

THE INTERPLAY OF REGULARITY METHOD AND COMPUTER SCIENCE

MOHAMMAD BAVARIAN

ABSTRACT. In this report, we review the connections between regularity theory, low rank approximation of matrices and tensors, and approximate constraint satisfaction problems. Our presentation of the algorithmic aspects of regularity theory follows parts of [1, 2, 7]; wherever possible we cast the material in the modern language advocated by Tao [15].

One main motivation is to understand the power and limitations of the classical regularity theory, and to identify new algorithmic possibilities offered by the new developments in regularity theory (i.e. the relative Szemerédi theorem which first appeared in Green-Tao [11], and hypergraph regularity of Gowers [9], and Rödl et al [14]).

1. INTRODUCTION

Szemerédi’s regularity lemma is a structural theorem in extremal graph theory developed by Endre Szemerédi on the way to his celebrated proof of Erdős-Turán conjecture on the existence of arbitrary long arithmetic progression in subsets of integers with positive (upper) density. Although the influence of regularity theory has been first and foremost in extremal graph theory, by now it has become well-established that the scope of regularity theory is much wider than that of extremal graph theory; in some sense the main point of the regularity theory is to study dense combinatorial objects (be it graphs/hypergraphs or polynomials over finite fields).

The main signature of the regularity theory is a kind of decomposition theorem which allows one to write an object f (think of it as a graph adjacency matrix) as the sum of three components as follows:

$$f = f_{str} + f_{sml} + f_{psr}$$

where f_{str} is of bounded complexity, f_{sml} is small in some universal norm say L^2 , and f_{psr} ’s behavior is similar to a random object in a suitable sense. In a typical application of regularity lemma to computer science, one starts from an input f and then uses the algorithmic version of the above decomposition to reduce the optimization problem at hand to optimization over f_{str} (by giving up some approximation factor ϵ incurred because of presence of $f_{err} = f_{psr} + f_{sml}$). Finally, since f_{str} is of bounded complexity the related optimization over f_{str} can be done rather quickly.

Let us start by a simple example which though is not of much use from algorithmic perspective, it is the cleanest case to demonstrate the main aspects the regularity lemma.

Claim 1.1. *Let $f : \{-1, 1\}^n \rightarrow [-1, 1]$ be a bounded function over the discrete cube. For any $\epsilon > 0$ and any growth function $F : \mathbb{Z}^+ \rightarrow \mathbb{R}^+$ of at least linear growth¹ (typical choices are $F(x) = x^{O(1)}$ or $F(x) = e^{cx}$), there exists $M = O_{F,\epsilon}(1)$ and integer $k \leq M$ such that*

$$f = f_{str} + f_{sml} + f_{psr},$$

MASSACHUSETTS INSTITUTE OF TECHNOLOGY, CAMBRIDGE, MA, USA.

E-mail address: bavarian@mit.edu .

¹In what follows, unless otherwise stated, all the growth functions is assumed to be at least of linear growth to avoid trivial problems.

satisfying

- (1) f_{psr} is psuedorandom in the sense that $|\widehat{f_{psr}}(S)| \leq \frac{1}{F(k)}$ for all $S \subseteq [n]$.
- (2) $\|f_{sml}\|_2 \leq \epsilon$.
- (3) f_{str} has a k -sparse Fourier representation.

Moreover, even for those k sets S with $\widehat{f_{str}}(S) \neq 0$, we still have $|\widehat{f_{str}}(S)| \leq 1$.

In the above example we have quantified psuedorandomness by $\|g\|_{psr} := \max_{S \subseteq [n]} |\widehat{g}(S)|$. Though, this norm may sound unfamiliar at first, in fact it has an operational meaning in terms of the amount by which g has advantage over a random function in passing the Blum-Luby-Rubinfeld linearity test.

Remark 1.2. Almost all major elements of the (classical) regularity theory is present in this example:

- (1) We had to make a choice on how to quantify psuedorandomness (this usually depends on the application one has in mind); here we did so using the maximal correlation of the function with Fourier characters. Some other typical choices over the hypercube are the maximal correlation with the set of AC0 functions, degree d -threshold functions or CNFs of certain size.
- (2) One usually also has the freedom for choosing the growth function $F : \mathbb{Z}^+ \rightarrow \mathbb{R}^+$. Choosing a faster growing function such as $F(x) = e^{cx}$ is necessary when one needs a greater control on the psuedorandom part of the function. On the other hand, the complexity bound on f_{str} (roughly speaking) behaves like $M = F(F(\dots(F(1))\dots))$ iterated $\text{poly}(1/\epsilon)$ times. Hence, the great control over the f_{psr} comes at a price.

As usual in regularity theory, the proof is a potential function argument.

Proof of Claim 1.1. List the subsets $S \subseteq [n]$ in the order of size, i.e.

$$|\widehat{f}(S_1)| \geq |\widehat{f}(S_2)| \geq |\widehat{f}(S_3)| \geq \dots$$

Start from $M_0 = 1$. Having defined M_{i-1} , we define M_i to be the minimum index $1 \leq M_i \leq 2^n$ such that $|\widehat{f}(S_{1+M_i})| \leq \frac{1}{F(M_{i-1})}$. Now notice that for $M_{i-1} + 1 \leq r \leq M_i$ we have $|\widehat{f}(S_r)| > \frac{1}{F(M_{i-1})}$. Since $\|f\|_2^2 \leq 1$, we see that $M_i - M_{i-1} \leq F(M_{i-1})^2$. Consider taking $f_{str} = \sum_{r \leq M_{i-1}} \widehat{f}(S_r) \chi_{S_r}$, $f_{sml} = \sum_{M_{i-1}+1 \leq r \leq M_i} \widehat{f}(S_r) \chi_{S_r}$ and $f_{psr} = \sum_{r \geq M_{i+1}} \widehat{f}(S_r) \chi_{S_r}$. This choice immediately satisfies property (1) and (3) of the decomposition in the claim; hence for this decomposition to fail we must have

$$\|f_{sml}\|_2^2 = \sum_{M_{i-1}+1 \leq r \leq M_i} |\widehat{f}(S_r)|^2 > \epsilon^2. \quad (1)$$

Since f_{sml} 's in the above have disjoint Fourier supports for different i 's, and $\|f\|_2^2 \leq 1$, equation (1) does not happen for some $i \leq 1/\epsilon^2$. This proves the desired result for $M = G(G(\dots G(1)\dots))$, iterated $1/\epsilon^2$ times; here $G(x) = x + F(x)^2$, and thus $M = O_{\epsilon, F}(1)$ as desired. \square

Let us note that plugging in $F(x) = \epsilon^{-2}2^{x+1}$ in Claim 1.1 and simple Fourier analysis allows one to recover the arithmetic regularity of Green [10] over \mathbb{F}_2^n . Both in the case of Green's result and strong graph regularity lemma one needs the local control offered by choice $F(x) = e^{\Theta(x)}$. In the next section, we discuss and compare the weak versus strong regularity lemmas in the context of graphs; hopefully this clarifies in which contexts one can get away with a weak regularity lemma

(which has much better bound, and can be done faster algorithmically) and when this is not possible.

2. REGULARITY LEMMA AS LOW-RANK APPROXIMATION TO MATRICES AND TENSORS

In this section we show the proof of matrix and tensor decomposition theorems due to Frieze and Kannan [7]. The proof of the matrix case and the tensor case are very similar, and in fact we state a more general theorem which gives both at the same time; this theorem is a reformulation of regularity lemma due to Tao [15]. We consider this abstract form of regularity lemma of Tao extremely powerful: for one the abstract approach has the advantage that it captures what is common among all forms of regularity lemma, hence saving time and effort spent redoing essentially the same proof.

Theorem 2.1 (Strong regularity–Tao formulation). *Let f be a vector of at most unit length in a Hilbert space H . Let $S \subset H$ be a set of structured or special vectors (elements of S are not necessarily unit norm). For any $\epsilon > 0$ and any growth function $F : \mathbb{Z}^+ \rightarrow \mathbb{R}^+$ there exists $M = O_{F,\epsilon}(1)$ and $k \leq M$ such that $f = f_{str} + f_{sml} + f_{psr}$ where $\|f_{sml}\|_H \leq \epsilon$ and f_{str} can be written as the sum at most M vector in S with coefficient at most of size M . f_{psr} is $\frac{1}{F(M)}$ -pseudorandom via the following norm:*

$$\|g\|_{psr} = \max_{u \in S} \langle u, g \rangle_H.$$

The proof of this theorem is very similar to Claim 1.1; we shall note the slight modifications to the proof of Claim 1.1 one needs in the case of Theorem 2.1 in the appendix ². The versatility of the above theorem is remarkable: Taking $F(x) = \epsilon^{-2}x$ gives us the “weak regularity lemma” of Frieze and Kannan. Plugging in $F(x) = \epsilon^{-3}4^x$ on the other hand gives the strong graph regularity lemma. Also it can easily be seen to imply Claim 1.1 which is equivalent to Green’s arithmetic regularity lemma. Let us demonstrate how one arrives at the strong and weak graph regularity from this theorem:

For simplicity assume $f : V_1 \times V_2 \rightarrow [0, 1]$ represents the edge weights of a bipartite graph. Consider $H = L^2(V_1 \times V_2)$ given by the inner product norm

$$\langle f, g \rangle_H = \mathbf{E}_{(v_1, v_2) \in (V_1, V_2)} [f(v_1, v_2)g(v_1, v_2)].$$

Take the set of structured vectors S to consist of the following $2^{|V_1|+|V_2|}$ functions which correspond to uniform bipartite graphs:

$$S = \{u \in L^2(V_1 \times V_2) : u(v_1, v_2) = 1_A(v_1)1_B(v_2) \mid A \subseteq V_1, B \subseteq V_2\}. \quad (2)$$

Notice that elements of S are not normalized, and $\|1_A 1_B\|_H = \sqrt{\frac{|A||B|}{|V_1||V_2|}}$ is much smaller for functions with smaller support. This means functions of large support tend to affect $\|\cdot\|_{psr}$ more often. Notice that

$$f_{str} = \sum_{i \leq M} c_i 1_{A_i} 1_{B_i} \quad A_i \subseteq V_1, B_i \subseteq V_2.$$

The M sets A_i ’s induce a partition \mathcal{P}_1 on the left side of the graph of size at most 2^M given by

$$\mathcal{P}_1 = \{U \subset V_1 : \exists I \subseteq [M], U = \cap_{i \in I} A_i\}.$$

²essentially the difference is akin to the difference between the orthogonal matching pursuit (OMP) with the matching pursuit. More specifically, for Theorem 2.1, in each iteration of the proof, we do a complete orthogonalization updating all the coefficients in f_{str} as opposed to the case of Claim 1.1 where once a coefficient is chosen in f_{str} it does not ever change.

Similarly, one gets a partition \mathcal{P}_2 of size at most 2^M on V_2 . The “atoms” arising in \mathcal{P}_1 and \mathcal{P}_2 are exactly the partition given by the classical regularity lemma ³.

Let us now mention how the properties of the partition induced by $(\mathcal{P}_1, \mathcal{P}_2)$ on $V_1 \times V_2$ arises from the properties of our decomposition. For that we need a definition.

Definition 2.2 (regularity). Let $g : U_1 \times U_2 \rightarrow \mathbb{R}$ be a function. g is called ϵ -regular (or quasirandom) if

$$\mathbf{E}_{(x,y) \in U_1 \times U_2} [g(x,y)u(x)v(y)] \leq \epsilon$$

for any choice of $u : V_1 \rightarrow [-1, 1]$ and $v : V_2 \rightarrow [-1, 1]$. Here $\mathbf{E}[\cdot]$ is with respect to uniform probability measure on $V_1 \times V_2$.

One can see that the following holds:

Fact 2.3 (Equivalence of different notions of regularity). *Let $g : U_1 \times U_2 \rightarrow \mathbb{R}$ be a ϵ -regular function. For any $U'_1 \subseteq U_1$, $U'_2 \subseteq U_2$ with $|U'_i| \geq \epsilon^{1/3}|U_i|$ for $i = 1, 2$, we have*

$$\left| \sum_{x \in U'_1, y \in U'_2} f(x,y) \right| \leq \epsilon^{1/3}|U'_1||U'_2|.$$

Now finally we can see the implications of the decomposition $f = f_{str} + f_{sml} + f_{psr} = f_{str} + f_{err}$ guaranteed by Theorem 2.1 on the partition $(\mathcal{P}_1, \mathcal{P}_2)$ arising on $V_1 \times V_2$. The following shows that we have a great control on the resulting “graph” f_{err} coming from taking the difference between f and its approximation by f_{str} . The main point is that (after some refinement and garbage collection) we can assume all elements in $U \in \mathcal{P}_1$ and $R \in \mathcal{P}_2$ satisfy $|U| = \Omega(2^{-M}|V_1|)$ and $|R| = \Omega(2^{-M}|V_2|)$ which is the bound we expect because there are at most 2^M atomic sets induced by A_i 's and B_i 's. Now the fact is that by the choice of $F(\cdot)$ we know f_{psr} is $\epsilon^{-3}4^M$ -psuedorandom. This gives us a great local type of control: let us imagine we zoom in induced subgraph over (U, R) . Since we have zoomed in by a factor 4^M , our psuedorandomness parameter gets worsen by a factor 4^M on the graph induced on (U, R) considered in isolation. However, the fact that $F(x) = \epsilon^{34^x}$ still means that even this local subgraph induced on (U, R) by f_{psr} is ϵ -regular in the sense of Szemerédi even under this magnified resolution (the transition from ϵ^3 to ϵ is because of Fact 2.3). The effects of f_{sml} part of f_{err} results in the following:

(*) The bound $\|f_{sml}\|_2 \leq \epsilon$ exactly translates to the fact that all but a small minority of atomic paris U and U' arising in our regularity partition are 2ϵ -irregular:

$$\sum_{(U,R): 2\epsilon\text{-irregular}} |U||R| \leq \epsilon^2|V_1||V_2|.$$

Tensor decompositions. So far we have seen the power of Tao’s formulation of regularity theory in redriving various known regularity statements (weak, strong, arithmetic) uniformly and rather easily. In most applications one just needs to adjust the definition of the set S to derive the desired result. To derive the low rank tensor decomposition of Frieze-Kannan, one work a Hilbert space $H = L^2(V_1 \times V_2 \times \dots \times V_r)$ with inner product induced by uniform probability measure on $V_1 \times \dots \times V_r$. and the set S of “generalized cut function” (or cylinder functions):

$$S = \{u \in H : u(x_1, x_2, \dots, x_r) = 1_{A_1}(x_1)1_{A_2}(x_2) \dots 1_{A_r}(x_r), A_i \subseteq V_i\}.$$

³Further refinement to make sure all the atoms are of equal size can be achieved by throwing atomic sets that are to small into a garbage set, and then refining the partitions of \mathcal{P}_i to equal parts with the hall of the garbage set. But these technicalities do not pose a real threat.

The pseudorandomness norm induced by S above given by $\|g\|_{psr} = \max_{u \in S} \langle u, g \rangle$ is usually called the “cut-norm” in the literature. Applying Theorem 2.1 to the above choice of (H, S) leads to F-K tensor decomposition:

Theorem 2.4 (F-K). *Let $\epsilon > 0$ and let $A : V_1 \times V_2 \times \dots \times V_r \rightarrow \mathbb{R}$ be a tensor. There exists $t \leq \frac{1}{\epsilon^2}$ such that*

$$\left\| A - \sum_{i=1}^t d_i u_i \right\|_{cut} \leq \epsilon \|A\|_F \leq \epsilon \|A\|_\infty,$$

where $\|A\|_F = (\mathbf{E}_{x \in V_1 \times \dots \times V_r} A(x_1, x_2, \dots, x_r)^2)^{1/2}$, and all u_i 's are cut functions and d_i 's are scalar coefficients. Moreover, $\sum_{i=1}^t d_i^2 = O(\|A\|_F^2)$.

3. ALGORITHMIC FORMS AND CONSTRAINT SATISFACTION PROBLEMS

The basic implication of regularity theory to constraint satisfaction is more or less clear: let us for simplicity work in the setting of MAX-3SAT. We write the objective function of our CSP as a polynomial over variables $x_1, x_2, \dots, x_n, y_1, y_2, \dots, y_n$ where $y_i = 1 - x_i$. Let us denote $x_{n+i} = y_i$. Thus, we have a polynomial optimization problem over Boolean variables $\{x_i\}$'s:

$$\text{CSP-Val} = \max_{\substack{x \in \{0,1\}^{2n} \\ x_i = 1 - x_{n+i}}} p(x), \quad p(x) = \sum_{(i,j,k) \in [2n] \times [2n] \times [2n]} A(i, j, k) x_i x_j x_k.$$

Assuming the instance is dense, and all constraints have equal worth, we see that $\|A\|_\infty = O(1)$ and $\|A\|_F = \Omega(1)$. Assuming we have an algorithmic form of Frieze-Kannan Theorem, we can reduce our problem to a much simpler problem of bounded complexity. Notice that $p(x)$ is evaluated on Boolean points, hence the approximation in Theorem 2.4 guarantees that we make at most $\epsilon \|A\|_\infty n^r$ in our approximation by reducing our problem to the A_{str} . This gives a PTAS, because we know $\text{CSP-VAL} = \Omega(n^r)$; so by choosing ϵ sufficiently small we can reduce the error incurred by replacing A by A_{str} in the optimization problem to get as close as desired to the value of OPT. Now we have to deal with the following issues:

- (1) What does it take to make Theorem 2.1 algorithmic? In particular we need an algorithmic form of the weak regularity of Frieze Kannan, Theorem 2.4.
- (2) If we are just content with a PTAS, the above plus the algorithmic weak regularity is sufficient for us; because even enumerating over all possible subdivision of atoms gives a bound of type $2^{2^{\Theta(1/\epsilon^2)}}$. But can we improve on this?

The main point for making Theorem 2.1 algorithmic is that we need to be able to find a $u \in S$ such that $\langle u, f \rangle_H \geq \epsilon$, or prove that such u does not exist. Even when S is a very nice set, this problem is usually hard, but the next best thing is usually possible: in the Frieze-Kannan setting if there exists $u \in S$ with $\langle u, f \rangle \geq \epsilon$ we can find some $u' \in S$ with $\langle u', f \rangle \geq c\epsilon^3$ for some constant $c > 0$. If for all $u \in S$ we have $\langle u, f \rangle \leq \epsilon^3 c$ we can prove that (In the intermediate cases we might do one or the other). This means the number of iterations of the algorithm might increase from $O(1/\epsilon^2)$ to $O(1/\epsilon^6)$, but otherwise there is no other problem.

For the cut-matrix case, we can even do better because as Alon and Naor [4] observed the Grothendieck's inequality from functional analysis can become algorithmic to give a constant factor approximation for the cut-norm. For the tensor case, there seem to be no known constant factor approximation for the cut-norm. But a polynomial time algorithm is possible via a generalized version of the following :

Definition 3.1. Let $f : V_1 \times V_2 \rightarrow \mathbb{R}$. For $Q_1 \subseteq V_1$ we define $P(Q_1) \subseteq V_2$ by $P(Q_1) = \{x \in V_2 : A(x, y) > 0 \forall y \in Q_1\}$. $P(Q_2)$ for $Q_2 \subseteq V_2$ is defined analogously.

Theorem 3.2 (see [2]). *For appropriate chosen polynomial $q(\cdot)$, if $Q_1 \subseteq V_1$ and $Q_2 \subseteq V_2$ are chosen (with replacement) as subsets of size $q(1/\epsilon, 1/\delta)$, then with probability $1 - \delta$ there exists $Q'_1 \subseteq Q_1$ and $Q'_2 \subseteq Q_2$ such that*

$$\langle A, 1_{P(Q'_1)} 1_{P(Q'_2)} \rangle = \Omega(\|A\|_{cut}) - \frac{\epsilon}{2} \|A\|_F.$$

Finally, we note that improving the item (2) above from $2^{2^{\Theta(1/\epsilon^2)}}$ to $2^{\Theta(1/\epsilon^2)}$ is possible by noticing that one does not have to go down to the atomic representation; see [2].

4. SUMMARY AND FUTURE DIRECTIONS

In this work, we tried to unify various aspects of regularity theory under the framework advocated by Tao [15]. For example, we saw many different forms of regularity statements can be seen an application of Theorem 2.1 applied to various type of structured set S . Another advantage of Tao's framework is to remedy the strange situation where one had two forms of regularity lemma in the literature, weak and strong, with no clear way to tradeoff between them. Given his formulation, it is clear that there is indeed a spectrum of intermediate regularity lemmas available in between the two. We also briefly mentioned how one goes about making the steps in regularity lemma algorithmic.

What's next? The main conceptual consequence of regularity theory for computer science is the fact that dense MAX-rCSP problems (i.e. with $\Omega(n^r)$ constraints) admit a polynomial time approximation scheme. It is natural to ask whether regularity theory offers any algorithmic implication in the sparse setting. 10 years ago, this would have been a very strange thing to consider, but since the appearance of Green-Tao theorem [11], it has become apparent that under some suitable conditions regularity theorem *transfer* to the sparse setting.

Dense model philosophy: Let $H \subset [n]$ be a sparse but strongly pseudorandom subset of integer in a suitable sense. Then the additive behavior of a dense (relative to $|H|$) subset $A \subset H$ can be modeled by a dense subset A' of $[n]$. Hence, Szemerédi's theorem on the existence of long arithmetic progression of dense subsets of $[n]$ for large n applies, via the transference principle, equally well to the setting of dense subsets of H .

A rigorous form of the above statement was proved by Green-Tao allowing them to establish their result on arithmetic progression on primes. Such results belong to the expanding area of relative Szemerédi theory [8, 6]. There perhaps may be further applications of this theory to constraint satisfaction problems.

Problem 4.1. *Is it possible to apply an algorithmic form of the relative Szemerédi theory to obtain interesting results in the sparse constraint satisfaction setting (i.e. when the interaction graph of the CSP is dense in some host graph with great pseudorandom properties)?*

Let us note that a somewhat related line of research in the sprit of identifying easy instances of UG and partitioning problems has seen some advances; see [5, 12, 13].

REFERENCES

- [1] N. ALON, *The algorithmic aspects of the regularity lemma*, in FOCS, 1992.
- [2] N. ALON, W. F. DE LA VEGA, R. KANNAN, AND M. KARPINSKI, *Random sampling and approximation of max-csps*, Journal of Computer and System Sciences, 67 (2003).
- [3] N. ALON, E. FISCHER, I. NEWMAN, AND A. SHAPIRA, *A combinatorial characterization of the testable graph properties: it's all about regularity*, SIAM Journal on Computing, 39 (2009), pp. 143–167.
- [4] N. ALON AND A. NAOR, *Approximating the cut-norm via grothendieck's inequality*, SIAM Journal on Computing, 35 (2006), pp. 787–803.
- [5] Y. BILU AND N. LINIAL, *Are stable instances easy*, in 1st Symposium on Innovations in Computer Science (ITCS), 2010.
- [6] D. CONLON, J. FOX, AND Y. ZHAO, *A relative szemerédi theorem*, arXiv preprint arXiv:1305.5440, (2013).
- [7] A. FRIEZE AND R. KANNAN, *Quick approximation to matrices and applications*, Combinatorica, 19 (1999).
- [8] W. GOWERS, *Decompositions, approximate structure, transference, and the hahn-banach theorem*, Bulletin of the London Mathematical Society, 42 (2010), pp. 573–606.
- [9] W. T. GOWERS, *Quasirandomness, counting and regularity for 3-uniform hypergraphs*, Combinatorics, Probability & Computing, 15 (2006).
- [10] B. GREEN, *A szemerédi-type regularity lemma in abelian groups, with applications*, Geometric & Functional Analysis GAFA, 15 (2005), pp. 340–376.
- [11] B. GREEN AND T. TAO, *The primes contain arbitrarily long arithmetic progressions*, Annals of mathematics, 167 (2008), pp. 481–547.
- [12] A. KOLLA, K. MAKARYCHEV, AND Y. MAKARYCHEV, *How to play unique games against a semi-random adversary: Study of semi-random models of unique games*, in Foundations of Computer Science (FOCS), 2011 IEEE 52nd Annual Symposium on, IEEE, 2011, pp. 443–452.
- [13] K. MAKARYCHEV, Y. MAKARYCHEV, AND A. VIJAYARAGHAVAN, *Approximation algorithms for semi-random partitioning problems*, in Proceedings of the 44th symposium on Theory of Computing, ACM, 2012, pp. 367–384.
- [14] V. RÖDL, B. NAGLE, J. SKOKAN, M. SCHACHT, AND Y. KOHAYAKAWA, *The hypergraph regularity method and its applications*, Proceedings of the National Academy of Sciences of the United States of America, 102 (2005), pp. 8109–8113.
- [15] T. TAO, *Structure and randomness in combinatorics*, in FOCS, 2007.

A NOTE ABOUT THE PROOF OF THEOREM 2.1.

Proof of Theorem 2.1 is quite similar to that of Claim 1.1, and hence we shall not repeat all the details. Essentially the same iterative procedure works using $\|f\|_H$ as the potential function; the only difference is that in the case of Claim 1.1 the set of structured vectors S consisted of an orthogonal set of vectors in H . As a result the bound $|c_i| \leq M$ in the expansion

$$f_{str} = \sum_{i \leq M} c_i g_i \quad , \quad \{g_i\} \subseteq S,$$

came for free. This in general is not the case, e.g. in the graph case the cut functions $S = \{1_A 1_B\}_{A \subseteq V_1, B \subseteq V_2}$ form an overcomplete basis for $H = L^2(V_1 \times V_2)$. For things to work out more smoothly in the non-orthogonal case, we change our update procedure: when we find a new vector $u_{new} \in S$ with large $\langle f_{i-1}, u_{new} \rangle$, instead of updating by $f_i \leftarrow f_{i-1} - \frac{\langle f_{i-1}, u_{new} \rangle}{\|u_{new}\|_H^2} u_{new}$, we do a full orthogonal projection. With this change, things works out without much difficulty. However, now the resulting coefficient in c_i 's in the expansion of f_{str} may grow as opposed to remaining uniformly bounded (but, they still remain bounded by M ; moreover, a slightly more careful argument can avoid this by instead working with Tao's measure theoretic version of regularity).

WHY WE NEED STRONG REGULARITY?

Given the fact one can go quite far with the weak regularity in many applications, say in dense constraint satisfactions, one may wonder why we even need the “strong regularity lemma”. It

turns out the type of local control alluded to in the discussion after Fact 2.3 is necessary both in applications to property testing and also in extremal graph theory. The weak regularity lemma seem simply not strong enough to conclude from the fact that G has few triangles the fact that G is in Hamming distance close to a graph G' completely devoid of triangles. The fact is that weak regularity may allow you to decrease the triangles but making it completely devoid of triangles needs a type of local control which seem to be offered by the strong version not by the weak version.

One possible application of the new development in hypergraph regularity to computer science is to characterize testable hypergraph properties in the spirit of Alon et al [3]. This direction has the advantage that there is almost guaranteed progress one can make; however, it may be a less exciting direction theoretically.