Proving regularity lemmas via online optimization

Ignacio S. Aleman

EPFL

ignacio.aleman@epfl.ch

July 22, 2020

Abstract

Szemerédi's regularity lemma is a fundamental result in graph theory, in this work we revisit proofs of the lemma that are done via methods borrowed from online optimisation [1], [2]. We will try to give an intuition of why the particular method from online optimisation might be well suited for this problem. We also give a proof for a slight variant of the regularity closely following the strategy in [2].

Contents

The author would like to thank Juhan Aru for valuable comments and discussion. This work was done under the supervision of Juhan Aru as part of a semester project at EPFL. The author is supported by a grant from his Family. The author would also like to thank Nawel Naas for helping him with the mysteries of Tikz.

1 Introduction

Structural results are a very pleasing and useful type of property to have; In essence, it tells us that a certain object of study can be decomposed into low complexity (**structured**) and high complexity (**random like**) parts (See [3] for a brief survey).

When the objects are graphs, Szemerédi's regularity lemma is an example of such a result; It tells us that for any large graph there exists a very large partition, such that between most of these sets, the edges appear to be pseudo-random[1] (See [4]).

While the applications of Szemerédi's regularity lemma are rich[2] and vast, in this report we will be interested instead in how to prove the regularity lemma. As it turns out, it is possible to formulate the regularity lemma in terms of functions, which opened the doors to give fourier theoretic proofs (as done by Gowers in [5]) as well via ergodic theory (by Furstenberg in [6]).

An interesting generalisation was given by Tao in [7], where he gives a probabilistic formulation of the regularity lemma. The proof is surprisingly simple and follows a so called energy increment algorithm; this has two ingredients:

1. Show that if the approximation is bad one can improve it.
2. Show that it will take at most some fixed number of steps to get a good approximation.

Perhaps inspired by this kind of methodology, people started to look into online optimisation, there the objective is to iteratively update a solution until it satisfies some criterion; Indeed, this sounds very similar to what the energy increment method is doing!

As we will see, it is not very hard to prove a weaker version of the regularity lemma[3] via the so called Follow the Regularized Leader (FTRL), a popular framework in online optimisation. The weak version was first showed by Trevisan, Tulsiani and Vadhan in [1], and with a little more work Skorski in [2] gave a proof for the strong version. In this report we follow the proofs found in [2] for the strong regularity lemma, and we will use it to show a different variant[4].

[1] We will later formalise what we mean by this; for now you can think of a graph being pseudo-random if we might confuse it for a random graph (where the edges were added independently with some probability d).

[2] The most notable application of the regularity lemma is in the field of arithmetic progressions (which is also how the regularity lemma originated). Interestingly, graphs are rich enough to encode information about arithmetic progressions. See [4] for a compact tutorial.

[3] When we refer to the strong regularity lemma we will be refering to the original regularity lemma.

[4] Skorksi in [2] gives a partition where each set differs in size by at most one element, we will prove the variant where all sets are of the same size. This difference will be made clear in the preliminaries.

1.1 Outline

The article will be organised as follows: first we will review the necessary vocabulary to state the regularity lemma; we will then state the various formulations of the lemma that we will use. Next we will give an intuition of how FTRL works and why it might be useful to our case; as we will not need any of the theory of FTRL, we will state some results that motivate it without proof.

The end of the preport will be used to give proofs of the various regularity lemmas using FTRL. We note that we have found a potential mistake in [2]; details of this are explored in Appendix A.

1.2 Preliminaries

We start by formalizing what we meant in the introduction by edges appearing to look random. Note that when we talk about a random graph, we will mean one generated by the Erdős–Rényi model[5]; where edges are sampled independently with some probability d.

First we recall the notion of edge density:

Definition 1. ***(Edge Density).*** *Given a graph $G = (V, E)$, and two disjoint subsets $X, Y \subset V$, we define the density as follows*

$$d_G(X,Y) = \frac{E(X,Y)}{|X||Y|} = \frac{|\{(x,y) : x \in V, y \in Y, \{x,y\} \in E\}|}{|X||Y|} \tag{1}$$

One way to formalise the notion of a pseudo-random graph is to find some characteristic that we expect with high probability from a random graph. Say we are given two disjoint sets of vertices X and Y, and suppose that between these sets we add an edge independently with probability d; then, with high probability we will have the following:

Definition 2. *(**ϵ-regularity**). Given a fixed $\varepsilon > 0$ and $G = (V, E)$, let $A, B \subset V$. We say the pair (A, B) is ϵ-regular if, for every subset $X \subseteq A$ and $Y \subseteq B$, s.t. $|X| \geq \varepsilon|A|$ and $|Y| \geq \varepsilon|B|$ the following holds:*

$$|d(X,Y) - d(A,B)| \leq \varepsilon \tag{2}$$

Thus if (A, B) is ϵ-regular, we might be tempted to say that it was generated by some random process as describe above (even if it is deterministic), which is the motivation for calling such pairs pseudo-random (when we look at the crossing edges). We can now state Szemerédi's regularity lemma[6]:

[5]For our purposes we will be only dealing with bipartite graphs; we will start with two vertex sets of equal size, and then add edges (between the sets) independently and with the same probability d.

[6]A variant of the lemma is mostly the same but with a slight difference: instead of the partitions being of the same size, we allow them to differ by at most 1 element (this is the variant that was proved in [2]).

Theorem 1 (**Strong regularity lemma**)**.** *For every $G = (V, E)$, there exists a partition $V_1, ..., V_k$ with the following properties:*

1. *Exceptional set:* $V_1 \leq \epsilon|V|$
2. *Equipartition:* $|V_2| = |V_3| = ... = |V_k|$
3. *Regularity: For all but at most* ϵk^2 *pairs,* (V_i, V_j) *is* ϵ*-regular;* $1 < i, j \leq k$.

Where k is at most a power of twos of height poly$(1/\epsilon)$.

We next give a weak version of the regularity lemma:

Theorem 2 (**Weak regularity lemma**)**.** *Given $G = (V, E)$, and some fixed $\varepsilon > 0$, there exists a partition $V_1, ...V_k$ and numbers $d_{ij} \in \mathbb{R}$ for each pair of sets. Where $k \in 2^{O(d\varepsilon^{-2})}$, with $d = \frac{|E|}{|V|^2}$;s.t. the following holds for all $T, S \subseteq V$:*

$$\left| \sum_{i,j} E\left(S \cap V_i, T \cap V_j\right) - \sum_{i,j} d_{i,j} |S \cap V_i||T \cap V_j| \right| \leqslant \epsilon |V|^2 \tag{3}$$

We remark that in both the strong and weak regularity lemma what is being said is that one can approximate a graph by a much simpler one; in the case of the weak regularity, we can approximate it by a graph that is fully connected between the partitions (with appropriate weights d_{ij}), s.t. on average if we take arbitrary cuts the approximation is good. The strong regularity lemma is slightly different, but as we will see, we can reformulate it to be more similar to the weak one; what the strong version is saying is essentially the same, except that the *average absolute distance* of cuts inside all pairs of partitions is now small (a much stronger notion).

We will end this section by showing how to cast the weak regularity lemma as an expression only involving functions.

Lemma 1 (**Functional weak regularity formulation**)**.** *The weak regularity is equivalent to the following statement:*

Given some $g\colon V \times V \to \mathbb{R}$, and a set $\mathcal{F}$ as given bellow. Then there exists some $h\colon V \times V \to \mathbb{R}$ s.t.

$$h = \sum_{i,j} d_{ij} \mathbf{1}_{V_i \times V_j}$$

For some partition $V_1, ...V_k$ of V and numbers $d_{ij} \in \mathbb{R}$ with $k \in 2^{O(\varepsilon^{-2})}$ s.t. the following holds:

$$\forall f \in \mathcal{F} : |\langle f, g - h \rangle| \leqslant \epsilon |V|^2 \tag{4}$$

$$\mathcal{F} = \{f : f = \pm \mathbf{1}_{T \times S},\, T, S \subseteq V\} \tag{5}$$

Proof. Let $g = \mathbf{1}_E$,

$$
\begin{aligned}
|\langle f, g-h\rangle| &= |\langle \mathbf{1}_{S\times T}, \mathbf{1}_E - h\rangle| && (6)\\
&= |\langle \mathbf{1}_{S\times T}, \mathbf{1}_E\rangle - \langle \mathbf{1}_{S\times T}, \sum_{i,j} d_{ij}\mathbf{1}_{V_i\times V_j}\rangle| && (7)\\
&= |E(S,T) - \sum_{i,j} d_{ij}\langle \mathbf{1}_{S\times T}, \mathbf{1}_{V_i\times V_j}\rangle| && (8)\\
&= |\sum_{i,j} E\left(S\cap V_i, T\cap V_j\right) - \sum_{i,j} d_{ij}|S\cap V_i||T\cap V_j|| && (9)
\end{aligned}
$$

The series of equalities follow by the linearity of the inner product, and by noting that the indicator inner products are counting either edge sets or the intersection of sets. The rest of the proof follows directly from the definitions.

□

2 Online optimisation: FTRL

The standard framework in online optimisation can be viewed as finding an algorithm to solve the following game:

In each turn $t = 1, 2, ...$ you — i.e. the algorithm — propose/s a move, say $x_t \in K$, where K is the set of permissible moves. In turn, an adversary chooses some cost function say $f_t \in \mathcal{F}$ and reveals it to us at the end of that turn ($\mathcal{F}$ is not known a priori), and we will pay penalty $f_t(x_t)$.

The objective will be to find a solution such that the total cost at the end of the game is as low as possible. To see how well we are doing we will compete against someone who has the ability to see the future, but who can only make one choice for the whole game, in particular, he can pick a move with the lowest loss[7]. Against such an adversary we can define the regret against him as follows (assuming he made some decision g for the whole game):

$$
\mathrm{Regret}_T(g) := \sum_{t=1}^{T} f_t\left(x_t\right) - f_t(g) \tag{10}
$$

There is a general framework known as Follow the Regularized leader[8] (FTRL) that allows to give bounds on very general settings (i.e. there are no assumptions

[7]Competing against someone who at each steps picks the best move would lead to an arbitrary difference in performance.

[8]An excellent in depth treatment of FTRL can be found in the lecture notes of Alexander Rakhlin in [8]; for a more brief overview with the type of application that we will use we recommend the very accessible blog series by Luca Trevisan found in [9].

on $\mathcal{F}$ and K) on the above game. What's important for our case is that if $\mathcal{F}$ is a set of linear functions and $K = \mathbb{R}^n$, then FTRL gives us a specific algorithm and bound as we will now see.

FTRL for the aforementioned case:

FTRL

Input: c

1. $h_0 = 0$
2. $h_t = h_{t-1} - cf_{t-1}$, after which adversary reveals f_t.

If we stop the FTRL at any time T, then we get the following bound:

$$\text{Regret}_T(g) = \sum_{t=1}^{T} \langle f_t, h_t - g \rangle = O(\sqrt{T}) \tag{11}$$

This is an amazing result! It says that if we play enough time, we can on average perform as well as an opponent with the ability of foresight (albeit with a fixed move constraint).

The power is that we can be competitive against *any* adversary[9], in particular we can specify an adversary to help us build an approximator.

With this in mind, we define the FTRL with a bespoke adversary:

Bespoke FTRL

Input: c

1. $h_0 = 0$
2. If there exists some $f \in \mathcal{F}$ s.t. $\mathcal{A}(f, h_t)$ is true, let $f_t = f$, else halt
3. $h_t = h_{t-1} - cf_{t-1}$, go to 2.

In our application we define $\mathcal{A}$ as follows:

$$\mathcal{A}(f, h_t) = |\langle f, g - h_t \rangle| > \epsilon |V|^2$$

Thus we can view FTRL as a turn based game (as suggested in [9]), where each turn a builder proposes a solution h, and the adversary $\mathcal{A}$ checks if the solution is valid in some criterion and returns a fix $f \in \mathcal{F}$ if necessary, see Figure 1 for a simple schematic of this dynamic.

There are other examples of using FTRL — and variants of it[10] — to build

[9] By adversary we mean here the agent in charge of giving us the cost functions at each turn.

[10] We remark that in some cases FTRL is equivalent to a technique in convex optimisation known as Bregman projections (See [8]); there is also a specific instance of FTRL known as multiplicative weight updates which also yields many such results (See [10]).

approximators to combinatorial structures; such as for graph specifiers as done in [11], and for the famous Impagliazzo hard-core lemma in [12] which is a a fundamental result in complexity theory that deals with the difficulty of computing boolean functions.

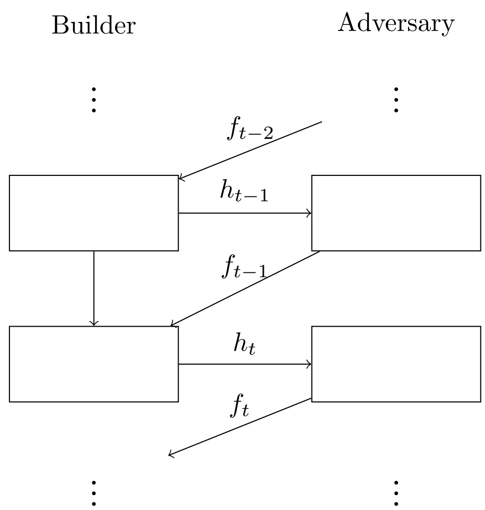

Figure 1: Schematic of FTRL

3 The regularity lemmas

We will now give proofs of both the weak and strong version using the FTRL; as stated in the beginning, the weak lemma will require little work, but we will need to do a bit more for the strong lemma.

3.1 The weak version

Based on the intuition gained in the previous section, we prove the functional weak regularity using the besoke FTRL.

Proof of theorem 2. We apply the bespoke FTRL on $\mathcal{F}$, with some $g : |V|^2 \to \mathbb{R}$ and c to be chosen later.

Suppose that at some iteration, there exists some $f \in \mathcal{F}$ s.t. the following holds

$$|\langle f, g - h \rangle| > \epsilon |V|^2$$

We then form $h^{'} = h - cf$ as prescribed. A simple calculation now shows that:

$$\|h^{'} - g\|_2^2 = \|h - g\|_2^2 - 2c\langle f, g - h\rangle + c^2\|f\|_2^2$$

Using the above two equations and setting $c = \epsilon$ we get

$$\|h^{'} - g\|_2^2 \leq \|h - g\|_2^2 - \epsilon^2|V|^2$$

Thus, at each iteration we are decreasing the distance to g by $\epsilon^2|V|^2$.

Since we start with $h_0 = 0$, we find that the initial distance is

$$\|h_0 - g\|_2^2 = \|g\|_2^2 = d|V|^2$$

for some $d \in [0, 1]$ since $\|g\|_2^2 \leqslant |V|^2$.

Thus we take at most $d\epsilon^{-2}$ iterations to terminate.

Observe also that the approximator h will have the following form:

$$h = \sum_{i=1}^{k} a_i \mathbf{1}_{T_i \times S_i}$$

With $k = O(d\epsilon^{-2})$, in order to get a partition it suffices to take all possible intersections yielding $O(2^{d\epsilon^{-2}})$ partitions[11].

Thus we have proven the functional formulation of the weak regularity, and coupled with Lemma 1, this implies the graph version of the weak regularity.

□

We remark that this is very similar — if not identical — to an energy increment argument. The key difference is in the way that updates are performed, here we use the FTRL approach. We note that we used very simple calculations to get bounds on the run time; in contrast to give bounds for energy increment arguments as in the probabilistic version of Tao [7] more complicated arguments are needed.

It's not clear why FTRL results in a cleaner and easier analysis, one reason might be that FTRL is optimal in some sense as described in the previous section.

[11]Note that before we take intersections, the partition pairs are approximated by constant densities $a_i \in \{-1, 1\}$.

3.2 The strong version

While it was very straight forward to prove the weak version (once we know about the functional formulation), it is not very clear how to apply FTRL to the strong version. We will use a slight variant that is more amenable to a functional formulation similar to that used for the weak version.

Theorem 3 (**Strong regularity lemma, version 2**). *For any graph G and any $\epsilon > 0$, there exists a partition $V_1, ..., V_k$ of V, and constants $d_{ij} \in \mathbb{R}$ s.t.*

$$\sum_{i<j} \max_{S \subset V_i, T \subset V_j} |E(S,T) - d_{i,j}|S||T|| \leqslant \epsilon |V|^2$$

Where k is at most a power of twos of height poly$(1/\epsilon)$.

While it's not obvious that this variant implies the usual strong regularity lemma, we can get some intuition of why this would be the case. What the variant is saying is the following: take any 2 partitions, and any cut for each partition, then if we look at the *absolute difference* between the actual edge set (focused on this subset) and a fully connected edge set with some constant density, then on average over all pairs this is smaller than ϵ.

In particular, if this holds, most pairs must behave pseudo randomly, i.e. they are approximated by constant density by taking any cuts. Finally, the other difference is that we do not have an equipartition in this variant, but as it turns out this is not an issue. A key property of ϵ-regularity is that it is preserved — albeit with a worsening of ϵ — when we shrink the original pairs. This is what will allow us to create a new partition from the original one with equally sized sets.

3.2.1 Proof of the strong variant

We will first give a proof using FTRL of the second variant of the regularity lemma; we then conclude by showing that the variant implies the original strong regularity lemma.

In the proof of the weak lemma we used the adversary to verify if our approximation was good by checking if there existed some cut (i.e. $f \in \mathcal{F}$) that gave a counterexample. Now instead we need the adversary to find a counterexample by finding multiple cuts, one for *each pair* of partitions. However the number of partitions we have changes at each iteration, thus we will allow $\mathcal{F}$ to vary, and we will call it $\mathcal{F}_k$ at iteration k.

Proof of Theorem 3. We start by building an $\mathcal{F}$ suited for the task that will vary with each iteration:

$$\mathcal{F}_k = \left\{ f = \sum_{i \leqslant j} a_{i,j} \mathbf{1}_{S_{i,j} \times T_{i,j}} : \quad a_{i,j} = \pm 1, \quad S_{i,j} \subset V_i,\, T_{i,j} \subset V_j \right\}$$

Similarly to the weak case — but with some slight modifications —, the setup will be the following:

At the beginning we specify some $g : V^2 \to \mathbb{R}$ and $c \in \mathbb{R}$.

- At step 0, $V_0 = V$ and $h_0 = 0$.

- At step k suppose we have a partition $V_1, ..., V_n$ and some h_k. The adversary then checks if there exists some $f \in \mathcal{F}_k$ s.t. $|\langle f, g - h \rangle| > \epsilon |V|^2$, then return such f as a cost, else it terminates and returns h.

- If we are not done, then we use the FTRL recipe to perform an update $h_{k+1} = h_k - cf$, after which we update the number of partitions[12].

First we will show how this implies the strong variant:

Let $g = \mathbf{1}_E$ and assume that we finished executing FTRL as described above, with output h s.t. $\forall f \in \mathcal{F}_k$ we have that

$$|\langle f, g - h \rangle| \leq \epsilon |V|^2$$

where,

$$h = \sum_{i \leq j} d_{ij} \mathbf{1}_{V_i \times V_j}$$

The proof is similar to that of lemma 1:

$$\begin{aligned}
|\langle f, g - h \rangle| &= |\langle \sum_{i \leqslant j} a_{i,j} \mathbf{1}_{S_{i,j} \times T_{i,j}}, \mathbf{1}_E - h \rangle| && (12)\\
&= |\sum_{i \leqslant j} a_{i,j} (\langle \mathbf{1}_{S_{i,j} \times T_{i,j}}, \mathbf{1}_E \rangle - \langle \mathbf{1}_{S_{i,j} \times T_{i,j}}, h \rangle |)| && (13)\\
&= |\sum_{i \leqslant j} a_{i,j} (E(S_{i,j}, T_{i,j}) - d_{ij} |S_{i,j}||T_{i,j}|)| && (14)\\
&= \sum_{i \leqslant j} |(E(S_{i,j}, T_{i,j}) - d_{ij} |S_{i,j}||T_{i,j}|)| && (15)
\end{aligned}$$

The first equality follows by definition of the objects, the second by linearity of the inner product. The third equality follows by recalling that h is constant in

[12] Observe that this is why we get a tower exponential in the number of partitions as we will see in the analysis.

$S_{i,j}, T_{i,j}$. Finally, it suffices to pick a_{ij} so as as to make the difference positive. Since this is true for all $S_{i,j}$ and $T_{i,j}$, it is also true for the maximum over all such pairs.

The final thing we need to show is an upper bound on the number of partitions k. First note that the algorithm will run for at most $d\epsilon^{-2}$, as we can use verbatim the run time proof used for the weak lemma.

The key difference is that we need to keep track of how big the number of partitions grows at each iteration. Note that in the weak case we only created partitions at termination.

Say we have a partition $V_1, ..., V_k$ at some iteration, and that we find some counterexample f. Then this means that for each partition k, we will have $k+1$ intersecting cuts added to h via f. In particular this means that we will have at most $k2^{k+1}$ partitions in the next iteration.

Hence at termination there are at most $d\epsilon^{-2}$ applications of the map $k \to k2^{k+1}$, i.e. we have a partition size of at most a power of 2 of height $d\epsilon^{-2}$.

□

3.2.2 Proof that the strong variant implies the strong regularity lemma

We show that the strong regularity (version 2) with ϵ^4 implies the usual strong regularity via a series of 2 propositions. First we show that we get ϵ-regularity

Proposition 1 (**Strong regularity version 2 with ϵ^4 implies 2ϵ-regularity**). *The strong regularity (version 2) of Theorem 3 with ϵ^4 implies that we have a partition V_1, ..., V_k, where all but at most $\epsilon|V|^2$ pairs are 2ϵ-regular; with the following additional bound (where I is the set of pairs that are not 2ϵ-regular):*

$$\sum_{(i,j)\in I} |V_i||V_j| \leqslant \epsilon|V|^2$$

We follow the same steps as done in [2]:

Proof of Proposition 1. We start by invoking Theorem 3 with ϵ^4; recall that this gives a partition V_i with the following property:

$$\sum_{i<j} \max_{S\subset V_i, T\subset V_j} |E(S,T) - d_{i,j}|S||T|| \leqslant \epsilon^4|V|^2$$

Dividing by $|V|^2$ and letting S_i', T_j' be the maximizers in the sum we get:

$$\sum_{i<j} \frac{|S_i'||T_j'|}{|V|^2} \left|d(S_i', T_j') - d_{i,j}\right| \leqslant \epsilon^4$$

Then, noting that when $|S'_i| \geqslant \epsilon|V_i|, |T'_j| \geqslant \epsilon|V_j|$ this gives:

$$\sum_{i<j} \frac{|V_i||V_j|}{|V|^2} \left|d(S'_i, T'_j) - d_{i,j}\right| \leqslant \epsilon^2$$

Next, if we consider sampling vertices independently and uniformly, the probability of sampling from V_i and then independently from V_j would be $p_{ij} = \frac{|V_i||V_j|}{|V|^2}$.

Thus applying the Markov inequality — with the previous probablistic setup — we get that:

$$\mathbb{P}(\left|d(S'_i, T'_j) - d_{ij}\right| > \epsilon) \leqslant \epsilon$$

This then gives that there is a set I s.t.

$$\sum_{(i,j)\in I} |V_i||V_j| \leqslant \epsilon|V|^2 \tag{16}$$

and if $(i, j) \notin I$, then we get that:

$$\left|d(S'_i, T'_j) - d_{ij}\right| \leqslant \epsilon$$

Since this was true for the maximizers, this is also true for any subsets S, T from V_i, V_j s.t. $|S| \geqslant \epsilon|V_i|$, and$|T| \geqslant \epsilon|V_j|$.

Finally, we get the last piece of the puzzle by using the triangle inequality:

For all $(i, j) \notin I$:

$$\begin{align}
|d(S, T) - d(V_i, V_j)| &\leqslant |d(S, T) - d_{ij}| + |d_{ij} - d(V_i, V_j)| \tag{17} \\
&\leqslant 2\epsilon \tag{18}
\end{align}$$

□

Before we move on, it's worth noting here that it's a simple application of the triangle inequality (at the end of the previous proof) that allows us to replace the actual densities $d(V_i, V_j)$ with some constants $d_{ij} \in \mathbb{R}$; the only price we need to pay is an ϵ increase in the regularity parameter. This is very useful as it is much easier to work with the relaxed[13] version.

[13] Note that in the proofs of FTRL we get arbitrary constants d_{ij} for the approximator, and it is in this sense that the problem is a relaxation; we do not need to care about the actual densities.

Now that we have established ϵ-regularity, it suffices to show that we can also build equipartitions to conclude.

Proposition 2 (**Partitions implies the existence of equipartition**). *ϵ-regular partitions implies the existence of an ϵ'-regular equipartition.*

The main ingredient to prove this Proposition is the following lemma (see appendix A for a proof):

Lemma 2 (**Regularity preserved under refinements**). *If (S, T) is ϵ-regular, for any $X \subset S$ and $Y \subset T$, then (X, Y) is ϵ'-regular*[14].

Using this lemma we can now prove Proposition 2.

Proof of Proposition 2. Given $G = (V, E)$ and a partition V_1, ..., V_k, where all but at most ϵk^2 pairs are ϵ-regular; We first build an equipartition:

Let $l > 0$, to be picked later, and let $s = \left\lceil \frac{|V|}{l} \right\rceil$

Next, for each $i \in [k]$, partition each set V_i into a number of disjoint sets of equal size s, and place the remaining elements into a special set:

$$V_i = \bigcup_{j=1}^{n_i} V_{i,j} \cup V_{i,r}$$

Where $|V_{i,1}| = |V_{i,2}| = ... = |V_{i,n_i}| = s$, and where $V_{i,r}$ is the set with the remaining vertices. Let also $V'_2 = V_{1,1}$, $V'_3 = V_{1,2}, ..., V'_{n_1+1} = V_{1,n_1}, V'_{n_1+2} = V_{2,1}$, ..., $V'_l = V_{k,n_k}$.

We next build the exceptional set by grouping together all the remainders of each partition:

$$V'_1 = \bigcup_{i=1}^{k} V_{i,r}$$

Then, by noting that $V_{i,r} < \left\lceil \frac{|V|}{l} \right\rceil$, and by setting $l = \frac{k}{\epsilon}$ we get:

$$|V'_1| \leqslant k\frac{|V|}{l} = \epsilon|V|$$

The last thing to do is to give an upper bound on the number of new pairs that are not ϵ'-regular, we will denote the set of such pairs as $I_{V'}$; similarly, let I_V be the set of pairs that are not ϵ-regular in the original partition $V_1, ..., V_k$.

[14] As discussed in the Appendix, this lemma as it is, is not correct; A variant might be true for some $\epsilon' = poly(\epsilon)$

First observe the following: Given a new pair (V_i', V_j') with $i, j > 1$, there are integers q, t s.t. $V_i' \subset V_q$ and $V_j' \subset V_t$. If (V_q, V_t) is ϵ-regular then (V_i', V_j') is ϵ'-regular by Lemma 2. In the derivation found bellow, this observation implies the first equality; the second line follows from noting that V_i''s are disjoint partitions of V_q. The last inequality follows from Proposition 1.

$$\sum_{(i,i)\in I_{V'}} |V_i'||V_j'| = \sum_{(q,t)\in I_V} \sum_{\substack{(i,j)\in I_{V'} \\ V_i'\subset V_q \\ V_j'\subset V_t}} |V_i'||V_j'| \tag{19}$$

$$\leqslant \sum_{(q,t)\in I_V} |V_q||V_t| \tag{20}$$

$$\leqslant \epsilon |V|^2 \tag{21}$$

Finally, note that $\frac{|V|}{l} \leqslant |V_i'|$, which implies that $|I_{V'}| \leqslant \epsilon l^2$.

Thus we have built a partition $V = V_1' \cup V_2' \cup ... \cup V_l'$ with the following properties:

1. (Bound on partition size) $l \leqslant \epsilon^{-1} k$
2. (Exceptional set) $V_1' \leqslant \epsilon |V|$
3. (Equipartiton) $|V_i'| = |V_j'|$, $1 < i < j \leqslant l$
4. (Regularity) All but at most ϵl^2 of the pairs are ϵ-regular[15].

$\square$

4 Conclusion

We have seen that using FTRL we can give very straight forward proofs for the regularity lemmas. Indeed, the only difficulty was encountered in the strong lemma in connecting the two different variants; note that we only used basic properties of regularity to do so. To give bounds on the run time required basic algebra, in contrast doing so for the energy increment as in the probabilistic version of Tao [7] requires more complicated arguments that rely on the clever use of Cauchy–Schwarz among other things[16]. Perhaps one reason that the analysis is simple is because as we saw, FTRL is optimal in a similar setting.

An interesting observation is that if we take theorem 1, and divide by V^2 (and by taking each vertex uniformly and independently) we get the following expression:

[15] Note that they are actually ϵ'-regular, but by an appropriate change of variable we get the result, assuming some variant of lemma 2 is true

[16] This should be taken with a grain of salt as there might easier proofs for the energy increment argument.

$$|\mathbb{E}[f(g-h)]| \leqslant \epsilon$$

Further, note that there is nothing specific here to the graph problem, instead of V^2, we could be looking at some set $\mathcal{X}$, and $\mathcal{F}$ could be a set of functions on some sigma-algebra of some events in $\mathcal{X}$. We can in this way get a more general statement as done in [1]. It is natural to ask whether it is possible to prove Tao's probabilistic generalisation in [7] using FTRL; in there he shows that one can approximate a random variable with a conditional expectation with bounded complexity. However when using additive updates, our solution might cease to be a conditional expectation for the the random variable that we wish to approximate, so in addition, some sort of projection after each update is necessary.

One of the more attractive characteristics of using FTRL (aside from the simpler derivation), is the adversarial dynamics of builder vs adversary. Since we know that it is optimal against any adversary, we can take it is a black box to design on the adversary/builder components.

An example of how one might further tweak the adversary of FTRL for the strong regularity is in trying to control the partition size. Recall that in the derivation of the strong regularity, the number of partitions grows in a tower exponential like fashion since we cannot know how many intersections the updates will have with the approximator. Unfortunately, there are examples where the bound matches the partition size (see [13]), so we cannot do anything in the general setting. However we conjecture that there must exists a rich enough family of graphs, where a much more economic partition exists. One approach would be to add an additional constraint in the adversary $\mathcal{A}$ and have him add a solution with at most a certain number of intersections.

References

[1] L. Trevisan, M. Tulsiani, and S. Vadhan, "Regularity, boosting, and efficiently simulating every high-entropy distribution," in *2009 24th Annual IEEE Conference on Computational Complexity*. IEEE, 2009, pp. 126–136.

[2] M. Skórski, "A cryptographic view of regularity lemmas: Simpler unified proofs and refined bounds," in *International Conference on Theory and Applications of Models of Computation*. Springer, 2017, pp. 586–599.

[3] T. Tao, "Structure and randomness in combinatorics," in *48th Annual IEEE Symposium on Foundations of Computer Science (FOCS'07)*. IEEE, 2007, pp. 3–15.

[4] D. Shagnik, "Applications of the szemerédi regularity lemma," in *Extremal Combinatorics*. FU Berlin.

[5] W. T. Gowers, "Hypergraph regularity and the multidimensional szemerédi theorem," *Annals of Mathematics*, pp. 897–946, 2007.

[6] H. Furstenberg and Y. Katznelson, "An ergodic szemer'edi theorem for commuting transformations," in *J. Analyse Math.* Citeseer, 1979.

[7] T. Tao, "Szemerédi's regularity lemma revisited," *arXiv preprint math/0504472*, 2005.

[8] A. Rakhlin, "Lecture Notes on Online Learning," 2009. [Online]. Available: http://www.mit.edu/~rakhlin/papers/online_learning.pdf

[9] L. Trevisan, "Online Optimization Post 4: Regularity Lemmas," 2019. [Online]. Available: https://lucatrevisan.wordpress.com/2019/05/16/online-optimization-post-4-regularity-lemmas/

[10] S. Arora, E. Hazan, and S. Kale, "The multiplicative weights update method: a meta-algorithm and applications," *Theory of Computing*, vol. 8, no. 1, pp. 121–164, 2012.

[11] Z. Allen-Zhu, Z. Liao, and L. Orecchia, "Spectral sparsification and regret minimization beyond matrix multiplicative updates," in *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*. ACM, 2015, pp. 237–245.

[12] B. Barak, M. Hardt, and S. Kale, "The uniform hardcore lemma via approximate bregman projections," in *Proceedings of the twentieth annual ACM-SIAM symposium on Discrete algorithms*. SIAM, 2009, pp. 1193–1200.

[13] W. T. Gowers, "Lower bounds of tower type for szemerédi's uniformity lemma," *Geometric & Functional Analysis GAFA*, vol. 7, no. 2, pp. 322–337, 1997.

Appendices

A Remark of Lemma 2

First note that Lemma 2, with $\epsilon' = \epsilon$ — as stated and used in [2] — cannot be true; a simple counterexample is to let S and T be singletons. It is not obvious how to correct the proof of Proposition 2 without this lemma.

In general, what is known as the **slicing lemma** is used in order to preserve ϵ-regularity when taking subsets. However, this lemma can only be applied if we have some lower bound on the size of the subset we are considering. Another angle of attack is to build a partition before we go to ϵ-regularity. In what follows we sketch such an approach, although as we will see it has it's own problem.

We start by using Theorem 3 with ϵ^4, as done in the proof of Proposition 1. The idea is to first build an equipartition from the given partition $V_1, ..., V_k$. We will use the same construction as done for Proposition 2, so we omit the details; recall that we built the partition as follows:

$$V_i = \bigcup_{j=1}^{n_i} V_{i,j} \cup V_{i,r}$$

where all $V_{i,j}$ are of the same size except for the remainder terms $V_{i,r}$ (one for each $i \in [k]$. Note that we will only group the remainders $V_{i,r}$ together after we have shown the regularity property.

Observe that if for the equipartition $V_{i,j}$ we have the following inequality:

$$\sum_{\substack{i<j \\ t,p}} p_{ij}(z,t) \max_{\substack{S \subset V_{i,t} \\ T \subset V_{j,p}}} |d(S,T) - d_{i,j}| \leqslant \epsilon^2 \tag{22}$$

where $p_{ij}(z,t) = \frac{|V_{iz}||V_{jt}|}{|V|^2}$,

then we are done, as we can follow the same approach as in the end of the proof of Proposition 1; i.e. since we have a bound on the expectation of the absolute deviation of the densities, we can use Markov's inequality to also bound the probability of this happening (At this point we could group the remainders to build the exceptional set).

We next give a partial proof of this statement:

$$
\begin{aligned}
\sum_{\substack{i<j\\ z,t}} p_{ij}(z,t) \max_{\substack{S\subset V_{i,t}\\ T\subset V_{j,p}}} |E(S,T)-d_{i,j}| &\leqslant \sum_{\substack{i<j\\ z,t}} p_{ij}(z,t) \max_{\substack{S\subset V_i\\ T\subset V_j}} |E(S,T)-d_{i,j}| \\
&\leqslant \sum_{i<j} p_{ij} \max_{\substack{S\subset V_i\\ T\subset V_j}} |E(S,T)-d_{i,j}| \\
&\leqslant \sum_{i<j} \frac{\epsilon^2|S||T|}{|V|^2} \max_{\substack{S\subset V_i\\ T\subset V_j}} |E(S,T)-d_{i,j}| \\
&\leqslant \epsilon^2
\end{aligned}
\tag{23}
$$

The first inequality follows since we are optimising over a superset; the second follows by first recalling that $p_{ij} = \frac{|V_i||V_j|}{|V|^2}$, and noting that $p_{ij} = \sum_{z,t} p_{ij}(z,t)$ since the right hand side is a product of partitions of left hand side.

The main issue is in the third inequality; in order to use Theorem 3 we need to replace V with S, one way to do so is to say that $|S| \geqslant \epsilon|V|$. However, we cannot ensure that the maximum will occur in this constrained setting; thus an idea is to add this constraint to all maximisers. The problem is that in the constrained setting we no longer get the first inequality (in general); what we would need is the following:

$$
\max_{\substack{S\subset V_{i,t}\\ T\subset V_{j,p}\\ |S|\geqslant\epsilon|V_{i,t}|\\ |T|\geqslant\epsilon|V_{j,p}|}} |E(S,T)-d_{i,j}| \leqslant \max_{\substack{S\subset V_i\\ T\subset V_j\\ |S|\geqslant\epsilon|V_i|\\ |T|\geqslant\epsilon|V_j|}} |E(S,T)-d_{i,j}|
$$

It is clear that this follows if $|V_{ij}| \geqslant \epsilon|V_i|$, and in the right hand side we replace ϵ with ϵ^2; however, if we had such a bound, then we could have directly used the slicing lemma as a substitute for lemma 2.

As a final note, in the probabilistic version given by Tao in [7], we also start with a partition; in there, the type of bound we get are different, so it's not clear how to mimic their approach for this case.