Analysis of Boolean Functions	(CMU 18-859S, Spring 2007)

Lecture 28: Szemerédi's Regularity Lemma in $\mathbb{F}_2^n$

May 1, 2007

Lecturer: Ryan O'Donnell *Scribe: Ryan O'Donnell*

1 Triangle-freeness in $\mathbb{F}_2^n$

In the last lecture we saw Roth's Theorem in $\mathbb{F}_3^n$, which gives an upper bound on how many elements a subset of $\mathbb{F}_3^n$ can have while remaining 3-AP-free. Asking about 3-AP-free sets in $\mathbb{F}_2^n$ doesn't make a lot of sense, since $x, x+d, x+2d = x$ are never all distinct. However it *does* make sense to talk about "triangles":

Definition 1.1 *A* triangle *in $\mathbb{F}_2^n$ (or in any additive group) is a triple x, y, z satisfying $x+y+z = 0$.*

Remark 1.2 *In $\mathbb{F}_2^n$, this is equivalent to satisfying $z = x + y$. In general, a triple $(x, y, x + y)$ is called a "Schur triple".*

We investigated Schur triples (equivalently, triangles) when studying the BLR test for linearity in Lecture 2. There we considered picking $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{F}_2^n$ at random, letting $\boldsymbol{z} = \boldsymbol{x} + \boldsymbol{y}$, and checking whether $f(\boldsymbol{x}) + f(\boldsymbol{y}) + f(\boldsymbol{z}) = 0$ (mod 2). If we were interested in triangle-freeness, we might instead check that $f(\boldsymbol{x}) = f(\boldsymbol{y}) = f(\boldsymbol{z}) = 1$; equivalently, that $f(\boldsymbol{x})f(\boldsymbol{y})f(\boldsymbol{z}) = 1$. Recall from Lecture 2 (and similarly in Lecture 27):

Proposition 1.3 *Let $f : \mathbb{F}_2^n \to \{0, 1\}$. If $\boldsymbol{x}, \boldsymbol{y}$ are chosen randomly from $\mathbb{F}_2^n$ and $\boldsymbol{z} = \boldsymbol{x} + \boldsymbol{y}$, then*

$$\mathbf{Pr}[f(\boldsymbol{x}) = f(\boldsymbol{y}) = f(\boldsymbol{z}) = 1] = \mathbf{E}[f(\boldsymbol{x})f(\boldsymbol{y})f(\boldsymbol{z})] = \sum_{\alpha \in \mathbb{F}_2^n} \hat{f}(\alpha)^3.$$

Consider now the problem analogous to Roth's for triangle-freeness — namely, how big can a subset of $\mathbb{F}_2^n$ be while remaining triangle-free? It may seem like the proof we gave for Roth's Theorem (which also analyzed $\sum_\alpha \hat{f}(\alpha)^2$) should work exactly equivalently here — i.e., if f is a little bit dense and also uniform, then it contains a triangle; otherwise, it has a large Fourier coefficient, hence it's denser on an affine subspace of one fewer dimension, and we can iterate. The catch is that if we find a triangle in the subspace, it doesn't necessarily give a triangle in the original space. Specifically, if T is an invertible affine map, and $x+y+z = 0$ and T is an invertible affine map, then $T^{-1}x + T^{-1}y + T^{-1}z = 0$ only if T is in fact linear (not just affine).

Indeed, there are extremely dense triangle-free subsets of $\mathbb{F}_2^n$:

Example 1.4 *Let $A = \{x \in \mathbb{F}_2^n : x_1 = 1\}$, a set of density $1/2$. Then A is triangle-free, since*

$$x, y, z \in A \Rightarrow x_1 = y_1 = z_1 = 1 \Rightarrow x_1 + y_1 + z_1 = 1 \Rightarrow x + y + z \neq 0.$$

1.1 Testing triangle-freeness

There is a very interesting related question though, which can be motivated by property testing: Suppose one has access to an unknown function $f : \mathbb{F}_2^n \to \{0,1\}$ (thought of as indicating a subset A of $\mathbb{F}_2^n$) and one wants to test that it is triangle-free. The natural algorithm, again, would be to choose $\boldsymbol{x}, \boldsymbol{y}$ randomly and check that not all of $\boldsymbol{x}, \boldsymbol{y}$, and $\boldsymbol{x} + \boldsymbol{y}$ are in A (three queries). Clearly, if A is triangle-free this test will always pass. We would like to show that if A is ϵ-far from being triangle-free then the test will have at least a slight chance of failing. If the probability is some function $\delta(\epsilon)$, then we can boost it up to $2/3$ and get a one-sided test for triangle-freeness using $O(1/\delta(\epsilon))$ queries.

As usual, to analyze the soundness. we look at the contrapositive:

Question 1.5 *Suppose* $\mathbf{Pr}[f(\boldsymbol{x}) = f(\boldsymbol{y}) = f(\boldsymbol{x} + \boldsymbol{y})] = 1] < \delta$. *Must f be close to being triangle-free, where the closeness only depends on δ?*

Since being triangle-free is an anti-monotone property, to measure closeness one only needs to consider how many points must be *deleted* from A to make it truly triangle-free. In other words, we may equivalently ask:

Question 1.6 *Suppose* $\mathbf{Pr}[f(\boldsymbol{x}) = f(\boldsymbol{y}) = f(\boldsymbol{x}+\boldsymbol{y})] = 1] < \delta$. *Can we delete some $c(\delta)2^n$ points from f and make it triangle-free?*

We will show that the answer to this question is "yes". However, the only bound known on $c(\delta)$ is extremely bad.

Definition 1.7 *The function $2 \uparrow\uparrow m$ is defined to be an exponential tower of 2's of height m. Its inverse function is denoted* $\log^*$.

The best bound known for $c(\delta)$ satisfies $c(\delta) < (\log^*(1/\delta))^{-\Omega(1)}$:

Theorem 1.8 *Suppose* $\mathbf{Pr}[f(\boldsymbol{x}) = f(\boldsymbol{y}) = f(\boldsymbol{x} + \boldsymbol{y})] = 1] < 1/(2 \uparrow\uparrow \mathrm{poly}(1/\epsilon))$. *Then one can delete $\epsilon 2^n$ points from f and make it triangle-free.*

Corollary 1.9 *There is a $(2 \uparrow\uparrow \mathrm{poly}(1/\epsilon))$-query test (with one-sided error) for a function $f : \mathbb{F}_2^n \to \{0,1\}$ being triangle-free.*

Theorem 1.8 was first proved by Green in 2004. The analogous result for triangles in graphs (i.e., 3-cycles) had long been known, with the proof using the famous Szemerédi's Regularity Lemma (SzRL) for graphs. Green's method involves proving a SzRL for functions on $\mathbb{F}_2^n$; we will see this in the next section.

We end this section by remarking that it is unknown, and an interesting open problem, if a non-tower-type bound is possible in Theorem 1.8. It may even be possible to get a polynomial relationship. (In the graph setting, Alon has shown that there are N-vertex graphs, for N arbitrarily large, which are ϵ-far from being triangle-free and yet contain only $(1/\epsilon)^{O(\log(1/\epsilon))}$ many triangles.)

2 Szemerédi's Regularity Lemma in $\mathbb{F}_2^n$

To analyze the triangle-freeness test, we follow [Green 2004] and first prove a "Szemerédi Regularity Lemma (SzRL) in $\mathbb{F}_2^n$". The usual SzRL is a structure theorem for all *graphs*; it says that any graph can be decomposed into a constant number of pieces so that the edge structure between almost all pairs of pieces is "pseudorandom". The SzRL for functions on $\mathbb{F}_2^n$ is somewhat similar; it says that for any function on $\mathbb{F}_2^n$, $f : \mathbb{F}_2^n \to [0,1]$ say, one can decompose $\mathbb{F}_2^n$ into constantly many pieces so that on almost all pieces, f is "pseudorandom" (specifically, uniform). Further, these pieces have a nice structure: they are all the cosets of a subspace of small codimension. Recall from Lecture 9:

Definition 2.1 *Let H be a subspace of $\mathbb{F}_2^n$. Then the sets $\{x + H : x \in \mathbb{F}_2^n\}$ are pairwise either equal or disjoint. These are the* cosets *of H, and together they partition $\mathbb{F}_2^n$.*

We say that the codimension *of H is $n - \dim(H)$. A subspace of codimension k is equivalent to the the set of all x satisfying the AND of k linearly independent constraints $\{\alpha_i \cdot x = 0\}_{i=1\dots k}$. A coset of this subspace is equivalent to all all x satisfying an AND of these constraints with different RHS's, $\{\alpha_i \cdot x = c_i\}_{i=1\dots k}$.*

Working with cosets is slightly annoying because for a typical H, there is no "canonical" choice of representatives for the cosets. For example, if $n = 2$ and $H = \{(0,0),(1,1)\}$, then one of the cosets of H is $(1,0) + H = \{(1,0),(0,1)\} = (0,1) + H$, and there is no natural way to choose between $(1,0)$ and $(0,1)$.

There is one case where the situation is nice, though; our oft-used setting of "restrictions" to coordinates:

Example 2.2 *Let $J \subseteq [n]$ and let H denote the span of all vectors e_i, $i \notin J$. This is a subspace of codimension $|J|$ (with the constraints being $\{x_j = 0 : j \in J\}$). The cosets of H are formed by all possible restrictions of the coordinates in J, and each is naturally identified with a vector $y \in \mathbb{F}_2^J \cong \{y : y_j = 0\ \forall j \notin J\}$.*

Just using these sorts of cosets, we can prove a significant warmup for the SzRL in $\mathbb{F}_2^n$.

2.1 The "degree-1" case

Let's introduce a weakening of the notion of being uniform:

Definition 2.3 *Let $f : \mathbb{F}_2^n \to \mathbb{R}$. We say that f is* $(\epsilon, 1)$-uniform *if $|\hat{f}(\alpha)| \leq 1$ for all $\alpha = e_i$ for some $i \in [n]$ (i.e., $|\hat{f}(S)| \leq \epsilon$ for all $|S| = 1$ in our old notation).*

Note that for monotone, boolean-valued functions, this is equivalent to all influences being smaller than ϵ. The notion of $(\epsilon, 1)$-uniformity has had some use in property testing; e.g., in the testing of LTFs.

The following "degree-1" variant on the SzRL in $\mathbb{F}_2^n$ contains most of the main ideas of the proof:

Lemma 2.4 *Let $f : \mathbb{F}_2^n \to \mathbb{R}$ satisfy $\|f\|_2 \le 1$, and let $\epsilon > 0$ be given. Then there is a decision tree of size at most $K(\epsilon) = 2 \uparrow\uparrow (1/\epsilon^3)$, where each internal node is labeled by a coordinate and where each leaf is labeled by the natural restricted subfunction of f, such that:*

$$\text{with prob.} \ge 1-\epsilon\text{, a rand. path down the tree ends at at an } (\epsilon, 1)\text{-uniform function.} \qquad (1)$$

Proof: Given a DT describing restricted subfunctions, we'll say that "splitting" a leaf involves replacing it with an internal node (querying some coordinate), and giving this node two leaf children representing the subfunctions on one more restricted variable.

We will start with the DT having no internal nodes and one leaf, labeled by f. Our plan is to grow the required DT in stages. At each stage, we ask whether (1) holds. If so, we conclude. Otherwise, we split all leaf-subfunctions g having a large $\hat{g}(e_i)$ on their coordinate of maximal Fourier coefficient and move to the next stage. Each stage at most doubles the size (number of leaves) in the tree. We will also show that the leaf-splitting in any stage increases a certain "progress measure" by at least ϵ^3. The progress measure will be bounded in $[0, 1]$ by definition, and hence the number of stages can be at most $1/\epsilon^3$. This will complete the proof of the theorem.

The progress measure we use is:

$$\mathrm{ExpImb} := \mathop{\mathbf{E}}_{\text{rand. path } \boldsymbol{y}}[\widehat{f_{\boldsymbol{y}}}(0)^2].$$

This measures the average extent to which f is imbalanced under a random path (restriction) in the current DT. We have

$$0 \le \mathrm{ExpImb} \le \|f\|_2^2 \le 1.$$

The left inequality is obvious; the rightmost one is by assumption. The inequality $\mathrm{ExpImb} \le \|f\|_2^2$ is a simple exercise (use induction). (Also, $\mathrm{ExpImb} \le 1$ is obvious if f's range is $\{0, 1\}$ or $[0, 1]$, the main cases of interest to us.)

Let's now analyze the effect of a split on ExpImb. Suppose we are splitting some leaf-subfunction g on a coordinate i with $|\hat{g}(e_i)| = \eta$. Write $\mu = \hat{g}(0)$. Before splitting, g contributes μ^2 to ExpImb (times the probability a random path reaches g). After splitting, we get two restricted functions g_0 and g_1, with $\hat{g_0}(0) = \mu + \eta$ and $\hat{g_1}(0) = \mu - \eta$. Their collective contribution to ExpImb will be $\frac{1}{2}(\mu+\eta)^2 + \frac{1}{2}(\mu-\eta)^2 = \mu^2 + \eta^2$ (again, times the probability a random path reaches their parent).

Hence splitting on a subfunction g with a "degree-1" Fourier coefficient at least ϵ in magnitude increases g's weighted contribution to ExpImb by at least ϵ^2. Now we see that if (1) does not hold, at least an ϵ fraction of the leaves must have a degree-1 Fourier coefficient with magnitude at least ϵ. So indeed, splitting all of these subfunctions will increase ExpImb by at least ϵ^3. This completes the proof. $\square$

It will be convenient to modify this result so that the DT is more structured; specifically so that it is a full DT of depth at most $K(\epsilon)$ in which all nodes in the same level query the same variable. In this case, it is just as though we are considering restrictions to some $K(\epsilon)$ coordinates.

Corollary 2.5 *In the setting of Lemma 2.4, there is a subset $J \subseteq [n]$ of cardinality at most $K(\epsilon)$ such that for at least a $1-\epsilon$ fraction of the restrictions y to J, the subfunction $f_{y \to H}$ is $(\epsilon, 1)$-uniform.*

Proof: To achieve this, we modify how we split at each stage. Suppose we have just finished all the splits from one stage in the above proof. We now further split each leaf-subfunction on all of the other coordinates that were used as splits in that stage. Since a split can never decrease ExpImb, after this additional splitting we still have that ExpImb increases by at least ϵ^3. We can now rearrange all of the newly added subtrees so that the coordinates are split in the same order in each.

This new process indeed generates full DTs in which the same variable is queried at each level. It remains to check that the depth after t stages is at most $2 \uparrow\uparrow t$. This is straightforward; if d_t is the depth after t stages, we have $d_{t+1} \leq d_t + 2^{d_t}$, and the bound follows because the base case has slack (after 1 stage the depth is at most $1 \leq 4 = 2 \uparrow\uparrow 1$). □

2.2 The general case

We would now like to extend this result to the full SzRL in $\mathbb{F}_2^n$, getting subfunctions that are η-uniform, not just $(\eta, 1)$-uniform. The same proof structure works, except that we must handle subfunctions g with $\hat{g}(\alpha)$ large for some $\alpha \notin \{e_1, \ldots, e_n\}$. It's possible to simply forge ahead, having the DT's internal nodes query arbitrary linear constraints, and having the leaves represent restrictions of f to cosets. However defining Fourier coefficients consistently for functions on cosets is slightly messy; hence we will use a twist:

Theorem 2.6 *Let $f : \mathbb{F}_2^n \to \mathbb{R}$ satisfy $\|f\|_2 \leq 1$, and let $\epsilon > 0$ be given. Then there is an invertible linear transformation T on $\mathbb{F}_2^n$ such that, if we define $h : \mathbb{F}_2^n \to \mathbb{R}$ by $h = f \circ T$, then the following holds:*

For at least a $1-\epsilon$ fraction of the restrictions y to the coordinates $J := \{1, 2, \ldots, K(\epsilon) = 2 \uparrow\uparrow (1/\epsilon^3)\}$, the subfunction $h_{y \to J}$ is ϵ-uniform.

Proof: We will follow the proof of Corollary 2.5. However, at each stage we will apply an invertible linear transformation to the input space in such a way that the set of coordinates the tree queries is always $\{1, 2, \ldots, d\}$ for some $d \in \mathbb{N}$.

Specifically, at a given stage we are considering all restrictions to the first d coordinates of some function $f \circ T$, where T is an invertible linear transformation. For each restriction $g : \mathbb{F}_2^{n-d} \to \mathbb{R}$ which is not ϵ-uniform, we select a Fourier coefficient $\alpha \neq 0$ for which $|\hat{g}(\alpha)| \geq \epsilon$. Let H be the subspace spanned by all of the α's selected, and let $\beta_1, \ldots, \beta_{d'}$ be a basis for this space. Note that $d' \leq 2^d$.

We now modify T applying an additional invertible linear transformation which maps β_i to e_{d+i} for each $i = 1 \ldots d'$; this is possible (and there will be many possible choices) since the β_i's are linearly independent. This brings us to a situation where for at least an ϵ fraction of the restrictions to the coordinates $\{1, \ldots, d\}$, the resulting subfunction g (of the new $f \circ T$) has the following property: $|\hat{g}(\gamma)| \geq \epsilon$ for some nonzero γ in the span of $\{e_{d+i}, \ldots, e_{d+d'}\}$.

We would like to show that splitting on all of the coordinates $d+1, \ldots, d+d'$ will increase ExpImb by at least ϵ^3. As before, it suffices to show that if $|\hat{g}(\gamma)| = \eta$, then splitting g on all of γ's nonzero coordinates increases its contribution to ExpImb by at least η^2. (We can afterwards rearrange the queries to the e_i's, as before.)

This is an easy exercise; in fact, one that we have implicitly seen in Lecture 7 on the Goldreich-Levin algorithm. Specifically, if $\boldsymbol{y}$ is a random restriction to the nonzero coordinates of γ, then $\mathbf{E}_{\boldsymbol{y}}[\widehat{g_{\boldsymbol{y}}}(0)^2] = \sum_{0 \leq \gamma' \leq \gamma} \hat{g}(\gamma')^2$, and this is clearly at least $\hat{g}(0)^2 + \hat{g}(\gamma)^2 \geq \hat{g}(0)^2 + \eta^2$. $\square$

Having proved this, we can now undo the linear transformation on the input space. Doing this converts restrictions to $K(\epsilon)$ coordinates into restrictions to the cosets of a subspace of codimension $K(\epsilon)$. Since each coset is affinely isomorphic to a copy of $F_2^{n-K(\epsilon)}$, there is a sensible notion of what it means for a function, restricted to a coset, to be ϵ-uniform. (Note that the notion of uniformity is invariant under invertible affine transformations.) This proves the statement which we consider to be the SzRL for $\mathbb{F}_2^n$:

Corollary 2.7 (SzRL in F_2^n) *Let $f : \mathbb{F}_2^n \to \mathbb{R}$ satisfy $\|f\|_2 \leq 1$, and let $\epsilon > 0$ be given. Then there is a subspace H of $\mathbb{F}_2^n$ of codimension at most $K(\epsilon) = 2 \uparrow\uparrow (1/\epsilon^3)$ such that for at least a $1-\epsilon$ fraction of the cosets of H, the function f when restricted to the coset is ϵ-uniform.*

3 Analyzing the triangle-freeness test

To derive Theorem 1.8 we will use the Theorem 2.6 version of the SzRL. A key point here is that applying invertible *linear* transformations (as opposed to *affine* transformations) does not affect triangles, since if $x + y + z = 0$ then $T^{-1}x + T^{-1}y + T^{-1}z = 0$. Thus to prove Theorem 1.8, we may first apply any invertible transformation we like to f's input space.

So let $f : \mathbb{F}_2^n \to \{0, 1\}$ satisfy $\mathbf{Pr}[f(\boldsymbol{x}) = f(\boldsymbol{y}) = f(\boldsymbol{x}+\boldsymbol{y}) = 1] \leq 1/(2 \uparrow\uparrow \mathrm{poly}(1/\epsilon))$. Apply Theorem 2.6 with the "ϵ" parameter set to $\epsilon^3/10$. Now we may assume that f has the property that for at least a $1 - \epsilon^3/10$ fraction of the restrictions to coordinates $J = \{1, 2, \ldots, K\}$ (where $K \leq 2 \uparrow\uparrow (10/\epsilon^9)$), the restricted function is $\epsilon^3/10$-uniform.

Write A for the set that f indicates. For each restriction y to J for which $f_{y \to J}$ is *not* $\epsilon^3/10$-uniform, delete all the points from $A_{y \to J}$. Further, for each restriction y for which $\mathbf{E}[f_{y \to J}] \leq \epsilon/2$, delete all the points from $A_{y \to J}$. The total number of points deleted from A is then at most $(\epsilon^3/10)2^n + (\epsilon/2)2^n \leq \epsilon 2^n$. We claim that after this deletion, the set A is triangle-free.

Suppose otherwise; say $a, b, c \in A$ satisfy $a + b + c = 0$. Letting f denote the indicator of the post-deletion A, by construction we must have that $f_{a_J \to J}, f_{b_J \to J}, f_{c_J \to J}$ are all $\epsilon^3/10$-uniform functions with density at least $\epsilon/2$ each. The proposition below then implies that if we pick a random $u, v \in \mathbb{F}_2^{n-K}$, there is at least an $\epsilon^3/40$ chance that $f_{a_J \to J}(u) = f_{b_J \to J}(v) = f_{c_J \to J}(w) = 1$. This immediately implies that $\mathbf{Pr}[f(\boldsymbol{x}) = f(\boldsymbol{y}) = f(\boldsymbol{x} + \boldsymbol{y}) = 1] \geq 2^{-K}(\epsilon^3/40) > 1/(2 \uparrow\uparrow \mathrm{poly}(1/\epsilon))$ for a large enough poly, a contradiction.

Proposition 3.1 *Let $f_1, f_2, f_3 : \mathbb{F}_2^n \to \mathbb{R}$ satisfy $\mathbf{E}[f_i] \geq \epsilon/2$ for each i, and further suppose that: (a) f_1 is $(\epsilon/10)^3$-uniform; and, (b) $\|f_2\|_2, \|f_3\|_2 \leq 1$. Then if $\boldsymbol{x}, \boldsymbol{y}$ are chosen independently at random from $\mathbb{F}_2^n$, we have $\mathbf{E}[f_1(\boldsymbol{x}) f_2(\boldsymbol{y}) f_3(\boldsymbol{x} + \boldsymbol{y})] \geq \epsilon^3/40$.*

Proof: By Fourier analysis, we know that the expectation is

$$
\begin{aligned}
& \sum_{\alpha} \widehat{f_1}(\alpha)\widehat{f_2}(\alpha)\widehat{f_3}(\alpha) \\
\geq\; & (\epsilon/2)^3 - \sum_{\alpha \neq 0} |\widehat{f_1}(\alpha)||\widehat{f_2}(\alpha)||\widehat{f_3}(\alpha)| \\
\geq\; & \epsilon^3/8 - (\epsilon^3/10) \cdot \sum_{\alpha \neq 0} |\widehat{f_2}(\alpha)||\widehat{f_3}(\alpha)| && \text{(by uniformity of } f_1\text{)} \\
\geq\; & \epsilon^3/8 - (\epsilon^3/10) \cdot \sqrt{\sum_{\alpha \neq 0} \widehat{f_2}(\alpha)^2} \sqrt{\sum_{\alpha \neq 0} \widehat{f_3}(\alpha)^2} && \text{(Cauchy-Schwarz)} \\
\geq\; & \epsilon^3/8 - \epsilon^3/10 && \text{(since } \mathbf{E}[f_2^2], \mathbf{E}[f_3^2] \leq 1\text{)} \\
=\; & \epsilon^3/40.
\end{aligned}
$$

□