Notes on Szemeredi's Regularity Lemma

Ernie Croot

July 14, 2004

In these notes we give the standard proof of the Szemeredi regularity lemma. The presentation we give here is slightly unconventional, as we will first show how a certain simpleminded probabilistic approach does not work, and then we will show how the insight used to prove the regularity lemma can be used to prove a certain result on the distribution of integers in intervals. The purpose of proving this distribution result is to give one a rough idea of what types of arguments go in to the proof of the (Szemeredi) lemma, to make it easier to learn the result. Finally, in the last section, we will prove the (Szemeredi) lemma.

First, we need to define some terms. Given a graph G, and given vertex sets X and Y in G, we let $e(X,Y)$ denote the number of edges connecting a vertex of X to a vertex of Y. We define the density

$$d(X,Y) \ = \ \frac{e(X,Y)}{|X||Y|}.$$

We note that this density satisfies

$$0 \ \leq \ d(X,Y) \ \leq \ 1.$$

We say that a pair of disjoint vertex sets V_1 and V_2 of G is ϵ-regular, if given vertex subsets

$$X \subseteq V_1, \text{ and } Y \subseteq V_2,$$

satisfying

$$|X| \ \geq \ \epsilon |V_1|, \text{ and } |Y| \ \geq \ \epsilon |V_2|,$$

we have that

$$|d(V_1,V_2) - d(X,Y)| \ < \ \epsilon.$$

We now state the regularity lemma (which we will call a theorem):

Theorem 1 *For every $\epsilon > 0$ and $m \geq 1$ there exist constants $K > 0$ and $M > 0$ such that the following holds: If G is a graph having vertex set V satisfying $|V| = k \geq K$, then there exists an integer m' satisfying*

$$m \ \leq \ m' \ \leq M,$$

such that there is a partition of G into vertex sets $V_0, V_1, ..., V_{m'}$, having the following properties:

1. *$|V_0| \leq \epsilon |V|$. This set is called the exceptional set.*
2. *$|V_1| = |V_2| = \cdots = |V_{m'}|$.*
3. *All but at most $\epsilon(m')^2$ of the pairs (V_i, V_j), $1 \leq i, j \leq m'$ are ϵ-regular.*

1 Proof of the Regularity Lemma

1.1 A First Attempt

As a first attempt at proving the regularity lemma, we try a probabilistic approach. Let $k = mq + r$, where $0 \leq r \leq m - 1$. Then, we select at random (with uniform probability) disjoint vertex subsets $V_1, ..., V_m$, each having q elements, and we let V_0 be the remaining vertices of G.

Random subsets of a given set often have "regularity"-like properties with high probability. Let us see that this is not the case for our random sets $V_0, ..., V_k$ for the type of regularity we are interested in: Suppose that we started with G as a complete bipartite graph having vertex sets A and B, each having k vertices, such that the set of edges are all pairs $(a, b) \in A \times B$. Thus, G has $2k$ vertices and k^2 edges. Now suppose $m = 2$. Then, our random procedure chops G up into two new vertex sets V_1 and V_2 (V_0 is empty in this case), where each set has roughly $k/2$ vertices from A, and $k/2$ vertices from B.

Such a pair of vertex sets (V_1, V_2) cannot be $(1/2 - c)$-regular. To see this, we first decompose $V_1 = X_A \cup X_B$, and $V_2 = Y_A \cup V_B$, where X_A is the set of vertices in V_1 coming from the set A, and X_B, Y_A, Y_B are defined analogously. Since

$$|X_A| \ \sim \ |X_B| \ \sim \ |Y_A| \ \sim \ |Y_B| \ \sim \ \frac{k}{2},$$

we find that

$$e(V_1, V_2) \ = \ e(X_A, Y_B) \ + \ e(X_B, Y_A) \ \sim \ \frac{k^2}{4} + \frac{k^2}{4} \ = \ \frac{k^2}{2}.$$

Thus,

$$d(V_1, V_2) \;=\; \frac{e(V_1, V_2)}{|V_1||V_2|} \;\sim\; \frac{k^2/2}{k^2} \;=\; \frac{1}{2}.$$

Now, if (V_1, V_2) *were* $(1/2 - c)$-regular, then we would have that

$$|d(V_1, V_2) \;-\; d(X_A, Y_A)| \;<\; \frac{1}{2} - c;$$

however,

$$d(V_1, V_2) \;-\; d(X_A, Y_A) \;=\; d(V_1, V_2) \;-\; 0 \;\sim\; \frac{1}{2}.$$

This argument generalizes to $m \geq 3$. Thus, we see that a "naive" probabilistic approach will not work. Moreover, it is difficult to imagine any simple modification of the argument working, as we know from the work of Gowers that the parameters in the conclusion of the regularity lemma must have "tower-type" dependence on ϵ and m; and so, it seems that one would have to use an *iterative* approach to prove the regularity lemma, rather than just a one-step approach as above.

1.2 The Idea

The basic idea for how to prove "Szemeredi regularity"-type theorems is to cook up a certain "$\mathcal{L}_2$ norm" such that if some partition of sets fails to be "regular", then we can refine the partition so that our norm increases by a small constant depending only on ϵ. By giving a good upper bound on the norm of any partition, we then deduce that the refinement cannot go on forever; and so, it must stop with a "regular partition".

Perhaps the best way to give a detailed description of the key ideas at play in the regularity lemma is to prove a much simpler theorem, which nonetheless requires some of the same methods to prove. In this subsection we will do just that.

First, we need a notion of regularity for our problem. Given a subset P of the integers from $\{1, 2, ..., k\}$, and given an interval $I \subseteq [1, k]$, we denote the set of integers in P belonging to I by P_I. Now, we will say that an interval I is ϵ-regular if the following holds: For every subinterval $J \subseteq I$ satisfying

$$|J| \;=\; \delta|I|, \quad \text{where } \delta > \epsilon,$$

we have

$$||P_J| \;-\; \delta|P_I|| \;<\; \epsilon|J|.$$

This is just saying that I has about the expected number of elements in every subinterval J when ϵ is small. Notice how similar-looking this notion of regularity is to the one appearing in Szemeredi's regularity lemma.

The theorem mentioned above is as follows:

Theorem 2 *Suppose that $0 < \epsilon < 1$, and $m \geq 1$ are given. Then, there exists an integer $M > m$ such that the following holds for all sufficiently large integers k: Suppose that $P \subseteq \{1, ..., k\}$. Then, for some m' satisfying*

$$m \ \leq \ m' \ \leq \ M$$

we will have that all but at most $\epsilon m'$ of the following consecutive intervals are ϵ-regular:

$$(0, k/m'], \ (k/m', 2k/m'], \ (2k/m', 3k/m'], \ ..., \ ((m'-1)k/m', k].$$

Proof. Before we launch into the heart of the proof, we need to define the function

$$f(t) \ = \ t \sum_{1 \leq j \leq t} |P_{((j-1)k/t, jk/t]}|^2.$$

That is, we are summing the square of the number of elements of P in each of t consecutive sub-intervals. This function $f(t)$ satisfies the uniform upper bound

$$f(t) \ \leq \ k^2; \tag{1}$$

and, if P contains θk elements in $\{1, ..., k\}$, then we have the easy-to-prove lower bound

$$f(t) \ \geq \ \theta^2 k^2. \tag{2}$$

If $m' = m$ satisfies the conclusion above, then we are done. So we may assume that at least ϵm of the intervals

$$(0, k/m], \ ..., \ ((m-1)k/m, k]$$

are not ϵ-regular.

Now let h be the least integer larger than ϵ^{-3}, and then let $n = mh$. Consider the value of $f(n)$: To calculate it, we will need a definition. If

$$I \ = \ ((j-1)k/m, jk/m], \tag{3}$$

then we define $I_1, ..., I_h$ to be consecutive subintervals of width $k/(mh)$ which tile I. With this notation, we have that

$$f(n) \ = \ mh \sum_{\substack{1 \le j \le m \\ I \ = \ ((j-1)k/m, jk/m]}} \sum_{i=1}^{h} |P_{I_i}|^2.$$

To see how $f(n)$ and $f(m)$ compare, suppose that I is as in (3). Then, from the Cauchy-Schwarz inequality we can conclude that

$$m|P_I|^2 \ \le \ mh \sum_{i=1}^{h} |P_{I_i}|^2;$$

and so, we have established

Fact. $f(n) \ge f(m)$. That is, when we refine our partition of P, the value of f cannot go down.

But we will show even more. We will now see that if there are ϵm intervals in our m partition that are not ϵ-regular (as we have assumed), then f increases by $c(\epsilon)k^2$, where the constant here depends only on ϵ. To see this, suppose that I is as in (3), and is not ϵ-regular. Then, there is some subinterval J of I, having length $\delta|I|$, where $\delta > \epsilon$, such that either

$$|P_J| \ > \ \delta|P_I| + \epsilon|J| \ > \ \delta|P_I| + \epsilon^2|I|, \tag{4}$$

or

$$|P_J| \ < \ \delta(1+\epsilon)^{-1}|P_I| \ < \ \delta|P_I| - \epsilon|J| \ < \ \delta|P_I| - \epsilon\delta|I|. \tag{5}$$

Suppose now that (4) holds. Then, if we chop up I into the intervals $I_1, ..., I_h$, the union of some $\sim \delta h$ of these intervals contain J. Since J has slightly more elements of P than expected, the elements of P_I are not uniformly distributed amongst $P_{I_1}, ..., P_{I_h}$, and from an application of Cauchy-Schwarz inequality we can show that for $\epsilon > 0$ sufficiently small,

$$mh \sum_{i=1}^{h} |P_{I_i}|^2 \ > \ m|P_I|^2 + m(\epsilon^4 + O(1/h^2))|I|^2$$

For h small enough this error $O(1/h^2)$ will be smaller than $\epsilon^4/2$ in absolute value.

Since there are ϵm intervals I of width k/m that are not ϵ-regular, we can deduce that

$$f(n) \; > \; f(m) + (\epsilon m)m(\epsilon^4/2)|I|^2 \; = \; f(m) + (\epsilon^5/2)k^2,$$

We get a similar conclusion if we assume (5); in particular, that

$$f(n) \; > \; f(m) + c(\epsilon)k^2,$$

where the constant here depends only on ϵ.

The idea of the proof is to now iterate the above argument: Starting with a partition of $[1, k]$ into m intervals, we either have that the conclusion of the theorem is satisfied; or else, there are at least ϵm of the intervals that are not ϵ-regular. If this is the case we iteratively refine the interval $[1, k]$ into a set of n_i intervals, where

$$m = n_1 \; < \; n_2 \; < \; n_3 \; < \; \cdots,$$

such that

$$f(n_i) \; > \; f(n_{i-1}) + c(\epsilon)k^2,$$

From (1) we must obviously have that the procedure terminates after at most

$$c(\epsilon)^{-1} \text{ steps}$$

with an m' satisfying the conclusion of the theorem.

1.3 The Proof of Szemeredi's Regularity Lemma

The analogous function (to the one in the previous subsection) f we use for Szemeredi's lemma is the following: Suppose we partition the set V of G as $V = V_0 \cup \cdots \cup V_{m_0}$, where each of the V_i's have the same size for $i \geq 1$, and where $|V_0| \leq \epsilon k$. Then, we define

$$f(V_0, ..., V_{m_0}) \; = \; \frac{1}{m_0^2} \sum_{1 \leq i < j \leq m_0} d(V_i, V_j)^2.$$

It is easily seen that

$$f(V_0, ..., V_{m_0}) \; < \; \frac{1}{2}.$$

Suppose now that we start with a partition $V_0, ..., V_{m_0}$ of V, where $|V_0| < \epsilon k/2$, and where $|V_1| = \cdots = |V_{m_0}|$. Further, suppose that this partition does not satisfy the conclusion of the regularity lemma. Thus, for at least ϵm_0^2 pairs (i,j), with $i < j$, there exist pairs of vertex sets $(X_i(j), X_j(i))$, $X_i(j) \subseteq V_i$, $X_j(i) \subseteq V_j$, such that either

$$d(X_i(j), X_j(i)) \; > \; d(V_i, V_j) + \epsilon, \tag{6}$$

or

$$d(X_i(j), X_j(i)) \; < \; d(V_i, V_j) - \epsilon. \tag{7}$$

If (V_i, V_j) *is* regular, then we will assume $X_i(j)$ and $X_j(i)$ are the empty set.

We use these vertex sets to refine our partition of V, where we will break each V_i up into 2^{m_0-1} pieces, possibly of unequal size: Basically, we partition V_i according to which combinations of the following vertex sets that a vertex $v \in V_i$ lies in

$$X_i(1), ..., X_i(i-1), X_i(i+1), ..., X_i(m_0).$$

There are 2^{m_0-1} combinations here, which includes those vertices that do not lie in any of the sets $X_i(j)$'s. Denote these new vertex sets by $V_{i,n}$, where $n = 1, ..., 2^{m_0-1}$. Note that V_i is the union of $V_{i,n}$ over all n.

These new vertex sets $V_{i,n}$ may be of unequal size, so we will further refine them: Let s be the least integer greater than $k/(m_0)^2 2^{m_0}$. Then, we break these $V_{i,n}$'s up into a series of even smaller sets, each of size s, in an arbitrary manner. As the sizes of the sets $V_{i,n}$ may not be evenly divisible by s, there will be some "scraps" that we shunt into the exceptional set. [1]

It is obvious that there at most the following number of "scrap vertices"

$$\sum_{i=1}^{m_0} \sum_{n=1}^{2^{m_0-1}} \frac{k}{m_0^2 2^{m_0}} \; < \; \frac{k}{m_0}.$$

As the exceptional set was initially (the first set V_0) of size at most $\epsilon k/2$, in the new partition it will have size at most $k(\epsilon/2 + O(1/m_0))$.

[1] For example, if a vertex set has $st + r$, $0 \le r \le s - 1$ vertices, then we break the set up into t vertex sets each of size s, and then we will be left with one "scrap" vertex set of size r.

We now consider how much f grows when we use this new vertex partition. We suppose that V_i, V_j is a pair of vertex sets that are not ϵ-regular. Then, when we pass from the partition $V_0, ..., V_{m_0}$ to this finer partition, the quantity $d(V_i, V_j)^2$ in the expression for f gets replaced with

$$\sum_{\substack{t_0 \leq i \leq t_1 \\ u_0 \leq i \leq u_1}} d(W_i, W_j)^2,$$

where $W_0, ..., W_r$ is the new vertex partition, and where $W_{t_0}, ..., W_{t_1}$ are those vertex sets that lie in V_i, and $W_{u_0}, .., W_{u_1}$ are those that lie in V_j.

Let us see how much the union of $W_{t_0}, ..., W_{t_1}$ differs from V_i: For each $n = 1, ..., 2^{m_0-1}$, when we break up the set $V_{i,n}$ into equal-sized parts W_i of size about $k/m_0^2 2^{m_0}$, each such n can contribute at most $k/m_0^2 2^{m_0}$ scrap vertices; so, V_i and the union of $W_{t_0}, ..., W_{t_1}$ differ by at most

$$2^{m_0} \times \frac{k}{2^{m_0} m_0^2} \;=\; \frac{k}{m_0^2} \text{ vertices.}$$

Likewise, for this fixed pair i, j if we let X denote those W_i's lying entirely in $X_i(j)$, and let Y be those lying entirely in $X_j(i)$, then we deduce that the union of the vertex sets in X has size

$$|X_i(j)|(1 + O(1/m_0)),$$

and the union of vertex sets in Y has size

$$|X_j(i)|(1 + O(1/m_0)).$$

It follows also that $e(X, Y)$ (through an abuse of notation this means $e(A, B)$, where A is the union of vertex sets $W \in X$ and B is the union of vertex sets $W' \in Y$) and $e(X_i(j), X_j(i))$ differ by at most

$$O(|X_i(j)||X_j(i)|/m_0).$$

Now suppose that (6) holds. Then, we have that

$$\begin{aligned} \sum_{W \in X, W' \in Y} d(W, W') &= \frac{e(X, Y)}{|W_1|^2} \\ &= \frac{|X_i(j)||X_j(i)| d(X_i(j), X_j(i))(1 + O(1/m_0))}{|W_1|^2} \\ &> \frac{|X_i(j)||X_j(i)|}{|W_1|^2}(d(V_i, V_j) + \epsilon + O(1/m_0)). \end{aligned}$$

Thus, as expected, the $d(W, W')$'s are (on average) a little bigger than they "should be" when $W \in X$ and $W' \in Y$. We also know that

$$\begin{aligned}\sum_{\substack{u_0 \le i \le u_1 \\ t_0 \le j \le t_1}} d(W_i, W_j) &= \frac{e(V_i, V_j)}{|W_1|^2}(1 + O(1/m_0)) \\ &= \frac{|V_1|^2 d(V_i, V_j)}{|W_1|^2}(1 + O(1/m_0)).\end{aligned}$$

Using these facts and the Cauchy-Schwarz inequality, it is not too difficult to show that

$$\sum_{\substack{u_0 \le i \le u_1 \\ t_0 \le j \le t_1}} d(W_i, W_j)^2 \;>\; \frac{|V_1|^2}{|W_1|^2}(d(V_i, V_j)^2 + c\epsilon^4 + O(1/m_0)).$$

for some absolute $c > 0$.

If (V_i, V_j) were regular, we would likewise get the bound

$$\sum_{\substack{u_0 \le i \le u_1 \\ t_0 \le j \le t_1}} d(W_i, W_j)^2 \;\ge\; \frac{|V_1|^2}{|W_1|^2}(d(V_i, V_j)^2 + O(1/m_0)).$$

It follows that

$$\begin{aligned}f(W_0, W_1, ..., W_r) \;\ge\; & \frac{|V_1|^2}{|W_1|^2 r^2} \sum_{(V_i, V_j) \text{ not regular}} (d(V_i, V_j)^2 + c\epsilon^4 + O(1/m_0)) \\ & + \frac{|V_1|^2}{|W_1|^2 r^2} \sum_{(V_i, V_j) \text{ regular}} (d(V_i, V_j)^2 + O(1/m_0)).\end{aligned}$$

Now, as

$$r \;=\; \frac{m_0 |V_1| (1 + O(1/m_0))}{|W_1|},$$

we deduce from this that

$$f(W_0, W_1, ..., W_r) \;\ge\; f(V_0, V_1, ..., V_{m_0}) + c\epsilon^5 + O(1/m_0).$$

The same bounds holds if we were to assume (7), rather than (6).

Just as with the regularity theorem in the previous sub-section, we iterate the refinement until we get a regular set: We construct a sequence of vertex partitions

$$\mathbf{P}_1 := \{V_0, ..., V_{m_0}\} \ \geq \ \mathbf{P}_2 \ \geq \ \cdots,$$

where $\mathbf{P}_i \geq \mathbf{P}_{i+1}$ means that the partition $\mathbf{P}_i$ is finer than $\mathbf{P}_{i+1}$ except for the exceptional sets. These partitions satisfy

$$f(\mathbf{P}_{i+1}) \ \geq \ f(\mathbf{P}_i) + c\epsilon^5 + O(1/|\mathbf{P}_i|)$$

for $i \geq 2$, and the number of vertex sets in $\mathbf{P}_{i+1}$ is at worst exponential in the number of vertex sets in $\mathbf{P}_i$. Obviously this procedure must terminate after at most $O(c^{-1}\epsilon^{-5})$ steps as f is always bounded from above by $1/2$.

Thus, our algorithm terminates with an ϵ-regular partition. Note here that it is easy to check that the exceptional set never has more than $3\epsilon k/4 < \epsilon k$ vertices, provided m_0, the size of our initial partition, is sufficiently large.