Szemerédi's Lemma for the Analyst

László Lovász and Balázs Szegedy
Microsoft Research

April 2005

Microsoft Research Technical Report # MSR-TR-2005-90

Abstract

Szemerédi's Regularity Lemma is a fundamental tool in graph theory: it has many applications to extremal graph theory, graph property testing, combinatorial number theory, etc. The goal of this paper is to point out that Szemerédi's Lemma can be thought of as a result in analysis, and show some applications of analytic nature.

1 Introduction

Szemerédi's Regularity Lemma was first used in his celebrated proof of the Erdős–Turán Conjecture on arithmetic progressions in dense sets of integers. Since then, the Lemma has emerged as a fundamental tool in graph theory: it has many applications in extremal graph theory, in the area called "Property Testing", combinatorial number theory, etc.

Roughly speaking, the Szemerédi Regularity Lemma says that the node set of every (large) graph can be partitioned into a small number of parts so that the subgraphs between the parts are "random-like". There are several ways to make this precise, some equivalent to the original version, some not (see Section 2 for an exact statement).

The goal of this paper is not to describe these applications (see [13, 12] for surveys and discussions of such applications); nor will we discuss extensions to sparse graphs by [8], or hypergraphs by Frankl–Rödl, Gowers and Tao [6, 11, 16]. Tao [17] describes the lemma as a result in probability. Our goal is to point out that Szemerédi's Lemma can be thought of as a result in analysis, and show some applications of analytic nature.

2 Weak and Strong Regularity Lemma

We start with stating a version of the Lemma that will be most suited for our discussion. Consider a bipartite graph $G = (V, E)$, with bipartition $\{U, W\}$. This means that V is the set of nodes of G, which is partitioned into two classes U and W so that every edge connects a node in U to a node in W. The ratio $d = \frac{|E|}{|U|\cdot|V|}$ can be thought of as the *edge density* of the bipartite graph. For every $X \subseteq U$ and $Y \subseteq W$, let $e_G(X, Y)$ denote the number of edges with one endnode in X and another in Y. On the average, we expect that

$$e_G(X, Y) \approx d|X| \cdot |Y|.$$

For a general graph, $e_G(X,Y)$ may be very far from this "expected value". If G is a random graph, however, it will be close; random graphs are very "regular" in this respect. So we can measure how non-random G is by its *irregularity*

$$\mathrm{irreg}(G) = \max_{X\subseteq U, Y\subseteq V} \big|e_G(X,Y) - d|X|\cdot|Y|\big|.$$

Clearly $\mathrm{irreg}(G) \le |X|\cdot|Y|$.

With this notation, we can state one version of the Regularity Lemma:

Lemma 1 (Szemerédi Regularity Lemma) *For every $\varepsilon > 0$ there is a $k(\varepsilon) > 0$ such that every graph $G = (V,E)$ has a partition $\mathcal{P}$ into $k \le k(\varepsilon)$ classes $V_1,\dots,V_k$ such that*

$$\sum_{1\le i<j\le k} \mathrm{irreg}(G[V_i,V_j]) \le \varepsilon|V|^2.$$

While this form is perhaps easiest to state and prove, there are equivalent forms that are more suited for applications. We say that a bipartite graph G with bipartition $\{U,W\}$ and edge-density d is *ε-regular*, if

$$d - \varepsilon < \frac{e_G(X,Y)}{|X|\cdot|Y|} < d + \varepsilon$$

holds for all subsets $X\subseteq U$ and $Y\subseteq W$ such that $|X| > \varepsilon|U|$ and $|Y| > \varepsilon|W|$. (So every induced subgraph with sufficiently many nodes in both classes has about the same edge-density.) Notice that we could not require the condition to hold for small X and Y: for example, if both have one element, then the quotient $e_G(X,Y)/(|X|\cdot|Y|)$ is either 0 or 1.

We say that a partition $V_1,\dots,V_k$ is into *almost equal parts* if $\big||V_i| - |V_j|\big| \le 1$ for all $1 \le i < j \le k$. Clearly, this means that $\lfloor |V|/k\rfloor \le |V_i| \le \lceil |V|/k\rceil$ for every $1\le i\le k$.

With these definitions, the Regularity Lemma can be stated as follows:

Lemma 2 (Szemerédi Regularity Lemma, second form) *For every $\varepsilon > 0$ there is a $k = k(\varepsilon)$ such that every graph $G = (V,E)$ has a partition into $k \le k(\varepsilon)$ almost equal parts $V_1,\dots,V_k$ such that for all but εk^2 pairs of indices $1 \le i < j \le k$, the bipartite graph $G[V_i,V_j]$ is ε-regular.*

The equivalence of the two forms is easy to prove. One can add further requirements (at the cost of increasing $k(\varepsilon)$), like a lower bound on the number of partition classes, or the requirement that $\{V_1,\dots,V_k\}$ refines a given partition.

One feature of the Regularity Lemma, which unfortunately forbids practical applications, is that $k(\varepsilon)$ is very large: the best proof gives a tower of height about $1/\varepsilon^2$, and unfortunately this is not far from the truth, as it was shown by Gowers [10].

A related result with a more reasonable threshold was proved by Frieze and Kannan [7], but they measure irregularity in a different way. For a partition $\mathcal{P} = \{V_1,\dots,V_k\}$ of V, define $d_{ij} = \frac{e_G(V_i,V_j)}{|V_i|\cdot|V_j|}$. For any two sets $S,T \subseteq V(G)$, we expect that the number of edges of G connecting S to T is about

$$e_{\mathcal{P}}(S,T) = \sum_{i=1}^{l}\sum_{j=1}^{k} d_{ij}|V_i\cap S|\cdot|V_j\cap T|.$$

So we can measure the irregularity of the partition by $\max_{S,T} |e_G(S,T) - e_{\mathcal{P}}(S,T)|$. The Weak Regularity Lemma [7] says the following.

Lemma 3 (Weak Regularity Lemma) *For every $\varepsilon > 0$ and every graph $G = (V,E)$, V has a partition $\mathcal{P}$ into $k \le 2^{1/\varepsilon^2}$ classes $V_1, \dots, V_k$ such that for all $S, T \subseteq V$,*

$$|e_G(S,T) - e_{\mathcal{P}}(S,T)| \le \varepsilon |V|^2.$$

The partition in the weak lemma has substantially weaker properties than the partition in the strong lemma; these properties are sufficient in some, but not all, applications. The bound on the number of partition classes is still rather large (exponential), but at least not a tower. We'll see that the proof obtains the partition as an "overlay" of only $1/\varepsilon^2$ sets, which in some applications can be treated as if there were only about $1/\varepsilon^2$ classes, which makes the weak lemma quite efficient (see e.g. its applications in [1]).

Other versions of the Regularity strengthen, rather than weaken, the conclusion (of course, at the cost of replacing the tower function by an even more huge value). Such a "super-strong" Regularity Lemma was proved and used by Alon and Shapira [3]. A probabilistic and information theoretic version was given by Tao [17].

3 A lemma about Hilbert space

Lemma 4 *Let $\mathcal{K}_0, \mathcal{K}_1, \dots$ be arbitrary nonempty subsets of a Hilbert space $\mathcal{H}$. Then for every $\varepsilon > 0$ and $f \in \mathcal{H}$ there is an $m \le \lceil 1/\varepsilon^2 \rceil$ and there are $f_i \in \mathcal{K}_i$ ($0 \le i \le m$) and $\gamma_0, \gamma_1, \dots, \gamma_m \in \mathbb{R}$ such that for every $g \in \mathcal{K}_{m+1}$*

$$\langle g, f - (\gamma_1 f_1 + \dots + \gamma_m f_m)\rangle \le \varepsilon \|g\| \cdot \|f\|.$$

Before proving this lemma, a little discussion is in order. Assume that the sets $\mathcal{K}_n$ are closed subspaces. Then a natural choice for the function $\gamma_1 f_1 + \dots + \gamma_m f_m$ is the best approximation of f in the subspace $\mathcal{K}_1 + \dots \mathcal{K}_m$, and the error $f - (\gamma_1 f_1 + \dots + \gamma_m f_m)$ is orthogonal to every $g \in \mathcal{K}_1 + \dots \mathcal{K}_m$. The main point in this lemma is that it is also almost orthogonal to the *next* set K_{m+1}.

Proof. Let

$$\eta_k = \inf_{\gamma, \{f_i\}} \|f - \sum_{i=1}^k \gamma_i f_i\|^2,$$

where the infimum is taken over all $\gamma_1, \dots, \gamma_k \in \mathbb{R}$ and $f_i \in \mathcal{K}_i$. Clearly we have $\|f\|^2 \ge \eta_1 \ge \eta_2 \ge \dots \ge 0$. Hence there is an $m \le \lceil 1/\varepsilon^2 \rceil$ such that $\eta_m < \eta_{m+1} + \varepsilon^2 \|f\|^2$. So there are $\gamma_1, \dots, \gamma_m \in \mathbb{R}$ and $f_i \in \mathcal{K}_i$ such that

$$\|f - \sum_{i=1}^m \gamma_i f_i\|^2 \le \eta_{m+1} + \varepsilon^2 \|f\|^2.$$

Let $f^* = \sum_i \gamma_i f_i$, and consider any $g \in \mathcal{K}_{m+1}$. By the definition of η_{m+1}, we have for every real t that

$$\|f - (f^* + tg)\|^2 \ge \eta_{m+1} \ge \|f - f^*\|^2 - \varepsilon^2 \|f\|^2,$$

or

$$\|g\|^2t^2 - 2\langle g, f - f^*\rangle t + \varepsilon^2\|f\|^2 \geq 0.$$

The discriminant of this quadratic polynomial must be nonpositive, which proves the lemma. □

To formulate some consequences of this Lemma, we need some definitions. We call a function $f \in \mathcal{W}$ *a stepfunction with at most m steps* if there is a partition $\{S_1, \dots, S_m\}$ of $[0,1]$ such that f is constant on every $S_i \times S_j$. For every $W \in \mathcal{W}$ and every partition $\mathcal{P} = \{P_1, \dots, P_k\}$ of $[0,1]$ into measurable sets, let $W_{\mathcal{P}} : [0,1]^2 \to \mathbb{R}$ denote the stepfunction obtained from W by replacing its value at $(x,y) \in P_i \times P_j$ by the average of W over $P_i \times P_j$. (This is not defined when $\lambda(P_i) = 0$ or $\lambda(P_j) = 0$, but this is of measure 0.)

First, let us apply this lemma to the case when the Hilbert space is $L^2([0,1]^2)$, and each $\mathcal{K}_n$ is the set of indicator functions of product sets $S \times S$, where S is a measurable subset of $[0,1]$. Let $f \in \mathcal{W}_0$, then $f^* = \sum_{i=1}^k \gamma_i f_i$ is a stepfunction with at most 2^k steps, and so we get

Corollary 5 *For every function $W \in \mathcal{W}_0$ and $\varepsilon > 0$ there is a stepfunction $W^* \in \mathcal{W}$ with at most $2^{\lceil 2/\varepsilon^2 \rceil}$ steps such that for every measurable set $S \subseteq [0,1]$,*

$$\left| \int_{S \times S} (W - W^*) \right| \leq \varepsilon.$$

It is easy to see that we can choose the stepfunction W^* in Corollary 5 to be of the form $W_{\mathcal{P}}$, where $\mathcal{P}$ is a partition of $[0,1]$ with at most $\lceil 1/\varepsilon^2 \rceil$ classes. In particular, we can choose $W^* \in \mathcal{W}_0$.

It is also easy to see that the conclusion implies that for any two measurable sets $S, T \subseteq [0,1]$,

$$\left| \int_{S \times T} (W - W^*) \right| \leq 4\varepsilon,$$

which is often a more convenient form of the conclusion.

To derive the weak form of Szemerédi's Lemma from this Corollary, we represent the graph G on n nodes by a stepfunction W_G: we consider the adjacency matrix $A = (a_{ij})$ of G, and replace each entry a_{ij} by a square of size $(1/n) \times (1/n)$ with the constant function a_{ij} on this square. Let $\mathcal{A}$ be the algebra of subsets of $[0,1]$ generated by the intervals corresponding to nodes of G. We let $\mathcal{K}_n$ be the set of indicator functions of product sets $S \times S$ ($S \in \mathcal{A}$). Analogously to Corollary 5, we get a partition $\mathcal{P} = \{S_1, \dots, S_m\}$ of $[0,1]$ into sets in $\mathcal{A}$ such that

$$\left| \int_{S \times T} (W_G(x,y) - (W_G)_{\mathcal{P}}(x,y))\, dx\, dy \right| \leq 4\varepsilon$$

for all sets $S, T \in \mathcal{A}$. This translates into the conclusion of Lemma 3.

To get the strong form of the Regularity Lemma 1, let us define a sequence $s(1), s(2), \dots$ by $s(1) = 1$ and $s(k+1) = 2^{s(1)^2 \cdots s(k)^2}$. Let us apply Lemma 4 to the same Hilbert space, but choose $\mathcal{K}_n$ to be the set of stepfunctions with steps in $\mathcal{A}$ and with at most $s(n)$ steps. As before, take $f = W_G$. The Lemma 1 gives us a function $f^* = \sum_{i=1}^k \gamma_i f_i$, which is a stepfunction with at most $m = s(1)s(2)\dots s(k)$ steps; let $S_1, \dots S_m$ be these steps. If we choose the function g by selecting a rectangle R_{ij} from each $S_i \times S_j$, and letting its value to be ± 1 on each R_{ij} and 0 otherwise, then we get

$$\sum_{i,j=1}^{m} \left| \int_{R_{ij}} (W(x,y) - W^*(x,y))\, dx\, dy \right| \le \varepsilon.$$

Hence we get the strong form of Szemerédi's Lemma.

We say that a partition $\mathcal{P}$ of $[0,1]$ is a *weak Szemerédi partition* for W with error ε, if

$$\left| \int_{S\times S} (W - W_{\mathcal{P}}) \right| \le \varepsilon$$

holds for every subset $S \subseteq [0,1]$. So every function has a weak Szemerédi partition with error ε, with at most $2^{1/\varepsilon^2}$ classes. We define *strong Szemerédi partitions* analogously, but we'll not have much to say about them in this note.

There may be further interesting choices of the subsets $\mathcal{K}_n$ other than stepfuctions. For example, we can take $\mathcal{H}$ to be $L_2[0,1]$, and $\mathcal{K}_n$, the set of polynomials of degree at most 2^n. Then we get:

Corollary 6 *For every $\varepsilon > 0$ and every function $f \in L_2[0,1]$ there is a polynomial $p \in \mathcal{R}[x]$ of degree $d \le 2^{\lceil 1/\varepsilon^2 \rceil}$ such that*

$$\langle g, f - p \rangle \le \varepsilon \|f\| \cdot \|g\|$$

for every polynomial g of degree at most $2d$.

4 Distances of 2-variable functions

Let $\mathcal{W}$ denote the set of all bounded symmetric measurable functions $W:\ [0,1]^2 \to \mathbb{R}$. Consider the following norm on $\mathcal{W}$:

$$\|W\|_\square = \sup_{S,T\subseteq[0,1]} \left| \int\limits_{S\times T} W(x,y)\, dx\, dy \right|.$$

For the case of matrices, this norm is called the "cut norm"; various important properties of it were proved by Alon and Naor [2] and by Alon, Fernandez de la Vega, Kannan and Karpinski [1]. Many of these extend to the infinite case without any change. In particular, $\|W\|_\square$ is within absolute constant factors to the $L_1 \to L_\infty$ norm of W as a kernel operator. For our purposes, the following inequality from [5] will be needed. For every $W \in \mathcal{W}$, let

$$\|W\|_2 = \left(\int_0^1 \int_0^1 W(x,y)^2\, dx\, dy \right)^{1/2}$$

denote its L_2-norm as a function, and let $W \circ W$ denote its square as a kernel operator, i.e.,

$$(W \circ W)(x,y) = \int_0^1 W(x,t) W(t,y)\, dt.$$

Then we have

$$\|W\|_\square^4 \le \|W \circ W\|_2^2 \le \|W\|_\square \|W\|_\infty^2 \|W\|_1. \tag{1}$$

The Weak Regularity Lemma in these terms asserts the following:

Lemma 7 (Weak Regularity Lemma, Analytic Form) *For every function $W \in \mathcal{W}_0$ and $\varepsilon > 0$ there is a stepfunction $W' \in \mathcal{W}_0$ with at most $\lceil 2^{1/\varepsilon^2} \rceil$ steps such that $\|W - W'\|_\square \le \varepsilon$.*

It would be very interesting to obtain an analogous formulation for the strong Regularity Lemma.

5 A compactness theorem

For two functions $U \in W$ and $\phi : [0,1] \to [0,1]$, define $U^\phi(x,y) = U(\phi(x), \phi(y))$. Define a "distance" of two functions $U, W \in \mathcal{W}$ by

$$\delta_\square(U,W) = \inf_{\phi,\psi} \|U^\phi - W^\psi\|_\square,$$

where ϕ and ψ range over all measure preserving maps $[0,1] \to [0,1]$.

It is not hard to check that the triangle inequality is satisfied. The $\delta_\square$ distance of two different functions can be 0; various characterizations of when this distance is 0 are given in [4]. We construct a metric space $\mathcal{X}$ from $(\mathcal{W}, \delta_\square)$ by identifying such pairs of functions. Let $\mathcal{X}_0$ denote the image of $\mathcal{W}_0$ under this identification.

The following fact can be proved by the methods in [14]; here we sketch the proof, to show the application of Szemerédi's Lemma.

Lemma 8 *The metric space $\mathcal{X}_0$ is compact.*

Proof. Let $W_1, W_2, \dots$ be a sequence of functions in $\mathcal{W}_0$. We want to construct a subsequence that has a limit in $\mathcal{X}_0$.

For each k and n, we find a stepfunction $W_{n,k} \in \mathcal{W}_0$ with $\lceil 2^{k^2} \rceil$ steps such that

$$\|W_n - W_{n,k}\|_\square \le 1/k.$$

We can assume that the steps of $W_{n,k+1}$ are subsets of the steps of $W_{n,k}$; otherwise, we consider the common refinement of the partitions for $W_{n,1}, W_{n,2}, \dots, W_{n,k}$ as the steps of $W_{n,k}$; this means that $W_{n,k}$ will have at most $m_k = 2^{k^3}$ steps; we may assume it has exactly this number of steps.

We'll only use that $\delta_\square(W_n, W_{n,k}) \le 1/k$, which means that we can rearrange the range of $W_{n,k}$ as we wish; in particular, we may assume that all steps are intervals.

Now we can select a subsequence of the W_n for which the length of the i-th interval of $W_{n,1}$ converges for every i, and also the value of $W_n^{(1)}$ on the product of the i-th and j-th intervals converges for every i and j (as $n \to \infty$). It follows then that the sequence $W_{n,1}$ converges to a limit U_1 almost everywhere, which itself is a stepfunction with m_1 steps that are intervals.

We repeat this for $k = 2, 3, \dots$, to get subsequences for which $W_{k,n} \to U_k$ almost everywhere, where U_k is a stepfunction with m_k steps that are intervals.

For every $k < l$, the partition into the steps of $W_{n,l}$ is a refinement of the partition into the steps of $W_{n,k}$, and hence it is easy to see that the same relation holds for the partitions into the steps of U_l and U_k. Furthermore, the function $W_{n,k}$ can be obtained from the function

$W_{n,l}$ by averaging its value over each step, and it follows that a similar relation holds for U_l and U_k. It follows by the Martingale Theorem that the sequence $(U_1, U_2, \dots)$ is convergent almost everywhere. Let U be its limit.

Fix any $\varepsilon > 0$. Then there is a $k > 3/\varepsilon$ such that $\|U - U_k\|_1 < \varepsilon/3$. Fixing this k, there is an n_0 such that $\|U_k - W_{n,k}\|_1 < \varepsilon/3$ for all $n \geq n_0$. Then

$$\begin{aligned}\delta_\square(U, W_n) &\leq \delta_\square(U, U_k) + \delta_\square(U_k, W_{n,k}) + \delta_\square(W_{n,k}, W_n) \\ &\leq \|U - U_k\|_1 + \|U_k - W_{n,k}\|_1 + \delta_\square(W_{n,k}, W_n) \\ &\leq \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon.\end{aligned}$$

This proves that $W_n \to U$ in the metric space $\mathcal{X}_0$. □

6 The Regularity Lemma and covering by small balls

Every function $W \in \mathcal{W}$ gives rise to a metric on $[0, 1]$ by

$$d^1_W(x_1, x_2) = \|W(x_1, .) - W(x_2, .)\|_2 = \left(\int_0^1 (W(x_1, y) - W(x_2, y))^2 \, dy\right)^{1/2}.$$

It turns out that for our purposes, the following distance function is more important: we square W as a kernel operator, and then consider the above distance. More precisely, we define

$$\begin{aligned}d_W(x_1, x_2) &= d^1_{W\circ W}(x_1, x_2) \\ &= \left(\int_0^1 \Big(\int_0^1 W(x_1, y)W(y, z)\, dy - \int_0^1 W(x_2, y)W(y, z)\, dy\Big)^2 dz\right)^{1/2}.\end{aligned}$$

Using this metric space, we one can prove the following characterization of weak Szemerédi partitions (the proof is given in the Appendix).

Theorem 9 *Let $W \in \mathcal{W}_0$ and let $\mathcal{P} = \{P_1, \dots, P_k\}$ be a partition of $[0, 1]$ into measurable sets.*

(a) *If $\mathcal{P}$ is a weak Szemerédi partition with error $\varepsilon^3/4$, then there is a set $S \subseteq [0, 1]$ with $\lambda(S) \leq \varepsilon$ such that for each partition class, $P_i \setminus S$ has diameter at most ε in the d_W metric.*

(b) *If there is a set $S \subseteq [0, 1]$ with $\lambda(S) \leq \varepsilon^2$ such that for each partition class, $P_i \setminus S$ has diameter at most ε^2 in the d_W metric, then $\mathcal{P}$ is a weak Szemerédi partition with error ε.*

Combining this fact with the existence of weak Szemerédi partitions, we get the following:

Corollary 10 *For every function $W \in \mathcal{W}$ and every $\varepsilon > 0$ there is a partition $\mathcal{P} = \{P_1, \dots, P_k\}$ of $[0, 1]$ into $k \leq 2^{16/\varepsilon^6}$ measurable sets and a set $S \subseteq [0, 1]$ with $\lambda(S) \leq \varepsilon$ such that for each partition class, $P_i \setminus S$ has diameter at most ε in the d_W metric.*

References

[1] N. Alon, W. Fernandez de la Vega, R. Kannan and M. Karpinski: Random sampling and approximation of MAX-CSPs, *J. Comput. System Sci.* **67** (2003), 212–243.

[2] N. Alon, A. Naor: Approximating the Cut-Norm via Grothendieck's Inequality (preprint)

`http://research.microsoft.com/research/theory/naor/homepage%20files/cutnorm.pdf`

[3] N. Alon and A. Shapira, Every monotone graph property is testable, Proc. of the 37 ACM STOC, Baltimore, ACM Press (2005), to appear.

`http://www.math.tau.ac.il/~nogaa/PDFS/MonotoneSTOC.pdf`

[4] C. Borgs, J. Chayes, L. Lovász: Unique limits of dense graph sequences (manuscript)

[5] C. Borgs, J. Chayes, L. Lovász, V.T. Sós, K. Vesztergombi: Convergent sequences of dense graphs

[6] P. Frankl,V. Rödl: Extremal problems on set systems, *Random Structures and Algorithms* **20** (2002), 131–164.

[7] A. Frieze and R. Kannan: Quick approximation to matrices and applications, *Combinatorica* **19**, 175–220.

[8] S. Gerke, A. Steger: The sparse regularity lemma and its applications (manuscript)

`http://www.ti.inf.ethz.ch/as/publications/szem-survey.ps`

[9] O. Goldreich, S. Goldwasser and D. Ron: Property testing and its connection to learning and approximation, *J. ACM* **45** (1998), 653–750.

[10] W.T. Gowers: Lower bounds of tower type for Szemerédi's Uniformity Lemma, *Geom. Func. Anal.* **7** (1997), 322–337.

[11] W.T. Gowers: Hypergraph regularity and the multidimensional Szemerédi theorem (manusript)

[12] Y. Kohayakawa, V. Rödl: Szemerédi's regularity lemma and quasi-randomness, in: *Recent Advances in Algorithms and Combinatorics,* CMS Books Math./Ouvrages Math. SMC **11**, Springer, New York (2003), 289–351.

[13] J. Komlós and M. Simonovits: Szemerédi's Regularity Lemma and its applications in graph theory, in: *Combinatorics, Paul Erdos is Eighty* (D. Miklos et. al, eds.), Bolyai Society Mathematical Studies **2** (1996), pp. 295–352.

[14] L. Lovász and B. Szegedy: Limits of dense graph sequences (MSR Tech Report)

`ftp://ftp.research.microsoft.com/pub/tr/TR-2004-79.pdf`

[15] V. Rödl, J. Skokan: Regularity lemma for k-uniform hypergraphs, *Random Structures and Algorithms* **25** (2004), 1–42.

[16] T.C. Tao: A variant of the hypergraph removal lemma (preprint)

`http://arxiv.org/abs/math.CO/0503572`

[17] T.C. Tao: Szemerédi's regularity lemma revisited (preprint)

http://arxiv.org/abs/math.CO/0504472

Appendix: Proof of Theorem 9

We start with the following bound on the average square distance in $[0,1]$ with respect to the metric d^1_W (the proof is easy).

Lemma 11 *For every $W \in \mathcal{W}$, we have*

$$\int_0^1 \int_0^1 d^1_W(x,y)^2 \, dx \, dy \le \|W\|_2^2.$$

We'll apply this lemma to $W \circ W$; in view of inequality (1), this will relate the average square distance with respect to d_W to $\|W\|_\square$.

Now we turn to the proof of Theorem 9.

(a) Suppose that $\mathcal{P}$ is a weak Szemerédi partition with error ε^2. Then $U = W_{\mathcal{P}}$ is a step-function, and we know that $\|W - U\|_\square < \varepsilon^2$. It is trivial that $d_U(x,y) = 0$ for any two points x, y in the same class P_i. Since trivially

$$d_W(x,y) \le d_U(x,y) + d_{W-U}(x,y),$$

for any two points, it suffices to consider the metric d_{W-U}. By Lemma 11, the average square distance with respect to d_{W-U} is at most $2\|(W-U)\circ(W-U)\|_2^2$, and by (1), this is $\le \|W-U\|_\square \le \frac{1}{4}\varepsilon^3$. Let S be the set of "far away" points, i.e.,

$$S = \Big\{ x \in [0,1] : \int_0^1 d_{W-U}(x,y)^2 \, dy > \frac{1}{2}\varepsilon^2 \Big\}.$$

Then trivially $\lambda(S) \le \varepsilon$. We claim that for any two points $x, y \in [0,1] \setminus S$, we have $d_{W-U}(x,y) \le \varepsilon$. Suppose not. Then for every $z \in [0,1]$, we have

$$d_{W-U}(x,z) + d_{W-U}(y,z) \ge d_{W-U}(x,y) > \varepsilon,$$

and so, using the Cauchy-Schwartz inequality twice,

$$\begin{aligned}
\int_0^1 d_{W-U}(x,z)^2 \, dz + \int_0^1 d_{W-U}(y,z)^2 \, dz &\ge \left(\int_0^1 d_{W-U}(x,z) \, dz \right)^2 + \left(\int_0^1 d_{W-U}(x,z) \, dz \right)^2 \\
&\ge \frac{1}{2} \left(\int_0^1 d_{W-U}(x,z) \, dz + \int_0^1 d_{W-U}(x,z) \, dz \right)^2 \\
&= \frac{1}{2} \left(\int_0^1 d_{W-U}(x,z) + d_{W-U}(x,z) \, dz \right)^2 \\
&\ge \frac{1}{2}\varepsilon^2,
\end{aligned}$$

which is a contradiction since both $x, y \in S$.

Thus for any two points $x, y \in P_i \setminus S$, we have

$$d_W(x,y) \le d_U(x,y) + d_{W-U}(x,y) \le 0 + \varepsilon = \varepsilon,$$

which proves (a).

(b) We may assume that $S = \emptyset$, since deleting a set of measure ε^2 does not change the conclusion. Let h_i be the indicator function of the set P_i, i.e.,

$$h_i(x) = \begin{cases} 1 & \text{if } x \in P_i, \\ 0 & \text{if } x \in S \setminus P_i \end{cases}$$

We want to show that $\|W - W_{\mathcal{P}}\|_\square < \varepsilon$. It suffices to show that for any 0-1 valued function f,

$$\langle f, (W - W_{\mathcal{P}})f \rangle \le \varepsilon. \tag{2}$$

Let us write $f = f_{\mathcal{P}} + g$, where $f_{\mathcal{P}}$ is obtained by replacing $f(x)$ by the average of f over the class P_i containing x. Then $\langle f_{\mathcal{P}}, (W - W_{\mathcal{P}})f_{\mathcal{P}} \rangle = 0$ and $\langle g, W_{\mathcal{P}} g \rangle = 0$ by these definitions, and so we have

$$\langle f, (W - W_{\mathcal{P}})f \rangle = \langle f_{\mathcal{P}}, (W - W_{\mathcal{P}})f_{\mathcal{P}} \rangle + \langle f + g, (W - W_{\mathcal{P}})g \rangle = \langle 2f_{\mathcal{P}} + g, Wg \rangle$$

By Cauchy-Schwartz,

$$\langle f + g, Wg \rangle \le \|f + g\| \cdot \|Wg\| \le 2\|Wg\|.$$

We have

$$\|Wg\|^2 = \langle g, W^2 g \rangle = \int_{[0,1]^3} g(x)W(x,y)W(y,z)g(z)\,dx\,dy\,dz.$$

For each x, let $\phi(x)$ be an arbitrary, but fixed, element of the class P_i containing x. Then

$$\begin{aligned} &\int_{[0,1]^3} g(x)W(x,y)W(y,z)g(z)\,dx\,dy\,dz \\ &\qquad = \int_{[0,1]^3} g(x)\big(W(x,y)W(y,z) - W(x,y)W(y,\phi(z))\big)g(z)\,dx\,dy\,dz \\ &\qquad + \int_{[0,1]^3} g(x)W(x,y)W(y,\phi(z))g(z)\,dx\,dy\,dz \end{aligned}$$

Here the last integral is 0, since the integral of g over each partition class is 0. Furthermore,

$$\begin{aligned} &\int_{[0,1]^3} g(x)\big(W(x,y)W(y,z) - W(x,y)W(y,\phi(z))\big)g(z)\,dx\,dy\,dz \\ &\qquad \le \left(\int_{[0,1]^3} g(x)^2 g(z)^2\,dx\,dy\,dz \right)^{1/2} \\ &\qquad \times \left(\int_{[0,1]^3} \big(W(x,y)W(y,z) - W(x,y)W(y,\phi(z))\big)^2\,dx\,dy\,dz \right)^{1/2} \end{aligned}$$

Here

$$\int_{[0,1]^3} g(x)^2 g(z)^2\,dx\,dy\,dz \le 1,$$

and

$$\int_{[0,1]^3} \big(W(x,y)W(y,z) - W(x,y)W(y,\phi(z))\big)^2 \, dx \, dy \, dz = \int_0^1 d_W(z,\phi(z))^2 \, dz \le \varepsilon^2.$$

Thus

$$\int_{[0,1]^3} g(x)W(x,y)W(y,z)g(z) \, dx \, dy \, dz \le \varepsilon,$$

which completes the proof.