \cup A \in \mathcal{K}_0$, then $\gamma'(A) = \gamma'(B)$ and $\gamma'(A) \in \mathcal{J}_4$, so $\gamma'(B) \in \mathcal{J}_4$, so $1 \cup B \in \mathcal{K}_0$. Therefore, by Lemma 3.3, $\omega' \in \mathrm{Hom}(\mathcal{K}_1, \mathcal{H}_{x,x'})$ if and only if both (x, ω') and (x', ω') belong to $\mathrm{Hom}(\mathcal{K}_0, \mathcal{H})$.

Now we define an r-partite chain $\mathcal{K}_2$ with vertex sets $\{1, 1'\}, V_2, \ldots, V_r$. The edges of $\mathcal{K}_2$ are all sets $A \in \mathcal{K}_1$ together with all sets of the form $1 \cup A$ or $1' \cup A$ such that $\gamma'(A) \in \mathcal{J}_4$. For each $\omega' \in \Omega'$ let (x, x', ω') denote the r-partite map that takes 1 to x, $1'$ to x' and a

to $\omega'(a)$ for all $a \in V_2 \cup \ldots \cup V_r$. Clearly (x, ω') and (x', ω) belong to $\mathrm{Hom}(\mathcal{K}_0, \mathcal{H})$ if and only if $(x, x', \omega') \in \mathrm{Hom}(\mathcal{K}_2, \mathcal{H})$. We have therefore shown that

$$\prod_{C' \in \mathcal{K}_1} (H_{x,x'})^{C'}(\omega') = \prod_{C \in \mathcal{K}_2} H^C(x, x', \omega')$$

for every $x, x' \in X_1$ and every $\omega' \in \Omega'$.

As for $(f_{x,x'})^{B'}(\omega')$, it equals the value of $f_{x,x'}(\phi')$ for any $\phi' \in \Phi'$ such that $\phi'([2,k]) = \omega'(B')$. But if ϕ' has this property then $(x, \phi')([k]) = (x, x', \omega')(1 \cup B')$ and $(x', \phi')([k]) = (x, x', \omega')(1' \cup B')$. It follows that

$$(f_{x,x'})^{B'}(\omega') = f_{x,x'}(\phi') = f(x, \phi') f(x', \phi') = f^{1 \cup B'}(x, x', \omega') f^{1' \cup B'}(x, x', \omega') \ .$$

Let $U_1 = \{1, 1'\}$ and let $\mathcal{B}$ be the set of all sets of the form $1 \cup B'$ or $1' \cup B'$ with $B' \in \mathcal{B}'$. Let Ω_2 be the set of all r-partite maps from $U_1 \cup V_2 \cup \ldots \cup V_r$ to $X_1 \cup \ldots \cup X_r$ and note that these are precisely the maps of the form (x, x', ω') with $\omega' \in \Omega'$. It follows that the right-hand side of the last inequality can be written

$$M^{2^{k-1}} |X_1|^{2^k - 2} \sum_{\omega \in \Omega_2} \prod_{C \in \mathcal{K}_2} H^C(\omega) \prod_{B \in \mathcal{B}} f^B(\omega) \ .$$

Step 5. *Reinterpreting the inequality we have just proved, in order to demonstrate that it proves the lemma.*

It remains to define a suitable chain $\mathcal{K}$ and rewrite the expression we have just obtained in terms of $\mathcal{K}$ in such a way that it has the form stated in the lemma. Then we must check that $\mathcal{K}$ has the properties claimed. What enables us to do this is the fact that $M^{2^{k-1}}$ and $|X_1|^{2^k-2}$ both count homomorphisms from suitable chains into $\mathcal{H}$. (Recall that $M = |\mathrm{Hom}(\mathcal{J}_1, \mathcal{H})|$.)

Define a chain $\mathcal{K}$ as follows. Let W_1 be a set of size $2^k - 2$ and for $2 \leqslant i \leqslant r$ let W_i be a set of size 2^{k-1}. Let the vertex sets of $\mathcal{K}$ be $F_1, \ldots, F_r$, where $F_1 = U_1 \cup W_1$ and $F_i = V_i \cup W_i$ for $i \geqslant 2$. Let $\mathcal{K}_3$ be the set of all singletons $\{a\}$ with $a \in W_1$. Let $\mathcal{K}_4$ be a union of 2^{k-1} disjoint copies of $\mathcal{J}_1$ living inside $W_2 \cup \ldots \cup W_r$. (More formally, let $Z_1, \ldots, Z_{2^{k-1}}$ be disjoint sets such that $|E_i \cap Z_j| = 1$ for every $2 \leqslant i \leqslant r$ and $1 \leqslant j \leqslant 2^{k-1}$. Then for each Z_j let ζ_j be the unique $(r-1)$-partite map from $[2,k]$ to Z_j and let $\mathcal{K}_4$ consist of all sets of the form $\zeta_j(A)$ with $A \in \mathcal{J}_1$.)

We now let $\mathcal{K} = \mathcal{K}_2 \cup \mathcal{K}_3 \cup \mathcal{K}_4$ and let Ω be the set of all r-partite maps ω from $F_1 \cup \ldots \cup F_r$ to $X_1 \cup \ldots \cup F_r$. Such a map is a homomorphism if and only if its restrictions to W_1 and to $W_2 \cup \ldots \cup W_r$ are homomorphisms from $\mathcal{K}_3$ to $\mathcal{H}$ and from $\mathcal{K}_4$ to $\mathcal{H}$ respectively. It follows easily that

$$M^{2^{k-1}}|X_1|^{2^k-2}\sum_{\omega\in\Omega_2}\prod_{C\in\mathcal{K}_2}H^C(\omega)\prod_{B\in\mathcal{B}}f^B(\omega)=\sum_{\omega\in\Omega}\prod_{C\in\mathcal{K}}H^C(\omega)\prod_{B\in\mathcal{B}}f^B(\omega)\ .$$

Step 6. *Checking that $\mathcal{K}$ has the properties claimed.*

It remains to prove that for each non-empty set $A \in \mathcal{J}$ there are exactly 2^k sets in $\mathcal{K}$ of index A, and that these sets are all the non-empty sets in $\mathcal{K}$. (It is clear that $\mathcal{K}$ depends only on $\mathcal{J}$.) So let $A \in \mathcal{J}$. If $1 \notin A$ then there are 2^{k-1} copies of A in $\mathcal{K}_4$ and no copies of A in $\mathcal{K}_3$. We also have that $A \in \mathcal{J}_1 \cap \mathcal{J} = \mathcal{J}_3$, so, as we showed earlier, there are 2^{k-1} sets C in $\mathcal{K}_1$ such that $\gamma'(C) = A$. Each of these sets belongs to $\mathcal{K}_2$. If $1 \in A$ then let $B = A \setminus \{1\}$. If B is non-empty then again there are 2^{k-1} sets C in $\mathcal{K}'$ such that $\gamma'(C) = B$. Since $1 \cup B \in \mathcal{J}$, we have $B \in \mathcal{J}_4$, so for each such C the sets $\{1\} \cup C$ and $\{1'\} \cup C$ belong to $\mathcal{K}_2$. Therefore, $\mathcal{K}_2$ contains 2^k sets with index A. Neither $\mathcal{K}_3$ nor $\mathcal{K}_4$ contains any sets with index A. If B is empty then $A = \{1\}$ and there are two sets C in $\mathcal{K}_2$ of index A, namely $\{1\}$ and $\{1'\}$. Together with the singletons $\{a\} \in \mathcal{K}_3$ this again makes 2^k sets in $\mathcal{K}$. $\square$

§5. A counting lemma for quasirandom chains.

In this section we shall prove that a sufficiently quasirandom r-partite chain $\mathcal{H}$ behaves like a random chain in the following respect: if $\mathcal{J}$ is a fixed chain (we imagine that $\mathcal{H}$ is varying and its size is tending to infinity), then the number of r-partite maps from the vertices of $\mathcal{J}$ to the vertices of $\mathcal{H}$ that are homomorphisms is roughly what would be expected if $\mathcal{H}$ had been chosen randomly.

The main result of the section is in fact a generalization of the counting lemma itself. We state the more general form because it lends itself more easily to an inductive proof, and is in fact a useful statement in its own right. A few preliminary words about the result may make the formal statement easier to take in.

Let F and f be as defined in the statement of Lemma 3.1. The main idea behind the counting lemma, and the justification for the definition of octahedral quasirandomness, is that if $\mathrm{Oct}(f)$ is small and $\mathcal{H}$ behaves like a random chain then $\sum_{\phi\in\Phi} F(\phi)$ is also small. One can give a heuristic argument for this using the inequality from part (iv) of Lemma 3.1 as follows. Let $\mathcal{K}_1$ be the chain of all subsets of all sets $B \in \mathcal{B}$. Then $\prod_{B\in\mathcal{B}} f^B(\omega)$ is non-zero only if $H^C(\omega) = 1$ for every $C \in \mathcal{K}_1$. Therefore, we can rewrite the right-hand side of the inequality as

$$\sum_{\omega\in\Omega} \prod_{C\in\mathcal{K}\setminus\mathcal{K}_1} H^C(\omega) \prod_{B\in\mathcal{B}} f^B(\omega) \ .$$

We would expect that if ω is chosen randomly then the probability that $\prod_{C\in\mathcal{K}\setminus\mathcal{K}_1} H^C(\omega) = 1$ given that $\prod_{B\in\mathcal{B}} f^B(\omega) \neq 0$ is around $\prod_{C\in\mathcal{K}\setminus\mathcal{K}_1} \delta_C$. Furthermore, since no set in $\mathcal{K}\setminus\mathcal{K}_1$ is contained in any set in $\mathcal{K}_1$, we expect this probability to be more or less independent of the value of $\prod_{B\in\mathcal{B}} f^B(\omega)$. In other words, we expect that

$$\sum_{\omega\in\Omega} \prod_{C\in\mathcal{K}\setminus\mathcal{K}_1} H^C(\omega) \prod_{B\in\mathcal{B}} f^B(\omega) \ \approx \ \prod_{C\in\mathcal{K}\setminus\mathcal{K}_1} \delta_C \sum_{\omega\in\Omega} \prod_{B\in\mathcal{B}} f^B(\omega) \ . \qquad (*)$$

We count $\mathrm{Oct}(f)$ as small if $\sum_{\omega\in\Omega}\prod_{B\in\mathcal{B}} f^B(\omega)$ is small compared with $|\Omega|\prod_{C\in\mathcal{K}_1}\delta_C$. Putting these two estimates together shows that $\left(\sum_{\phi\in\Phi} F(\phi)\right)^{2^k}$ is small compared with $|\Omega|\prod_{C\in\mathcal{K}}\delta_C$. Since for each set in $\mathcal{J}$ the chain $\mathcal{K}$ contains 2^k sets of the same index, this is equal to $\left(|\Phi|\prod_{A\in\mathcal{J}}\delta_A\right)^{2^k}$. If $\mathcal{H}$ is sufficiently quasirandom, then $|\Phi|\prod_{A\in\mathcal{J}}\delta_A$ will be roughly the size of $\mathrm{Hom}(\mathcal{J},\mathcal{H})$, so it is appropriate to claim that $\sum_{\phi\in\Phi} F(\phi)$ is small.

In order to make the above ideas rigorous we must somehow take each function $H^C(\omega)$ on the left-hand side of $(*)$ and replace it by the constant function δ_C. We do this by exploiting the quasirandomness of H^C. Roughly speaking, that tells us that the difference between H^C and δ_C is octahedrally quasirandom relative to the chain, so we can try to use Lemma 3.1 to prove that replacing H^C by δ_C will not make much difference. This seems to raise exactly the same difficulties that we are already trying to solve, but if we are careful we can assume that in this instance they have been dealt with by induction.

Theorem 5.1, then, is a general result that allows us, under appropriate circumstances, to take expressions involving functions H^A and replace some of them by constant functions δ_A. The proof is a more detailed version of the argument just sketched. In order for the proof to work, certain relationships must hold between the relative densities and the quasirandomness parameters. So that we do not have to keep repeating these, and so that the statements of our results are not too cumbersome, let us formulate a definition first.

Suppose that $\mathcal{J}$ and $\mathcal{H}$ are r-partite chains. For each $A \in \mathcal{J}$, let the relative density of $H(A)$ in $\mathcal{H}$ be δ_A and suppose that $H(A)$ is relatively η_A-quasirandom. For each s let $\delta_{\geqslant s}$ be the product of the relative densities δ_A over all $A \in \mathcal{J}$ of size at least s. Let $0 < \epsilon < 1$ and define a pair of sequences $\eta_2, \dots, \eta_k$ and $\epsilon_2, \dots, \epsilon_k$ inductively, by setting $\epsilon_k = \epsilon$ and for each $s \leqslant k$ setting $\eta_s = \frac{1}{2}(\epsilon_s\delta_{\geqslant s})^{2^s}$ and $\epsilon_{s-1} = \eta_s|\mathcal{J}|^{-1}\prod_{t=s}^k 2^{-t}$. We shall say that the chain $\mathcal{H}$ is $(\epsilon, \mathcal{J}, k)$-*quasirandom* if $\eta_A \leqslant \eta_s$ for each $s \leqslant k$ and each set A of size s. For much of the argument that follows, it is not necessary to remember the details of this definition: basically, $\mathcal{H}$ is $(\epsilon, \mathcal{J}, k)$-quasirandom if the various parts of $\mathcal{H}$ are sufficiently quasirandom to make the proof of our counting lemma work. Notice one feature of the definition: that if $\mathcal{H}$ is $(\epsilon, \mathcal{J}, k)$-quasirandom and $\epsilon_2, \ldots, \epsilon_k$ are as defined above, then $\mathcal{H}$ is $(\epsilon_s, \mathcal{J}, s)$-quasirandom for each s between 2 and k.

Theorem 5.1. *Let $\mathcal{J}_0$ and $\mathcal{H}$ be r-partite chains with vertex sets $E_1 \cup \ldots \cup E_r$ and $X_1 \cup \ldots \cup X_r$, respectively. Let $\mathcal{J}_1$ be a subchain of $\mathcal{J}_0$ and let $G = \prod_{A \in \mathcal{J}_0} g^A$ be a $\mathcal{J}_0$-function on $\mathcal{H}$ such that each g^A is an A-function supported in $\mathcal{H}$ and $g^A = H^A$ whenever $A \in \mathcal{J}_0 \setminus \mathcal{J}_1$. Let k be the size of the largest set in $\mathcal{J}_0 \setminus \mathcal{J}_1$ and suppose that $\mathcal{H}$ is $(\epsilon, \mathcal{J}_0, k)$-quasirandom. Then*

$$\Big|\sum_{\phi \in \Phi} \prod_{A \in \mathcal{J}_0} g^A(\phi) - \prod_{A \in \mathcal{J}_0 \setminus \mathcal{J}_1} \delta_A \sum_{\phi \in \Phi} \prod_{A \in \mathcal{J}_1} g^A(\phi)\Big| \leqslant \epsilon |\mathcal{J}_0 \setminus \mathcal{J}_1| |\Phi| \prod_{A \in \mathcal{J}_0} \delta_A \ .$$

Proof. Our argument will be by a double induction. The outer induction will be on k and the inner one will be on the number $|\mathcal{J}_0 \setminus \mathcal{J}_1|$. If $\mathcal{J}_1 = \mathcal{J}_0$ then the result is trivial. Otherwise, let Φ be the set of r-partite maps from $E_1 \cup \ldots \cup E_r$ to $X_1 \cup \ldots \cup X_r$ and let A_0 be a set of size k in $\mathcal{J}_0 \setminus \mathcal{J}_1$. For each $\phi \in \Phi$ let $f(\phi) = H^{A_0}(\phi) - \delta_{A_0} \prod_{A \subsetneq A_0} H^A(\phi)$. Then let $F(\phi) = f(\phi)G(\phi)$. (This is a quick way of defining the function that has the same formula as G except that g^{A_0}, which equals H^{A_0}, is replaced by f. It works because $f = f g^{A_0}$.) Then the conditions of Lemma 3.1 are satisfied for $\mathcal{J}_0$, $\mathcal{H}$, A_0 and F: since A_0 is a maximal element of $\mathcal{J}_0 \setminus \mathcal{J}_1$ and $\mathcal{J}_1$ is a chain, $A_0 = A$ whenever $A_0 \subset A$, so $g^A = H^A$; the other conditions are obvious.

Now let $\mathcal{J}$ be the set of all edges of $\mathcal{J}_0$ of cardinality less than k. Let $\mathcal{K}$ be the chain provided for us by Lemma 3.1, so that, in the notation of that lemma, we have the inequality

$$\Big(\sum_{\phi \in \Phi} F(\phi)\Big)^{2^k} \leqslant \sum_{\omega \in \Omega} \prod_{C \in \mathcal{K}} H^C(\omega) \prod_{B \in \mathcal{B}} f^B(\omega) \ .$$

Letting $\mathcal{K}_1$ be the subchain of $\mathcal{K}$ consisting of all subsets of sets in $\mathcal{B}$, we can rewrite this as

$$\begin{aligned}\Big(\sum_{\phi \in \Phi} F(\phi)\Big)^{2^k} &\leqslant \sum_{\omega \in \Omega} \prod_{C \in \mathcal{K} \setminus \mathcal{K}_1} H^C(\omega) \prod_{B \in \mathcal{B}} f^B(\omega) \prod_{C \in \mathcal{K}_1 \setminus \mathcal{B}} H^C(\omega) \\ &= \sum_{\omega \in \Omega} \prod_{C \in \mathcal{K} \setminus \mathcal{K}_1} H^C(\omega) \prod_{B \in \mathcal{B}} f^B(\omega) \ ,\end{aligned}$$

since whenever $f_B(\omega) \neq 0$ we have $H^C(\omega) = 1$ for every $C \subset B$.

But the largest set in $\mathcal{K} \setminus \mathcal{K}_1$ has size less than k, and for each $s \leqslant k-1$ we have $\epsilon_{s-1} \leqslant \eta_s |\mathcal{J}_0|^{-1} \prod_{t=s}^{k} 2^{-t} \leqslant \eta_s |\mathcal{K}|^{-1} \prod_{t=s}^{k-1} 2^{-t}$, since $|\mathcal{K}| \leqslant 2^k |\mathcal{J}_0|$. This tells us that $\mathcal{H}$ is $(\epsilon_{k-1}, \mathcal{K}, k-1)$-quasirandom, which puts us in a position to apply our inductive hypothesis. This tells us that

$$\Big|\sum_{\omega\in\Omega} \prod_{C\in\mathcal{K}\setminus\mathcal{K}_1} H^C(\omega) \prod_{B\in\mathcal{B}} f^B(\omega) - \prod_{C\in\mathcal{K}\setminus\mathcal{K}_1} \delta_C \sum_{\omega\in\Omega} \prod_{B\in\mathcal{B}} f^B(\omega)\Big| \leqslant \epsilon_{k-1} |\mathcal{K}\setminus\mathcal{K}_1| |\Omega| \prod_{C\in\mathcal{K}} \delta_C \ .$$

It does not quite tell us this directly, since $\mathcal{B}$ is not a chain, but it does so if in addition we use the equality $\prod_{B\in\mathcal{B}} f^B(\omega) = \prod_{B\in\mathcal{B}} f^B(\omega) \prod_{C\in\mathcal{K}_1\setminus\mathcal{B}} H^C$, which we observed earlier.

By hypothesis,

$$\sum_{\omega\in\Omega} \prod_{B\in\mathcal{B}} f^B(\omega) \leqslant \eta_k |\Omega| \prod_{B\in\mathcal{K}_1} \delta_B \ ,$$

from which it follows that

$$\prod_{C\in\mathcal{K}\setminus\mathcal{K}_1} \delta_C \sum_{\omega\in\Omega} \prod_{B\in\mathcal{B}} f^B(\omega) \leqslant \eta_k |\Omega| \prod_{C\in\mathcal{K}} \delta_C \ ,$$

and therefore that

$$\Big|\sum_{\omega\in\Omega} \prod_{C\in\mathcal{K}\setminus\mathcal{K}_1} H^C(\omega) \prod_{B\in\mathcal{B}} f^B(\omega)\Big| \leqslant (\epsilon_{k-1} |\mathcal{K}\setminus\mathcal{K}_1| + \eta_k) |\Omega| \prod_{C\in\mathcal{K}} \delta_C \ .$$

Since to each set $A \in \mathcal{J}$ we associate 2^k sets $C \in \mathcal{K}$ with the same index, this is equal to $(\epsilon_{k-1}|\mathcal{K}\setminus\mathcal{K}_1| + \eta_k)\Big(|\Phi| \prod_{A\in\mathcal{J}} \delta_A\Big)^{2^k}$. But $|\mathcal{K}\setminus\mathcal{K}_1| < 2^k |\mathcal{J}_0|$ and so, by our assumptions about ϵ_{k-1} and η_k, we may conclude that

$$\Big|\sum_{\phi\in\Phi} F(\phi)\Big| \leqslant (\epsilon_{k-1}|\mathcal{K}\setminus\mathcal{K}_1| + \eta_k)^{1/2^k} |\Phi| \prod_{A\in\mathcal{J}} \delta_A \leqslant (2\eta_k)^{1/2^k} |\Phi| \prod_{A\in\mathcal{J}} \delta_A \ .$$

Since $(2\eta_k)^{1/2^k} \leqslant \epsilon_k \delta_{\geqslant k}$, this is at most $\epsilon_k |\Phi| \prod_{A\in\mathcal{J}_0} \delta_A$.

But

$$\sum_{\phi\in\Phi} F(\phi) = \sum_{\phi\in\Phi} \prod_{A\in\mathcal{J}_0} g^A(\phi) - \delta_{A_0} \sum_{\phi\in\Phi} \prod_{A\in\mathcal{J}_0, A\neq A_0} g^A(\phi) \ .$$

By our second induction (on $|\mathcal{J}_0 \setminus \mathcal{J}_1|$) we have that

$$\Big|\sum_{\phi\in\Phi} \prod_{A\in\mathcal{J}_0, A\neq A_0} g^A(\phi) - \prod_{A\in\mathcal{J}_0\setminus\mathcal{J}_1, A\neq A_0} \delta_A \sum_{\phi\in\Phi} \prod_{A\in\mathcal{J}_1} g^A(\phi)\Big|$$
$$\leqslant \epsilon_k (|\mathcal{J}_0 \setminus \mathcal{J}_1| - 1) |\Phi| \prod_{A\in\mathcal{J}_0, A\neq A_0} \delta_A \ ,$$

which implies that

$$\Big|\delta_{A_0}\sum_{\phi\in\Phi}\prod_{A\in\mathcal{J}_0,A\neq A_0}g^A(\phi)-\prod_{A\in\mathcal{J}_0\setminus\mathcal{J}_1}\delta_A\sum_{\phi\in\Phi}\prod_{A\in\mathcal{J}_1}g^A(\phi)\Big|\leqslant\epsilon(|\mathcal{J}_0\setminus\mathcal{J}_1|-1)|\Phi|\prod_{A\in\mathcal{J}_0}\delta_A\ .$$

Hence, by the triangle inequality,

$$\Big|\sum_{\phi\in\Phi}\prod_{A\in\mathcal{J}_0}g^A(\phi)-\prod_{A\in\mathcal{J}_0\setminus\mathcal{J}_1}\delta_A\sum_{\phi\in\Phi}\prod_{A\in\mathcal{J}_1}g^A(\phi)\Big|\leqslant\epsilon_k|\mathcal{J}_0\setminus\mathcal{J}_1||\Phi|\prod_{A\in\mathcal{J}_0}\delta_A$$

and the proof is complete. □

In the case where $\mathcal{J}_0=\mathcal{J}$ and $\mathcal{J}_1=\emptyset$, the function g^A in the statement of Theorem 5.1 is equal to H^A, for every $A\in\mathcal{J}$. Then we have $\sum_{\phi\in\Phi}\prod_{A\in\mathcal{J}_0}g^A(\phi)=|\operatorname{Hom}(\mathcal{J},\mathcal{H})|$. To complete the section, let us state this case of Theorem 5.1 separately. This is our "counting lemma" for quasirandom chains.

Corollary 5.2. *Let $\mathcal{J}$ and $\mathcal{H}$ be r-partite chains with vertex sets $E_1\cup\ldots\cup E_r$ and $X_1\cup\ldots\cup X_r$, respectively. Let k be the size of the largest set in $\mathcal{J}$ and suppose that $\mathcal{H}$ is $(\epsilon,\mathcal{J},k)$-quasirandom. Then*

$$\Big||\operatorname{Hom}(\mathcal{J},\mathcal{H})|-|\Phi|\prod_{A\in\mathcal{J}}\delta_A\Big|\leqslant\epsilon|\mathcal{J}||\Phi|\prod_{A\in\mathcal{J}}\delta_A\ .$$ □

§6. Local increases in mean-square density.

All known proofs of Szemerédi's theorem use (explicitly or implicitly) an approach of the following kind. Given a dense set that fails to be quasirandom in some appropriate sense, one can identify imbalances in the set that allow one to divide it into pieces that "improve" in some way, on average at least, on the set itself. One then iterates this argument until one reaches sets that *are* quasirandom. At that point one uses some kind of counting lemma to prove that they contain an arithmetic progression of length k.

This proof is no exception. We have defined a notion of quasirandomness and proved a counting lemma for it. Now we must see what happens when some parts of a chain are *not* relatively quasirandom. We shall end up proving a *regularity lemma*, which says, roughly speaking, that any dense chain can be divided up into a bounded number of pieces, almost all of which are quasirandom. This generalizes Szemerédi's regularity lemma for graphs (which formed part of his proof of his theorem on arithmetic progressions).

Given a dense graph G and a positive real number ϵ, Szemerédi's regularity lemma asserts that the vertices of G can be partitioned into K classes of roughly equal size, with

K bounded above by a function of ϵ only, in such a way that, proportionately speaking, at least $1-\epsilon$ of the bipartite graphs spanned by two of these classes are ϵ-quasirandom. (One can insist that K is much bigger than ϵ^{-1}, so it is not necessary to worry about the case where the two classes are equal. Or it can be neater to say that two equal classes form a "regular pair" if they span a quasirandom graph.)

Very roughly, the proof is as follows. Suppose you have a graph G and a partition of its vertex set. Then either this partition will do or there are many pairs of cells from the partition that give rise to induced bipartite subgraphs of G that are not ϵ-quasirandom. If X and Y are two disjoint sets of vertices, write $G(X,Y)$ for the corresponding induced bipartite subgraph of G. Suppose that X and Y are two cells of the partition, for which $G(X,Y)$ is not ϵ-quasirandom. Then it is not hard to show that there are large subsets $X(0)\subset X$ and $Y(0)\subset Y$ for which the density of $G(X(0),Y(0))$ is substantially different from that of $G(X,Y)$. Letting $X(1)=X\setminus X(0)$ and $Y(1)=Y\setminus Y(0)$, we have obtained partitions of X and Y into two sets each, in such a way that the densities of the graphs $G(X(i),Y(j))$ are not almost all approximately the same as that of $G(X,Y)$. One can then define an appropriately weighted average of the squares of these four densities and show that this average is greater than the square of the density of $G(X,Y)$. Let us call this *stage one* of the argument, the stage where we identify a "local" increase in mean-square density.

It remains to turn these local increases into a global increase. This, which we shall call *stage two*, is quite simple. Denote the cells of the partition by $X_1,\dots,X_k$. For each pair (X_i,X_j) that fails to be ϵ-regular, use the above argument to partition X_i into two sets $X_{ij}(0)$ and $X_{ij}(1)$, and to partition X_j into two sets $X_{ji}(0)$ and $X_{ji}(1)$. Then for each i find a partition of X_i that refines all the partitions $\{X_{ij}(0),X_{ij}(1)\}$. The result is a partition into $m\leqslant k.2^k$ sets $Y_1,\dots,Y_m$ that refines the partition $\{X_1,\dots,X_k\}$. It can be shown that the average of the squares of the densities $G(Y_i,Y_j)$, again, with appropriate weights, is significantly greater than it was for the partition $\{X_1,\dots,X_k\}$. Therefore, if one iterates the procedure, the iteration must terminate after a number of steps that can be bounded in terms of ϵ. It can terminate only if almost all the graphs $G(X_i,X_j)$ are quasirandom, so the result is proved.

We have given this sketch since our generalized regularity lemma will be proved in a similar way. There are two main differences. First, it is an unfortunate fact of life that, when one is dealing with k-chains rather than graphs, simple arguments have to be expressed in terminology that can obscure their simplicity. For example, even defining the appropriate notion of a "partition" of a chain is somewhat complicated. Thus, stage two of our argument, although it is an "obvious" generalization of stage two of the proof of the usual regularity lemma, is noticeably more complicated to write down.

A more fundamental difference, however, is that our stage one is not completely straightforward, and here the difference is mathematical rather than merely notational. For an explanation of why it should be harder, see the beginning of §8, and also §10, of [G2]. (It is not, however, substantially harder for general k than it is for $k = 3$, the case dealt with in that paper.) This local stage will occupy us for the rest of this section.

To begin with, let us say what we mean by the mean-square density of a function with respect to a partition. Let U be a set of size n, let $f : U \to \mathbb{R}$ and let $B_1, \dots, B_r$ be sets that form a partition of U. Then the *mean-square density* of f *with respect to* the partition $\{B_1, \dots, B_r\}$ is

$$\sum_{i=1}^{r} \frac{|B_i|}{n} \Big(|B_i|^{-1} \sum_{x \in B_i} f(x)\Big)^2 .$$

If we write β_i for $|B_i|/n$ (which it is helpful to think of as the probability that a random $x \in U$ is an element of B_i) and δ_i for $|B_i|^{-1} \sum_{x \in B_i} f(x)$ (that is, the expectation, or "density", of f in B_i) then this sum is $\sum_{i=1}^{r} \beta_i \delta_i^2$, the average of the squared densities δ_i^2, with a natural system of weights β_i.

The following two simple lemmas are lifted straight out of [G2]. The first is our main tool, while the second is more of a technical trick that will be used in the main result of this section, Lemma 6.3.

Lemma 6.1. *Let U be a set of size m and let f and g be functions from U to the interval $[-1, 1]$. Let $B_1, \dots, B_r$ be a partition of U and suppose that g is constant on each B_i. Then the mean-square density of f with respect to the partition $B_1, \dots, B_r$ is at least $\langle f, g \rangle^2 / m \|g\|_2^2$.*

Proof. For each j let a_j be the value taken by g on the set B_j. Then, by the Cauchy-Schwarz inequality,

$$\begin{aligned}
\langle f, g \rangle &= \sum_j a_j \sum_{x \in B_j} f(x) \\
&\leqslant \Big(\sum_j |B_j| a_j^2\Big)^{1/2} \Big(\sum_j |B_j|^{-1} \sum_{x \in B_j} f(x)\Big)^{1/2} \\
&= \|g\|_2 \Big(\sum_j |B_j| \Big(|B_j|^{-1} \sum_{x \in B_j} f(x)\Big)^2\Big)^{1/2} .
\end{aligned}$$

But $\sum_j |B_j|\big(|B_j|^{-1}\sum_{x\in B_j} f(x)\big)^2$ is m times the mean-square density of f (by definition), so the lemma follows. $\square$

Lemma 6.2. *Let m and n be positive integers, let $0<\delta<1$ and let r be an integer greater than or equal to δ^{-1}. Let $v_1,\dots,v_n$ be vectors in ℓ_2^m such that $\|v_i\|^2\leqslant m$ for each i and such that $\|\sum v_i\|_2^2\leqslant \delta n^2 m$. Let r vectors $w_1,\dots,w_r$ be chosen uniformly and independently from the v_i. (To be precise, for each w_j an index i is chosen randomly between 1 and n and w_j is set equal to v_i.) Then the expectation of $\|\sum w_j\|_2^2$ is at most $2\delta r^2 m$.*

Proof. The expectation of $\|\sum w_j\|_2^2$ is the expectation of $\sum_{i,j}\langle w_i,w_j\rangle$. If $i\neq j$ then the expectation of $\langle w_i,w_j\rangle$ is $n^{-2}\|\sum v_i\|_2^2$ which, by hypothesis, is at most δm. If $i=j$, then $\langle w_i,w_j\rangle$ is at most m, again by hypothesis. Therefore, the expectation we are trying to bound is at most $(\delta r(r-1)+r)m$. Since $\delta r\geqslant 1$, this is at most $2\delta r^2 m$, as claimed. $\square$

Before we state the main result of this section, we need two definitions. The first is of a chain $\mathcal{D}$ that we shall call a *double octahedron*. We use this name for conciseness even though it is slightly misleading: in fact, $\mathcal{D}$ is the $(k-1)$-skeleton of a chain formed from two k-dimensional octahedra by identifying a face from one with the corresponding face from the other. To put this more formally, take the vertex set of $\mathcal{D}$ to be the set $[k]\times\{0,1,2\}$. For each i between 1 and k let V_i be the set $\{i\}\times\{0,1,2\}$ and for $j=0,1,2$ let B_j be the set $[k]\times\{j\}$. The edges of $\mathcal{D}$ are all sets B of cardinality at most $k-1$ such that $|B\cap V_i|\leqslant 1$ for every i and at least one of $B\cap B_1$ and $B\cap B_2$ is empty. (The two octahedra in question are O_1 and O_2, where O_j consists of all sets $B\subset B_0\cup B_j$ such that $|B\cap V_i|\leqslant 1$ for every i.)

Notice that if $A\subset[k]$ is a set of size at most $k-1$ then the number of edges in $\mathcal{D}$ of index A is $2^{|A|+1}-1$, since there are $2^{|A|}$ edges from each octahedron and one, namely $A\times\{0\}$, which is common to both.

For the second definition, suppose we have a k-partite $(k-1)$-chain $\mathcal{H}$ with vertex sets $X_1,\dots,X_k$. Recall from §2 that $H_*([k])$ is the collection of all sets A such that $|A\cap X_i|=1$ for every i and such that every proper subset of A belongs to $\mathcal{H}$. For this second condition to hold it is enough for C to be an edge of $\mathcal{H}$ whenever $C\subset A$ and $|C|=k-1$. Let H be the k-partite $(k-1)$-uniform hypergraph consisting of all edges of $\mathcal{H}$ of size $k-1$. For $1\leqslant i\leqslant k$ let H_i be the $(k-1)$-partite subhypergraph of H consisting of all edges of H that have empty intersection with X_i. We shall call the hypergraphs H_i the *parts* of H.

Each set $A \in H_*([k])$ has k subsets of size $k-1$. Each part H_i of H contains exactly one of these subsets, namely $A \setminus X_i$.

Suppose that each H_i is partitioned into subhypergraphs $H_{i1}, \dots, H_{ir_i}$. These partitions give rise to an equivalence relation $\sim$ on $H_*([k])$: we say that $A \sim A'$ if, for each $i \leqslant k$, the sets $A \setminus X_i$ and $A' \setminus X_i$ belong to the same cell H_{ij} of the partition of H_i. The corresponding partition will be called the *induced partition* of $H_*([k])$.

Lemma 6.3. *Let $\mathcal{H}$ be a k-partite $(k-1)$-chain with vertex sets $X_1, \dots, X_k$, let $\mathcal{D}$ be the double octahedron, let $\delta = \prod_{A \in \mathcal{D}} \delta_A$ and let $r \geqslant \delta^{-1}$ be a positive integer. Suppose that $\epsilon \leqslant |\mathcal{D}|^{-1}$, that $\mathcal{H}$ is $(\epsilon, \mathcal{D}, k-1)$-quasirandom and that $f : H_*([k]) \to [-1, 1]$ is a function that is not η-quasirandom relative to $\mathcal{H}$. Let H be the set of all edges of $\mathcal{H}$ of size $k-1$ and let $H_1, \dots, H_k$ be the k parts of H. Then there are partitions of the H_i into at most 3^r sets each such that the mean-square density of f with respect to the induced partition of $H_*([k])$ is at least $\eta^2/8$.*

We shall prove Lemma 6.3 in stages, by means of some intermediate lemmas. But to begin with, let us examine our main hypothesis, that f is not η-quasirandom relative to $\mathcal{H}$. For each $i \leqslant k$ let $U_i = \{i\} \times \{0, 1\}$ (so U_i consists of the "first two" of the three elements of V_i). As in §2, let $\mathcal{B}$ be the k-partite k-uniform hypergraph consisting of all sets $B \subset U_1 \cup \ldots \cup U_k$ such that $|B \cap U_i| = 1$ for every i, let $\mathcal{K}$ be the chain of all sets C that are proper subsets of some $B \in \mathcal{B}$ and let Ω be the set of all k-partite maps from $U_1 \cup \ldots \cup U_k$ to $X_1 \cup \ldots \cup X_k$. Then to say that f is not η-quasirandom relative to $\mathcal{H}$ is to say that

$$\mathrm{Oct}(f) = \sum_{\omega \in \Omega} \prod_{B \in \mathcal{B}} f^B(\omega) > \eta |\Omega| \prod_{A \in \mathcal{K}} \delta_A \ ,$$

where by $f^B(\omega)$ we mean $f(\omega(B))$ if $\omega(B) \in H_*([k])$ and 0 otherwise.

Let B_0 and B_1 be as defined earlier, so that $U_1 \cup \ldots \cup U_k = B_0 \cup B_1$. Let Φ and Ψ be the set of all k-partite maps from B_0 and B_1, respectively, to $X_1 \cup \ldots \cup X_k$. There is an obvious one-to-one correspondence between Ω and $\Phi \times \Psi$: given any $\omega \in \Omega$, associate with it the pair (ϕ, ψ) where ϕ and ψ are the restrictions of ω to B_0 and B_1. This procedure is invertible: given a pair (ϕ, ψ), define an r-partite map ω by setting $\omega(x) = \phi(x)$ if $x \in B_0$ and $\omega(x) = \psi(x)$ if $x \in B_1$. From now on we shall identify Ω with $\Phi \times \Psi$ and freely pass from one to the other.

Let us split the product $\prod_{B \in \mathcal{B}} f^B(\omega)$ into two parts. We shall write $F(\omega)$ for $f^{B_0}(\omega)$ and $G(\omega)$ for $\prod_{B \in \mathcal{B}, B \neq B_0} f^B(\omega)$. Now if $\omega = (\phi, \psi)$ then $F(\omega)$ does not depend on ψ (since it depends only on $\omega(B_0) = \phi(B_0)$). To emphasize this, we shall write $G(\phi, \psi)$ for

$G(\omega)$ and $F(\phi)$ for $F(\omega)$. Our hypothesis now becomes

$$\sum_{\phi\in\Phi}\sum_{\psi\in\Psi} F(\phi)G(\phi,\psi) > \eta|\Phi||\Psi|\prod_{A\in\mathcal{K}}\delta_A\ .$$

Let us see why this is useful. First, note that there is another obvious one-to-one correspondence, this time between Φ and $X_1\times\ldots\times X_k$. It associates with a map $\phi\in\Phi$ the k-tuple $(\phi(1,0),\ldots,\phi(k,0))$, and the inverse associates with a k-tuple $(x_1,\ldots,x_k)\in\prod_{i=1}^k X_i$ the map $\phi: B_0\to X_1\cup\ldots\cup X_k$ that takes $(i,0)$ to x_i for each $i\leqslant k$. Therefore, the function F is basically another way of thinking about f. The inequality above can be regarded as saying that, for an average $\psi\in\Psi$, F has a certain correlation with the function $G_\psi:\phi\mapsto G(\phi,\psi)$. This is significant, because the functions G_ψ have a special form, as the next lemma shows.

Lemma 6.4. *Each function $G_\psi:\Phi\to[-1,1]$ can be written as a product of A-functions over sets $A\subset B_0$ of size $k-1$.*

Proof. By definition, $G_\psi(\phi)=\prod_{B\in\mathcal{B},B\neq B_0} f^B(\phi,\psi)$. Now $f^B(\phi,\psi)$ depends on $(\phi,\psi)(B)=\phi(B\cap B_0)\cup\psi(B\cap B_1)$ only. Therefore, if ψ is fixed, $f^B(\phi,\psi)$ depends on $\phi(B\cap B_0)$ only. Thus, the function $\phi\mapsto f^B(\phi,\psi)$ is a $(B\cap B_0)$-function defined on Φ. Since $B\neq B_0$, $|B\cap B_0|\leqslant k-1$. This proves that G_ψ is a product of A-functions over sets A of size at most $k-1$. However, if $B\subset A$, then the product of a B-function with an A-function is still an A-function. From this simple observation it now follows that G_ψ is a product of A-functions over sets A of size equal to $k-1$. □

Our next task is to construct some new functions E_ψ out of the G_ψ that have very similar properties but take values 0, 1 and -1 only.

Lemma 6.5. *There exist functions $E_\psi:\Phi\to\{-1,0,1\}$, one for each $\psi\in\Psi$, with the following properties. First, $E_\psi(\phi)$ is non-zero only if $(\phi,\psi)\in\mathrm{Hom}(\mathcal{K},\mathcal{H})$. Second, each E_ψ can be written as a product of $\{-1,0,1\}$-valued A-functions over subsets $A\subset B_0$ of size $k-1$. Third,*

$$\sum_{\phi\in\Phi}\sum_{\psi\in\Psi} F(\phi)E_\psi(\phi) > \eta|\Phi||\Psi|\prod_{A\in\mathcal{K}}\delta_A\ .$$

Proof. Let us fix $\psi\in\Psi$ and consider the function $G=G_\psi$. By Lemma 6.4 we can write it as a product of A-functions, where each A in the product is a subset of B_0 of size $k-1$. There are k such sets, namely $A_1,\ldots,A_k$, where for each i we set $A_i=B_0\setminus\{(i,0)\}$. So we can write $G(\phi)=\prod_{i=1}^k g_i(\phi)$ with g_i an A_i-function for each i.

Now define an A_i-function $u_i : \Phi \to \{-1, 0, 1\}$ randomly in the following natural way. Say that two maps ϕ and ϕ' are *equivalent* if $\phi(A_i) = \phi'(A_i)$ and choose one map from each equivalence class. Let ϕ be one of these representatives. If $g_i(\phi) \geqslant 0$ then let $u_i(\phi)$ equal 1 with probability $g_i(\phi)$ and 0 with probability $1 - g_i(\phi)$. If $g_i(\phi) < 0$ then let $u_i(\phi)$ equal -1 with probability $-g_i(\phi)$ and 0 with probability $1 + g_i(\phi)$. Then the expectation of $u_i(\phi)$ is $g_i(\phi)$. If ϕ' is equivalent to ϕ then let $u_i(\phi') = u_i(\phi)$.

Do the same for each equivalence class and make all the random choices independently. Finally, for each $\phi \in \Phi$ let $E_\psi(\phi) = \prod_{i=1}^k u_i(\phi)$.

Now $E_\psi(\phi)$ can be non-zero only if $u_i(\phi) \neq 0$ for every i, and this is the case (with probability 1) only if $g_i(\phi) \neq 0$ for every i, and hence only if $G(\phi) \neq 0$. We defined $G(\phi)$ to be $G_\psi(\phi) = \prod_{B \in \mathcal{B}, B \neq B_0} f^B(\phi, \psi)$. But $f^B(\phi, \psi) = 0$ unless $(\phi, \psi)(B) \in H_*([k])$, and this is true only if $(\phi, \psi)(C) \in \mathcal{H}$ for every proper subset C of B. Therefore this product is non-zero only if (ϕ, ψ) is a homomorphism from $\mathcal{K}$ to $\mathcal{H}$.

Since the choices of the different functions u_i were made independently and the expectation of $u_i(\phi)$ is $g_i(\phi)$, the expectation of $u_1(\phi) \dots u_k(\phi)$ is $g_1(\phi) \dots g_k(\phi) = G_\psi(\phi)$. Therefore, by linearity of expectation, the expectation of $\sum_\phi \sum_\psi F(\phi) E_\psi(\phi)$ is $\sum_\phi \sum_\psi F(\phi) G_\psi(\phi)$, which we have assumed to be at least $\eta |\Phi||\Psi| \prod_{A \in \mathcal{K}} \delta_A$. It follows that we can choose functions E_ψ with the desired properties. □

Lemma 6.6. *For each $\psi \in \Psi$ let E_ψ be the function constructed in Lemma 6.5, and let $\mathcal{D}$ be the double octahedron chain introduced before the statement of Lemma 6.3. Then*

$$\sum_{\phi \in \Phi} \Big(\sum_{\psi \in \Psi} E_\psi(\phi) \Big)^2 \leqslant 2|\Phi||\Psi|^2 \prod_{A \in \mathcal{D}} \delta_A \ .$$

Proof. The left-hand side of the inequality we wish to prove can be rewritten

$$\sum_{\phi \in \Phi} \sum_{\psi_1, \psi_2 \in \Psi} E_{\psi_1}(\phi) E_{\psi_2}(\phi) \ .$$

By Lemma 6.5, $E_{\psi_1}(\phi) E_{\psi_2}(\phi)$ is non-zero if and only if (ϕ, ψ_1) and (ϕ, ψ_2) belong to $\mathrm{Hom}(\mathcal{K}, \mathcal{H})$. Therefore, this sum is at most the number of triples (ϕ, ψ_1, ψ_2) such that (ϕ, ψ_1) and (ϕ, ψ_2) belong to $\mathrm{Hom}(\mathcal{K}, \mathcal{H})$.

In order to estimate how many such triples there are, we shall apply the counting lemma to the chain $\mathcal{D}$. Every edge of $\mathcal{D}$ is a subset of either $B_0 \cup B_1$ or $B_0 \cup B_2$. Let $\mathcal{K}_1$ be the set of all edges of the first kind and let $\mathcal{K}_2$ be the set of all edges of the second kind. Both $\mathcal{K}_1$ and $\mathcal{K}_2$ are chains and they intersect in a chain that consists of all proper subsets of B_0. Moreover, $\mathcal{K}_1$ is essentially the same chain as $\mathcal{K}$ (formally, it has different vertex sets but the edges are the same). As for $\mathcal{K}_2$, it is isomorphic to $\mathcal{K}$ in the following sense. Let γ be the bijection from $B_0 \cup B_1$ to $B_0 \cup B_2$ that takes $(i,0)$ to $(i,0)$ and $(i,2)$ to $(i,1)$. Then A is an edge of $\mathcal{K}_2$ if and only if $\gamma(A)$ is an edge of $\mathcal{K}$.

Let Θ be the set of all k-partite functions from $V_1 \cup \ldots \cup V_k$ (the vertex set of $\mathcal{D}$) to $X_1 \cup \ldots \cup X_k$. There is a one-to-one correspondence between Θ and $\Phi \times \Psi \times \Psi$ taking $\theta \in \Theta$ to $(\phi, \psi_1, \psi_2 \circ \gamma)$, where ϕ, ψ_1 and ψ_2 are the restrictions of θ to B_0, B_1 and B_2, respectively. Since $\mathcal{D} = \mathcal{K}_1 \cup \mathcal{K}_2$, a map $\theta \in \Theta$ belongs to $\mathrm{Hom}(\mathcal{D}, \mathcal{H})$ if and only if (ϕ, ψ_1) belongs to $\mathrm{Hom}(\mathcal{K}_1, \mathcal{H})$ and (ϕ, ψ_2) belongs to $\mathrm{Hom}(\mathcal{K}_2, \mathcal{H})$. But this is true if and only if (ϕ, ψ_1) and $(\phi, \psi_2 \circ \gamma)$ belong to $\mathrm{Hom}(\mathcal{K}, \mathcal{H})$. (Note that $\psi_2 \circ \gamma$ here is the ψ_2 in the sum that we are estimating.)

What this shows is that the number of triples we wish to count is equal to the cardinality of $\mathrm{Hom}(\mathcal{D}, \mathcal{H})$. Since we are assuming that $\mathcal{H}$ is $(\epsilon, \mathcal{D}, k-1)$-quasirandom and that $\epsilon \leqslant |\mathcal{D}|^{-1}$, the counting lemma (Corollary 5.2) implies that this is at most $2|\Theta| \prod_{A \in \mathcal{D}} \delta_A = 2|\Phi||\Psi|^2 \prod_{A \in \mathcal{D}} \delta_A$, which proves the lemma. $\square$

Our next task is to show that we can make a small selection of the functions E_ψ and keep properties similar to those proved in the last two lemmas. The selection will be done in the obvious way: randomly.

Lemma 6.7. *Let $\delta = 2\prod_{A \in \mathcal{D}} \delta_A$, let $\beta = \prod_{A \in \mathcal{K}} \delta_A$ and let $r \geqslant \delta^{-1}$ be a positive integer. Then there exist functions $E_1, \ldots, E_r$ from Φ to $\{-1, 0, 1\}$ with the following three properties.*

(i) Each function E_i is a product of $\{-1,0,1\}$-valued A-functions over subsets $A \subset B_0$ of size $k-1$.

(ii) *For each i and each $\phi \in \Phi$, $E_i(\phi)$ is non-zero only if $\phi(B_0) \in \mathcal{H}$.*

(iii) $\sum_{i=1}^r \sum_{\phi \in \Phi} F(\phi) E_i(\phi) \geqslant (\eta/2)\beta r |\Phi|$.

(iv) $\sum_{\phi \in \Phi} \Big(\sum_{i=1}^r E_i(\phi) \Big)^2 \leqslant (4r\delta/\eta\beta) \sum_{i=1}^r \sum_{\phi \in \Phi} F(\phi) E_i(\phi)$.

Proof. For each i let E_i be one of the functions E_ψ, where ψ is chosen uniformly at random from Ψ. Let the choices be independent (so, in particular, the E_i are not necessarily distinct, though they probably will be). Then it follows from Lemma 6.5 that property (i) holds, and also that the expectation of $\sum_{i=1}^r \sum_{\phi \in \Phi} F(\phi) E_i(\phi)$ is at least $\eta\beta r|\Phi|$.

We now want to estimate the expectation of $\sum_{\phi \in \Phi} \Big(\sum_{i=1}^r E_i(\phi) \Big)^2$, and for this we shall use Lemma 6.2, the technical lemma from the beginning of the section. Set $n = |\Psi| = |\Phi|$ and let the vectors $v_1, \ldots, v_n$ be the functions E_ψ, which we regard as elements of $\ell_2(\Phi)$. Lemma 6.6 tells us that $\left\|\sum_{i=1}^n v_i\right\|_2^2 \leqslant \delta n^3$. Therefore, Lemma 6.2 tells us that the expectation of $\left\|\sum_{i=1}^r E_i\right\|_2^2$, which is the same as the expectation of $\sum_{\phi\in\Phi}\Big(\sum_{i=1}^r E_i(\phi)\Big)^2$, is at most $2\delta r^2 n = 2\delta r^2|\Phi|$.

It follows that the expectation of

$$4r\delta\sum_{i=1}^r\sum_{\phi\in\Phi} F(\phi)E_i(\phi) - \eta\beta\sum_{\phi\in\Phi}\Big(\sum_{i=1}^r E_i(\phi)\Big)^2$$

is at least $4\eta\beta\delta r^2|\Phi| - 2\eta\beta\delta r^2|\Phi| = 2\eta\beta\delta r^2|\Phi|$. It follows that there must be some choice of the functions $E_1,\dots,E_r$ such that the inequalities (iii) and (iv) are satisfied.

Since each E_i is one of the functions E_ψ, Lemma 6.5 implies that $E_i(\phi)$ is non-zero only if $(\phi,\psi)\in\mathrm{Hom}(\mathcal{K},\mathcal{H})$ for some $\psi\in\Psi$. But a necessary condition for this is that $\phi(B_0)\in\mathcal{H}$, so property (ii) is true as well. $\square$

Proof of Lemma 6.3. For each i let us write E_i as a product $\prod_{j=1}^k E_{ij}$, where E_{ij} is a $\{-1,0,1\}$-valued A_i-function. (As in the proof of Lemma 6.5, A_i is the set $B_0\setminus\{(i,0)\}$.)

For each $j\leqslant k$ we can partition the part H_j of H into at most 3^r sets, such that on each of these sets the function E_{ij} is constant for every $i\leqslant r$. Let $Z_1,\dots,Z_N$ be the corresponding induced partition of $H_*([k])$. Then every function E_i is constant on every cell Z_j, from which it follows that the function $g(\phi)=\sum_{i=1}^r E_i(\phi)$ is constant on every cell Z_j.

We are now in a position to apply Lemma 6.1. Property (ii) of Lemma 6.7 tells us that $\langle F,g\rangle\geqslant(\eta/2)\beta r|\Phi|$, and property (iii) tells us that $\langle F,g\rangle/\|g\|_2^2\geqslant\eta\beta/4r\delta$. Let U be the set of all $\phi\in\Phi$ such that $\phi(B_0)\in H_*([k])$. The map $\phi\mapsto\phi(B_0)$ is a bijection between U and $H_*([k])$, so we can regard $Z_1,\dots,Z_N$ as a partition of U.

Lemma 6.7 and these estimates tell us that the mean-square density of F with respect to this partition is at least

$$\frac{(\eta/2)\beta r|\Phi|}{|U|}\cdot\frac{\eta\beta}{4r\delta}=\frac{\eta^2\beta^2|\Phi|}{8\delta|U|}\ .$$

We also know that $|U|=|\Phi|\prod_{A\subsetneq B_0}\delta_A$. Recall that every set $A\subset B_0$ is the index of precisely $2^{|A|+1}-1$ sets in $\mathcal{D}$ and $2^{|A|}$ sets in $\mathcal{K}$. It follows that $\beta^2=\delta\prod_{A\subsetneq B_0}\delta_A$. Therefore, the mean-square density of F with respect to the partition $Z_1,\dots,Z_N$ is at least $\eta^2/8$. Since $F(\phi)=f(\phi(B_0))$, this statement is equivalent to the statement of Lemma 6.3. $\square$

Corollary 6.8. *Let $\mathcal{H}$ be a k-partite $(k-1)$-chain with vertex sets $X_1,\dots,X_k$, let $\mathcal{D}$ be the double octahedron, let $\delta=\prod_{A\in\mathcal{D}}\delta_A$ and let $r\geqslant\delta^{-1}$ be a positive integer. Suppose* *that $\epsilon \leqslant |\mathcal{D}|^{-1}$ and that $\mathcal{H}$ is $(\epsilon, \mathcal{D}, k-1)$-quasirandom. Let H^k be a k-partite k-uniform hypergraph with vertex sets $X_1, \dots, X_k$, let the density of H^k relative to $\mathcal{H}$ (that is, the quantity $|H^k|/|H_*([k])$) be $\delta_{[k]}$ and suppose that H^k is not η-quasirandom relative to $\mathcal{H}$. Let H be the set of all edges of $\mathcal{H}$ of size $k-1$ and let $H_1, \dots, H_k$ be the k parts of H. Then there are partitions of the H_i into at most 3^r sets each such that the mean-square density of (the characteristic function of) H^k with respect to the induced partition of $H_*([k])$ is at least $\delta_{[k]}^2 + \eta^2/8$.*

Proof. Let $f : H_*([k]) \to [-1, 1]$ be the function $H^k - \delta_{[k]}$. Then the statement that H^k is not η-quasirandom relative to $\mathcal{H}$ is, by definition, the statement that f is not η-quasirandom relative to $\mathcal{H}$. Therefore, by Lemma 6.3, we can find partitions of the required kind for which the mean-square density of f with respect to the induced partition of $H_*([k])$ is at least $\eta^2/8$.

Let $Z_1, \dots, Z_N$ be the induced partition of $H_*([k])$ and for each $(x_1, \dots, x_k) \in Z_i$ let $G(x_1, \dots, x_k) = |H^k \cap Z_i|/|Z_i|$. Then the mean of G is the same as the mean of H^k, namely $\delta_{[k]}$. The variance of G is the mean-square density of f. The result follows. □

§7. The statement of a regularity lemma for r-partite chains.

Corollary 6.8 is stage one of the proof of our regularity lemma. In this short section we will introduce some definitions and state the lemma. The proof (or rather, stage two of the proof) will be given in §9.

Broadly speaking, the lemma says that we can take a k-uniform hypergraph H, regard it as a chain (by adding all subsets of edges of H) and decompose that chain into subchains almost all of which are quasirandom. This is a useful thing to do, because Corollary 5.2 gives us a good understanding of quasirandom chains. Thus, the regularity lemma and counting lemma combine to allow us to decompose any (dense) k-uniform hypergraph into pieces that we can control. In the final section of the paper we shall exploit this by proving a generalization of Theorems 1.3 and 1.6 to k-uniform hypergraphs, which implies the multidimensional Szemerédi theorem.

Our principal aim will be to understand a certain $(k+1)$-partite k-uniform hypergraph. However, for the purposes of formulating a suitable inductive hypothesis it is helpful to prove a result that is more general in two ways. First of all, we shall look at r-partite k-uniform hypergraphs. Secondly, rather than looking at single hypergraphs we shall look at *partitions.* To be precise, let $X_1, \dots, X_r$ be a sequence of finite sets. Given any subset $A \subset [r]$, $A = \{i_1, \dots, i_s\}$, let $K(A)$ be the *complete s-uniform hypergraph* on the sets

$X_{i_1}, \ldots, X_{i_s}$, that is, the hypergraph consisting of all subsets of $X_1 \cup \ldots \cup X_r$ of index A. (It is part of the definition of "index" that such a set should intersect each X_i in at most a singleton.) For each $s \leqslant r$, the *complete s-uniform hypergraph* $K_s = K_s(X_1, \ldots, X_r)$ on the sets $X_1, \ldots, X_r$ is the union of the hypergraphs $K(A)$ over all sets $A \subset [r]$ of size s. Finally, the *complete k-chain* on $X_1, \ldots, X_r$, denoted $\mathcal{K}_k(X_1, \ldots, X_r)$, is the union of all $K(A)$ such that A has cardinality at most k: that is, it consists of all subsets of $X_1 \cup \ldots \cup X_r$ of size at most k that intersect each X_i at most once.

To form an arbitrary r-partite s-uniform hypergraph H with vertex sets $X_1, \ldots, X_r$, one can choose, for each $A \subset [r]$ of size s, a subset $H(A) \subset K(A)$ and let H be the union of these hypergraphs $H(A)$. If we want to, we can regard each $H(A)$ as a partition of $K(A)$ into the two sets $H(A)$ and $K(A) \setminus H(A)$. Our regularity lemma will be concerned with more general partitions, but it will imply a result for hypergraphs as an easy corollary.

Suppose now that for every subset $A \subset [r]$ of size at most k we have a partition of the hypergraph $K(A)$. If B and B' are two edges of this hypergraph (that is, if they are two sets of index A), let us write $B \sim_A B'$ if B and B' lie in the same cell of the partition, and say that B and B' are *A-equivalent.*

One can use these equivalence relations to define finer ones as follows. Given two sets B, B' of index A and given any subset $C \subset A$, there are unique subsets $D \subset B$ and $D' \subset B$ of index C. Let us say that B and B' are *C-equivalent* if D and D' are. Then let us say that B and B' are *strongly equivalent* if they are C-equivalent for every subset $C \subset A$.

Given this system of equivalence relations, we can define a collection of chains as follows. For every r-tuple $x = (x_1, \ldots, x_r) \in X_1 \times \ldots \times X_r$ and every set A of size at most k, let $x(A)$ be the set $\{x_i : i \in A\}$ and let $H(A, x)$ be the hypergraph consisting of all sets B that are strongly equivalent to $x(A)$.

Lemma 7.1. *The union $\mathcal{H} = \mathcal{H}(x)$ of the hypergraphs $H(A, x)$ over all sets A of size at most k is an r-partite k-chain.*

Proof. Let $B \in H(A, x)$ and let $D \subset B$. Let C be the index of D. Since B is strongly equivalent to $x(A)$, D is strongly equivalent to $x(C)$. Therefore $D \in H(C, x)$ and the lemma is proved. $\square$

Lemma 7.2. *Let $x = (x_1, \ldots, x_r)$ and $y = (y_1, \ldots, y_r)$ belong to the set $X_1 \times \ldots \times X_r$ and let $\mathcal{H}(x)$ and $\mathcal{H}(y)$ be the two chains constructed as above. Then for every set $A \subset [r]$ of size at most k, the hypergraphs $H(A, x)$ and $H(A, y)$ are either equal or disjoint.*

Proof. Suppose that B is a set of index A and that $B \in H(A, x) \cap H(A, y)$. Then B

is strongly equivalent to both $x(A)$ and $y(A)$, so these two sets are strongly equivalent to each other. It follows that $H(A, x) = H(A, y)$. $\square$

Let us call two r-partite k-chains $\mathcal{H}$ and $\mathcal{H}'$ with the same vertex sets $X_1, \dots, X_r$ *compatible* if, for every subset $A \subset [r]$ of size at most k, the hypergraphs $H(A)$ and $H'(A)$ are either equal or disjoint. By a *chain decomposition* of the complete r-partite k-chain $\mathcal{K}_k(X_1, \dots, X_r)$ we mean a set $\{\mathcal{H}_1, \dots, \mathcal{H}_N\}$ of chains with the following two properties:

(i) for every i and j the chains $\mathcal{H}_i$ and $\mathcal{H}_j$ are compatible;

(ii) for every sequence $(x_1, \dots, x_r) \in X_1 \times \dots \times X_r$ there is precisely one chain from the set $\{\mathcal{H}_1, \dots, \mathcal{H}_N\}$ that contains every subset of $\{x_1, \dots, x_r\}$ of size at most k.

Note that a chain decomposition is not a partition of $\mathcal{K}_k(X_1, \dots, X_r)$. There is no interesting way to partition $\mathcal{K}_k(X_1, \dots, X_r)$ into subchains, as a moment's thought will reveal. Lemmas 7.1 and 7.2 shows that the chains $\mathcal{H}(x)$ form a chain decomposition of $\mathcal{K}_k(X_1, \dots, X_r)$. (It may be that $\mathcal{H}(x) = \mathcal{H}(y)$, but this does not contradict (ii) because we have carefully defined a chain decomposition to be a *set* of chains rather than a *sequence* of chains.)

We are now ready to state our regularity lemma.

Theorem 7.3. *Let $\mathcal{J}$ be an r-partite k-chain with vertex sets $E_1, \dots, E_r$ and let $0 < \epsilon \leqslant |\mathcal{J}|^{-1}$. Let $X_1, \dots, X_r$ be a sequence of finite sets and for each subset $A \subset [r]$ of size at most k let $\mathcal{P}(A)$ be a partition of the hypergraph $K(A)$ into n_A sets. Then there are refinements $\mathcal{Q}(A)$ of the partitions $\mathcal{P}(A)$ leading to a chain decomposition of $\mathcal{K}_k(X_1, \dots, X_r)$ with the following property: if $x = (x_1, \dots, x_r)$ is a randomly chosen element of $X_1 \times \dots \times X_r$ then the probability that the chain $\mathcal{H}(x)$ is $(\epsilon, \mathcal{J}, k)$-quasirandom is at least $1 - \epsilon$. Moreover, $\mathcal{Q}(A) = \mathcal{P}(A)$ when $|A| = k$, and for general A the number of sets m_A in the partition $\mathcal{Q}(A)$ depends only on ϵ, $\mathcal{J}$, k and the numbers n_C.*

Before we start on the proof, let us comment on how we shall actually use Theorem 7.3. We will be presented with an r-partite k-uniform hypergraph H with vertex sets $X_1, \dots, X_r$. All the $\binom{r}{k}$ k-partite parts $H(A)$ of H will have density at least a certain fixed $\delta > 0$. We shall then apply Theorem 7.3 to the partitions $\mathcal{P}(A)$ defined as follows. If $|A| = k$ then $\mathcal{P}(A)$ will be $\{H(A), K(A) \setminus H(A)\}$. If $|A| < k$ then it will be the trivial partition $\{K(A)\}$. In this case, the result will tell us that we can find partitions $\mathcal{Q}(A)$ such that almost all edges of $\mathcal{H}$ lie in quasirandom chains from the decomposition determined by the partitions $\mathcal{Q}(A)$.

§8. Basic facts about partitions and mean-square density.

In order to prove a regularity lemma for systems of partitions, we need to generalize the notion of mean-square density as follows. Let $\mathcal{P} = \{X_1, \ldots, X_r\}$ and $\mathcal{Q} = \{Y_1, \ldots, Y_s\}$ be two partitions of a finite set U. Then the *mean-square density of* $\mathcal{P}$ *with respect to* $\mathcal{Q}$ is the quantity

$$\sum_{i=1}^{r}\sum_{j=1}^{s}\frac{|Y_j|}{|U|}\left(\frac{|X_i\cap Y_j|}{|Y_j|}\right)^2 ,$$

that is, the sum of all the mean-square densities of the sets X_i (that is, the mean-square densities of their characteristic functions, as defined in §6) with respect to $\mathcal{Q}$.

Since the numbers $|X_i \cap Y_j|/|Y_j|$ are non-negative and sum to 1, we have the simple upper bound

$$\sum_{i=1}^{r}\sum_{j=1}^{s}\frac{|Y_j|}{|U|}\left(\frac{|X_i\cap Y_j|}{|Y_j|}\right)^2 \leqslant \sum_{j=1}^{s}\frac{|Y_j|}{|U|} = 1$$

for this quantity. An alternative way of seeing this, which will be helpful later, is to notice that each $u \in U$ is contained in a unique X_i and a unique Y_j, and the mean-square density of $\mathcal{P}$ with respect to $\mathcal{Q}$ is the expected value of $|X_i \cap Y_j|/|Y_j|$.

Lemma 8.1. *Let $\mathcal{P} = \{X_1, \ldots, X_r\}$ and $\mathcal{Q} = \{Y_1, \ldots, Y_s\}$ be two partitions of a finite set U, and let $\mathcal{Q}'$ be a refinement of $\mathcal{Q}$. Then the mean-square density of $\mathcal{P}$ with respect to $\mathcal{Q}'$ is at least as great as the mean-square density of $\mathcal{P}$ with respect to $\mathcal{Q}$.*

Proof. Let the sets that make up $\mathcal{Q}'$ be called Y_{jk}, where $Y_j = \bigcup_k Y_{jk}$. For each j and k define γ_j and γ_{jk} by $|Y_j| = \gamma_j|U|$ and $|Y_{jk}| = \gamma_{jk}|U|$. For each i, j and k let $d_{ij} = |X_i \cap Y_j|/|Y_j|$ and let $d_{ijk} = |X_i \cap Y_{jk}|/|Y_{jk}|$. Then

$$\sum_k d_{ijk}|Y_{jk}| = \sum_k |X_i \cap Y_{jk}| = |X_i \cap Y_j| = d_{ij}|Y_j| ,$$

from which it follows that $\sum_k \gamma_{jk}d_{ijk} = \gamma_j d_{ij}$ for every i and j.

The mean-square density of $\mathcal{P}$ with respect to $\mathcal{Q}$ is $\sum_i\sum_j \gamma_j d_{ij}^2$, which is therefore equal to

$$\begin{aligned}\sum_i\sum_j \gamma_j^{-1}\Big(\sum_k \gamma_{jk}d_{ijk}\Big)^2 &= \sum_i\sum_j\Big(\sum_k \gamma_j^{-1/2}\gamma_{jk}d_{ijk}\Big)^2\\ &\leqslant \sum_i\sum_j\Big(\sum_k \gamma_j^{-1}\gamma_{jk}\Big)^2\Big(\sum_k \gamma_{jk}d_{ijk}^2\Big) ,\end{aligned}$$

by the Cauchy-Schwarz inequality. Since $\sum_k \gamma_j^{-1}\gamma_{jk} = 1$ for every j, this equals $\sum_i \sum_j \sum_k \gamma_{jk} d_{ijk}^2$, which is the mean-square density of $\mathcal{P}$ with respect to $\mathcal{Q}'$. □

The next lemma is a simple, but somewhat irritating, technicality.

Lemma 8.2. *Let $\epsilon > 0$, let $X_1, \dots, X_r$ be a sequence of finite sets, let $\mathcal{K}(X_1, \dots, X_r)$ be the complete r-partite k-chain with vertex sets $X_1, \dots, X_r$ and for each A of size at most k let $\mathcal{P}(A)$ be a partition of $K(A)$ into n_A sets. For each $x = (x_1, \dots, x_r) \in X_1 \times \dots \times X_r$ and each A of size at most k let $\delta_{A,x}$ be the relative density of the hypergraph $H(A, x)$ in the chain $\mathcal{H}(x)$ (defined in the previous section). Then if $(x_1, \dots, x_r)$ is chosen randomly from $X_1 \times \dots \times X_r$ and $A \subset [r]$ has size at most k, the probability that $\delta_{A,x} < \epsilon n_A^{-1}$ is at most ϵ.*

Proof. Let B and B' be two sets of index A. Let us call them *weakly equivalent*, and write $B \sim_* B'$, if B is C-equivalent to B' for every proper subset C. Then B is strongly equivalent to B' if and only if $B \sim_* B'$ and $B \sim_A B'$.

The relative density $\delta_{A,x}$ is simply the probability that a set B of index A is strongly equivalent to $x(A)$ given that it is weakly equivalent to A. Since $K(A)$ is partitioned into n_A sets, the number of strong equivalence classes in each weak equivalence class is at most n_A. Therefore, for any weak equivalence class T, the probability that $x(A)$ lies in a strong equivalence class of size less than $\epsilon n_A^{-1}|T|$ given that it lies in T is at most ϵ. If $x(A)$ lies in a strong equivalence class of size at least $\epsilon n_A^{-1}|T|$, then the probability that B is in the same strong equivalence class given that B is in T is at least ϵn_A^{-1}, which implies that $\delta_{A,x} \geqslant \epsilon n_A^{-1}$.

Therefore, for every T the conditional probability that $\delta_{A,x} < \epsilon n_A^{-1}$ given that $x(A) \in T$ is less than ϵ. The result follows. □

We now have all the ingredients needed to prove our regularity lemma.

§9. The proof of Theorem 7.3.

It will be convenient for the proof if $\mathcal{J}$ contains the double octahedron $\mathcal{D}$. Since the result for $\mathcal{J}$ follows from the result for $\mathcal{J} \cup \mathcal{D}$, we are free to assume that this is the case.

Let us describe an inductive procedure for producing better and better systems of partitions. Then we shall prove that the procedure terminates.

Suppose, then, that for all subsets $A \subset [r]$ of size at most k we have a partition $\mathcal{P}(A)$ of the hypergraph $K(A)$ into t_A sets. (For ease of notation, we are using the letter $\mathcal{P}$, but these partitions are not the initial partitions in the statement of the theorem.) Suppose, moreover, that if $x = (x_1, \ldots, x_r)$ is chosen randomly from $X_1 \times \ldots \times X_r$, then there is a probability of at least ϵ that the chain $\mathcal{H}(x)$ fails to be $(\epsilon, \mathcal{J}, k)$-quasirandom.

Let γ be defined by $2\gamma \sum_{i=1}^k \binom{r}{i} = \epsilon$. By Lemma 8.2, the probability that there exists a subset $A \subset [r]$ of size at most k such that $\delta_A < \gamma t_A^{-1}$ is at most $\gamma \sum_{i=1}^k \binom{r}{i} = \epsilon/2$. Therefore, with probability at least $\epsilon/2$, the chain $\mathcal{H}(x)$ fails to be $(\epsilon, \mathcal{J}, k)$-quasirandom but for each A the relative density $\delta_{A,x}$ is at least γt_A^{-1}.

In order to continue with the argument, we must use the detailed definition of "$(\epsilon, \mathcal{J}, k)$-quasirandom" that immediately precedes Theorem 5.1. Let $\eta_2, \ldots, \eta_s$ and $\epsilon_2, \ldots, \epsilon_k$ be the sequences that appear in that definition, and note that η_s depends only on ϵ and the densities $\delta_{A,x}$ with $|A| \geqslant s$. Since $\delta_{A,x} \geqslant \gamma t_A^{-1}$ for every A, it follows that η_s is bounded below by a function of ϵ and all those t_A for which $|A| \geqslant s$.

If $\mathcal{H}(x)$ fails to be $(\epsilon, \mathcal{J}, k)$-quasirandom, there must be a minimal s such that it fails to be $(\epsilon_s, \mathcal{J}, s)$-quasirandom, and for that s there must be a set A of size s such that $H(A, x)$ is not η_s-quasirandom relative to $\mathcal{H}(x)$, while $\mathcal{H}(x)$ is $(\epsilon_{s-1}, \mathcal{J}, s-1)$-quasirandom. Since there are at most $\sum_{i=1}^k \binom{r}{i}$ possibilities for this set A we may deduce from the previous paragraph that there exists a set A of size $s \leqslant k$ such that, with probability at least γ, the chain $\mathcal{H}(x)$ is $(\epsilon_{s-1}, \mathcal{J}, s-1)$-quasirandom but $H(A, x)$ is not η_s-quasirandom relative to $\mathcal{H}(x)$.

Let us call x *irregular* if $\mathcal{H}(x)$ has these two properties. Given an irregular x, let $\mathcal{H}_-(A, x)$ be the s-partite $(s-1)$-chain consisting of all strong equivalence classes $H(C, x)$ with $C \subsetneq A$. We can now apply Corollary 6.8 to the chain $\mathcal{H}_-(A, x)$ and to the s-uniform hypergraph $H(A, x)$. (Thus, the k of Corollary 6.8 is equal to s here.) Since $\epsilon_{s-1} \leqslant \epsilon \leqslant |\mathcal{J}|^{-1}$, the conditions hold for the corollary to be applicable. The hypergraphs $H_1, \ldots, H_k$ in the statement of Corollary 6.8 are, in this context, the hypergraphs $H(A', x)$, where A' ranges over all subsets of A of size $s-1$.

For each $C \subsetneq A$ we know that $\delta_{C,x} \geqslant \gamma t_C^{-1}$. Therefore, if r is a positive integer that is at least $\prod_{C \subsetneq A} \gamma^{-1} t_C$, then for each subset $A' \subset A$ of size $s-1$ we can find a partition of $H(A', x)$ into at most 3^r subsets, in such a way that the mean-square density of $H(A, x)$ with respect to the induced partition of $H_*(A, x)$ is at least $\delta_{A,x}^2 + \eta_s^2/8$. (Here, $H_*(A, x)$ denotes the hypergraph consisting of all sets Y of index A such that every proper subset of Y belongs to $\mathcal{H}_-(A, x)$.)

Let $\mathcal{H}(A, x)$ be the s-partite s-chain $H(A, x) \cup \mathcal{H}_-(A, x)$. The number of distinct such chains is at most $\prod_{C \subset A} t_C$. For each one such that x is irregular (if $\mathcal{H}(A, x) = \mathcal{H}(A, y)$

and x is irregular then y is irregular) choose a partition of the hypergraphs $H(A', x)$ as above. In general, it will often happen that $\mathcal{H}(A, x) \neq \mathcal{H}(A, y)$ but $H(A', x) = H(A', y)$, so each hypergraph $H(A', x)$ may be partitioned many times. However, the number of distinct chains $\mathcal{H}(A, x)$ is at most $T_A = \prod_{C \subset A} t_C$, so we can find a common refinement of all the partitions of $H(A', x)$ into at most 3^{rT_A} sets.

For each $A' \subset A$ of size $s - 1$ let $\mathcal{Q}(A')$ be the union of all these common refinements, over all the different sets $H(A', x)$. Then $\mathcal{Q}(A')$ is a partition of $K(A')$ into at most $T_{A'} 3^{rT_A}$ sets, and it refines $\mathcal{P}(A')$. For all other sets A, let $\mathcal{Q}(A) = \mathcal{P}(A)$.

By Lemma 8.1, given any irregular x, the mean-square density of $H(A, x)$ with respect to the partition of $H_*(A, x)$ that is induced by the refined partitions of the hypergraphs $H(A', x)$ is still at least $\delta_{A,x}^2 + \eta_s^2/8$. As for a regular x, Lemma 8.1 tells us that the mean-square density of $H(A, x)$ with respect to the refined partition of $\mathcal{H}(x)_*(A)$ is still at least $\delta_{A,x}^2$.

Let $\sigma_A(\mathcal{P})$ be the mean-square density of the partition $\mathcal{P}(A)$ with respect to the partition of $K(A)$ into weak equivalence classes coming from the partitions $\mathcal{P}(C)$. Let $\sigma_A(\mathcal{Q})$ be the mean-square density of $\mathcal{P}(A) = \mathcal{Q}(A)$ with respect to the partition of $K(A)$ arising from $\mathcal{Q}$ in the same way. By the remark preceding Lemma 8.1, ρ_A is the expectation of $\delta_{A,y}$ over all sequences $y = (y_1, \ldots, y_r)$. Let us write this as $\delta_{A,y}(\mathcal{P})$ since it depends on the system of partitions $\mathcal{P}(C)$. Then we can say that σ_A is the expectation of $\delta_{A,x}(\mathcal{Q})$.

What we have just shown is that if x is irregular, then $\mathbb{E}[\delta_{A,y}(\mathcal{Q}) | y \in H(A, x)]$ is at least $\delta_{A,y}(\mathcal{P})^2 + \eta_s^2/8$, which equals $\mathbb{E}[\delta_{A,y}(\mathcal{P}) | y \in H(A, x)] + \eta_s^2/8$. If x is regular, then this conditional expectation is at least $\delta_{A,y}(\mathcal{P})^2$, or $\mathbb{E}[\delta_{A,y}(\mathcal{P}) | y \in H(A, x)]$. Since the probability that x is irregular is at least γ, this shows that $\mathbb{E}[\delta_{A,y}(\mathcal{Q})] \geqslant \mathbb{E}[\delta_{A,y}(\mathcal{P})] + \gamma\eta_s^2/8$. In other words, $\sigma_A(\mathcal{Q}) \geqslant \sigma_A(\mathcal{P}) + \gamma\eta_s^2/8$.

To summarize: if the conclusion of Theorem 7.3 is not true for the partitions $\mathcal{P}(C)$ then there is a set A of size $s \leqslant k$ and a system of refinements $\mathcal{Q}(C)$ such that $\mathcal{Q}(C) = \mathcal{P}(C)$ except when C is a subset of A of size $s - 1$, such that $\sigma_A(\mathcal{Q}) \geqslant \sigma_A(\mathcal{P}) + \gamma\eta_s^2/8$. For a general C, we have $\sigma_C(\mathcal{Q}) \geqslant \sigma_C(\mathcal{P})$ except if $C \subset A$ and $|C| = s - 1$. This is because if C is any other set, then $\mathcal{Q}(C) = \mathcal{P}(C)$ and all other partitions have either been refined or stayed the same.

To see that we cannot iterate this process of successive refinement for ever, let us see why we are performing an induction over a complicated well-ordered set. The objects in the set will be functions $\zeta : [r]^{(\leqslant k)} \to \mathbb{N}$, where $[r]^{(\leqslant k)}$ stands for the collection of subsets of $[r]$ of size at most k. Let these be well-ordered as follows. First, let $\leqslant$ be an ordering on $[r]^{(\leqslant k)}$ such that $A \leqslant C$ whenever $|A| \geqslant |C|$. (In other words, large sets come first in the ordering.) Then let $\zeta < \zeta'$ if there exists a set $A \in [r]^{(\leqslant k)}$ such that $\zeta(C) = \zeta'(C)$ for every $C \leqslant A$ and $\zeta(A) < \zeta'(A)$. This is clearly a well-ordering: it is a lexicographical ordering with respect to the ordering $\leqslant$ on $[r]^{(\leqslant k)}$.

To each system of partitions $\mathcal{P}(A)$ we associate such a function $\zeta_\mathcal{P}$ as follows. Let t_A and γ be defined as they were earlier in this proof. For each $A \subset [r]$ of size at most k let $d_A = \gamma t_A^{-1}$. Then let $\eta_2, \dots, \eta_k$ be defined as they are in the definition of $(\epsilon, \mathcal{J}, k)$-quasirandomness, except that δ_A is replaced by d_A. (These numbers d_A are lower bounds for the densities associated with a typical chain $\mathcal{H}(x)$.) Finally, if $|A| = s$ then let $\zeta_\mathcal{P}(A)$ be the smallest integer greater than $8(1 - \sigma_A(\mathcal{P}))/\gamma\eta_s^2$.

The number $\zeta_\mathcal{P}(A)$ is chosen because it is an upper bound for the number of increases of $\gamma\eta_s^2/8$ that it is possible to make to $\sigma_A(\mathcal{P})$. Now when we replaced $\mathcal{P}$ by $\mathcal{Q}$ above, we had a number $s \leqslant k$ and a set A of size s such that $\mathcal{P}(C) = \mathcal{Q}(C)$ for every set C that is not a subset of A of size $s - 1$. In particular, the values of d_C for $\mathcal{Q}$ are as they were for $\mathcal{P}$ for every set C of size s or greater. It follows that η_s is the same for $\mathcal{Q}$ as it is for $\mathcal{P}$. Since $\sigma_A(\mathcal{Q}) \geqslant \sigma_A(\mathcal{P}) + \gamma\eta_s^2/8$, we therefore find that $\zeta_\mathcal{Q}(A) < \zeta_\mathcal{P}(A)$. For all $C \leqslant A$ we have $\zeta_\mathcal{Q}(C) = \zeta_\mathcal{P}(C)$, so we have shown that $\zeta_\mathcal{Q} < \zeta_\mathcal{P}$.

This proves that the iterations come to an end. To complete the proof of Theorem 7.3, we observe that at any given stage of the iteration, if we replace partitions $\mathcal{P}(A)$ of size $t_\mathcal{P}(A)$ by partitions $\mathcal{Q}(A)$ of size $t_\mathcal{Q}(A)$, then each number $t_\mathcal{Q}(C)$ is bounded above by a monotone increasing function of the $t_\mathcal{P}(A)$ only. So as the iteration proceeds we have a sequence of pairs of functions $(\zeta_\mathcal{P}, t_\mathcal{P})$ such that the $\zeta_\mathcal{P}$ form a decreasing sequence and such that if $\mathcal{Q}$ follows $\mathcal{P}$ then every $\zeta_\mathcal{Q}(C)$ and $t_\mathcal{Q}(C)$ is bounded above by a function of $(\zeta_\mathcal{P}, t_\mathcal{P})$ that is monotone increasing in each $t_\mathcal{P}(A)$. Therefore, the values of $t_\mathcal{Q}(C)$ when $\mathcal{Q}$ is the final system of partitions in the sequence are bounded above by a function that depends only on the values of the initial $t_\mathcal{P}$. $\square$

§10. Hypergraphs with few simplices.

Now that we have established counting and regularity lemmas we have the tools necessary to prove the generalization of Theorems 1.3 and 1.6 to k-uniform hypergraphs.

Theorem 10.1. *Let k be a positive integer. Then for every $a > 0$ there exists $c > 0$ with the following property. Let H be a $(k+1)$-partite k-uniform hypergraph with vertex sets $X_1, \dots, X_{k+1}$, and let N_i be the size of X_i. Suppose that H contains at most $c\prod_{i=1}^{k+1} N_i$ simplices. Then for each $i \leqslant k+1$ one can remove at most $a\prod_{j \neq i} N_j$ edges of H from*

$\prod_{j\neq i} X_j$ *in such a way that after the removals one is left with a hypergraph that is simplex-free.*

Proof. For each subset $A \subset [k+1]$ of size at most k, define a partition $\mathcal{P}(A)$ of $K(A)$ as follows. If $|A| < k$ then $\mathcal{P}(A)$ consists of the single set $K(A)$. If $|A| = k$ then it consists of the sets $H(A)$ and $K(A) \setminus H(A)$. Now apply Theorem 7.3 to this system of partitions, with $\mathcal{J} = [k+1]^{(\leqslant k)}$ and $\epsilon = \min\{|\mathcal{J}|^{-1}/2, a/2\}$, obtaining for each $A \in \mathcal{J}$ a partition $\mathcal{Q}(A)$ of $K(A)$ into m_A sets.

If $x = (x_1, \ldots, x_{k+1}) \in X_1 \times \ldots \times X_{k+1}$ and $\mathcal{H}(x)$ is not $(\epsilon, \mathcal{J}, k)$-quasirandom, then there must be some A of size $s \leqslant k$ such that $H(A, x)$ is not η_s-quasirandom relative to $\mathcal{H}(x)$. There must be some i such that $i \notin A$ and if $(y_1, \ldots, y_{k+1})$ is another sequence such that $y_j = x_j$ when $j \neq i$, then $H(A, y)$ will also not be η_s-quasirandom relative to $\mathcal{H}(y)$. Therefore, since $\mathcal{H}(x)$ is $(\epsilon, \mathcal{J}, k)$-quasirandom with probability at least $1 - \epsilon$, there are at most $\epsilon \prod_{j\neq i} N_j$ elements of $\prod_{j\neq i} X_j$ that can be extended to sequences x that are not $(\epsilon, \mathcal{J}, k)$-quasirandom. Remove from H any such element.

Let γ be defined by $\gamma \sum_{i=1}^{k} \binom{k+1}{i} = a/2$. Then Lemma 8.2 tells us that if $x = (x_1, \ldots, x_{k+1})$ is chosen randomly, then with probability at least $1 - a/2$, we have $\delta_{A,x} \geqslant \gamma m_A^{-1}$ for every $A \in [k+1]^{(\leqslant k)}$. Again, the event that this happens for a particular A does not depend on the x_i with $i \notin A$. So for each i there are at most $a \prod_{j\neq i} N_j/2$ elements of $\prod_{j\neq i} X_j$ that can be extended to sequences x for which $\delta_{A,x} < \gamma m_A^{-1}$ for some $A \subset [k+1]$ with $i \notin A$. Once again, remove all such elements from H.

For each i we have removed at most $a \prod_{j\neq i} N_j$ from $H \cap \prod_{j\neq i} X_j$. It remains to show that in the process we have either removed all simplices from H, or else, for some $c > 0$ that depends on a only, there were at least $c \prod_j N_j$ simplices to start with.

Suppose, then, that after the removals there is still a simplex $x = (x_1, \ldots, x_{k+1})$, and consider the chain $\mathcal{H}(x)$. Then for every $A \subset [k+1]$ of size k the following statements are true. First, the set $x(A)$ is an element of H (or else x would not be a simplex). Second, the hypergraph $H(A, x)$ is a subset of H (since $x(A) \in H$ and the partition into strong equivalence classes resulting from $\mathcal{Q}$ refines the partition $\mathcal{P}$). Third, $\delta_{C,x} \geqslant \gamma m_C^{-1}$ for every $C \subset A$ (or else we would have removed $x(A)$ from H). Finally, the chain $\mathcal{H}(x)$ is $(\epsilon, \mathcal{J}, k)$-quasirandom (or else for some A of size k we would have removed $x(A)$ from H).

We now apply Corollary 5.2, the counting lemma for quasirandom chains. It implies that the number of simplices in the chain $\mathcal{H}(x)$, which is the same as the number of homomorphisms from $\mathcal{J}$ to $\mathcal{H}(x)$, is at least $\prod_j N_j \prod_{A\in\mathcal{J}} \delta_{A,x}$, which is at least $\prod_j N_j \prod_{A\in\mathcal{J}} \gamma m_A^{-1}$. But γ and the m_A depend on a and k only, so the result is proved. $\square$

Finally, let us deduce from this a multidimensional Szemerédi theorem.

Theorem 10.2. *Let $\delta > 0$ and $k \in \mathbb{N}$. Then, if N is sufficiently large, every subset A of the k-dimensional grid $\{1, 2, \ldots, N\}^k$ of size at least δN^k contains a set of points of the form $\{a\} \cup \{a + de_i : 1 \leqslant i \leqslant k\}$, where $e_1, \ldots, e_k$ is the standard basis of $\mathbb{R}^k$ and d is a non-zero integer.*

Proof. Suppose that A is a subset of $\{1, 2, \ldots, N\}^k$ of size δN^k, and that A contains no configuration of the kind claimed. Define a $(k+1)$-partite k-graph F_k with vertex sets $X_1, \ldots, X_{k+1}$ as follows. If $j \leqslant k$ then the elements of X_j are hyperplanes of the form $P_{j,m} = \{(x_1, \ldots, x_k) : x_j = m\}$ for some integer $m \in \{1, 2, \ldots, N\}$. If $j = k+1$ then they are hyperplanes of the form $Q_m = \{(x_1, \ldots, x_k) : x_1 + \ldots + x_k = m\}$ where m is an integer between k and kN. The edges of F_k are sets of k hyperplanes from different sets V_j that intersect in a point of A.

If F_k contains a simplex with vertices P_{j,m_j} and Q_m, then the points $(m_1, \ldots, m_k)$ and $(m_1, \ldots, m_k) + (m - \sum_{i=1}^k m_i)e_j$ all belong to A. This gives us a configuration of the desired kind except in the degenerate case where $m = \sum_{i=1}^k m_i$, which is the case where all $k+1$ hyperplanes have a common intersection. By our assumption on A, all the simplices in F_k are therefore degenerate ones of this kind, which implies that there are at most δN^k of them.

Now $|X_i| = N$ if $i \leqslant k$ and $|X_{k+1}| = kN$. We can therefore apply (the contrapositive of) Theorem 10.1 with $c = N^{-1}k^{-1}$. If N is sufficiently large, then the resulting a is smaller than $\delta/2k$, which implies that we can remove fewer than δN^k edges from the hypergraph F_k and thereby remove all simplices. However, every edge of a degenerate simplex determines the point of intersection of the $k+1$ hyperplanes and hence the simplex itself. It follows that one must remove at least δN^k edges to get rid of all simplices. This contradiction proves the theorem. $\square$

The above result is a special case of the multidimensional Szemerédi theorem, but it is in fact equivalent to the whole theorem. This is a well-known observation. We give a (slightly sketchy) proof below.

Theorem 10.3. *For every $\delta > 0$, every positive integer r and every finite subset $X \subset \mathbb{Z}^r$ there is a positive integer N such that every subset A of the grid $\{1, 2, \ldots, N\}^r$ of size at least δN^r has a subset of the form $a + dX$ for some positive integer d.*

Proof. It is clearly enough to prove the result for sets X such that $X = -X$, so all we actually need to ensure is that $d \neq 0$. A simple averaging argument shows that we may also assume that X is not contained in any $(r-1)$-dimensional subspace of $\mathbb{R}^r$. Let the cardinality of X be $k+1$. Let ϕ be an affine map that defines a bijection from the set $\{0, e_1, \dots, e_k\} \subset \mathbb{R}^k$ to X, regarded as a subset of $\mathbb{R}^r$. Another simple averaging argument allows us to find a grid $\{1, 2, \dots, M\}^k$, where M tends to infinity with N, as well as a point $z \in \mathbb{Z}^r$ and a constant $\eta > 0$ depending on δ and X only, such that $z + \phi(x) \in A$ for at least ηM^k points in $\{1, 2, \dots, M\}^k$. Let B be the set of points with this property. Thus, B has density at least η and Theorem 10.2 shows that B contains a set of the form $w + c\{0, e_1, \dots, e_k\}$. But then $z + \phi(w + c\{0, e_1, \dots, e_k\})$ is a set of the form $a + dX$ and is also a subset of A. $\square$

Concluding Remarks.

I am very grateful indeed to Yoshiyasu Ishigami, who read carefully an earlier version of this paper and found an error which, though it did not invalidate the approach, occurred early in the argument and therefore required more than local changes to it. While thinking about how to deal with it, I discovered a way of substantially simplifying the proof of the counting lemma, and in the end it seemed best, even if depressing, to rewrite the whole paper (including the regularity part) from scratch.

Recently, Tao [T] has given another proof of the main result of this paper (Theorem 10.1), and indeed of a slight generalization. He also proves regularity and counting lemmas, by methods more closely related to those of Nagle, Rödl and Schacht, but with some new ideas and a different language that leads to considerably shorter proofs.

References.

[FK] H. Furstenberg, Y. Katznelson, *An ergodic Szemerédi theorem for commuting transformations*, J. Analyse Math. **34** (1978), 275-291.

[FKO] H. Furstenberg, Y. Katznelson, D. Ornstein, *The ergodic theoretical proof of Szemerédi's theorem*, Bull. Amer. Math. Soc. **7** (1982), 527–552.

[G1] W. T. Gowers, *A new proof of Szemerédi's theorem*, Geom. Funct. Anal. **11** (2001), 465-588.

[G2] W. T. Gowers, *Quasirandomness, Counting and Regularity for 3-Uniform Hypergraphs*, submitted.

[NRS] B. Nagle, V. Rödl and M. Schacht, *The counting lemma for regular k-uniform hypergraphs*, Random Structures and Algorithms (to appear).

[R] V. Rödl, *Some developments in Ramsey theory*, Proceedings of the International Congress of Mathematicians, Vol. I, II (Kyoto 1990), 1455-1466, Math. Soc. Japan, Tokyo, 1991.

[RS] I. Z. Ruzsa and E. Szemerédi, *Triple systems with no six points carrying three triangles*, Combinatorics (Proc. Fifth Hungarian Colloq., Keszthely, 1976), Vol. II, 939-945.

[S1] J. Solymosi, *Note on a generalization of Roth's theorem*, Discrete and Computational Geometry, 825-827, Algorithms Combin. **25**, Springer, Berlin 2003.

[S2] J. Solymosi, *A note on a question of Erdős and Graham*, Combin. Probab. Comput. **13** (2004), 263-267.

[Sz1] E. Szemerédi, *Integer sets containing no k elements in arithmetic progression*, Acta Arith. **27** (1975), 299-345.

[Sz2] E. Szemerédi, *Regular partitions of graphs*, in Problèmes Combinatoires et Théorie des Graphes, Proc. Colloque Inter. CNRS, (Bermond, Fournier, Las Vergnas, Sotteau, eds.), CNRS Paris, 1978, 399-401.

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