Electronic Colloquium on Computational Complexity, Report No. 45 (2008)

Dense Subsets of Pseudorandom Sets

Omer Reingold* Luca Trevisan† Madhur Tulsiani‡ Salil Vadhan§

April 23, 2008

Abstract

A theorem of Green, Tao, and Ziegler can be stated (roughly) as follows: if R is a pseudorandom set, and D is a dense subset of R, then D may be modeled by a set M that is dense in the entire domain such that D and M are indistinguishable. (The precise statement refers to"measures" or distributions rather than sets.) The proof of this theorem is very general, and it applies to notions of pseudorandomness and indistinguishability defined in terms of any family of distinguishers with some mild closure properties. The proof proceeds via iterative partitioning and an energy increment argument, in the spirit of the proof of the weak Szemerédi regularity lemma. The "reduction" involved in the proof has exponential complexity in the distinguishing probability.

We present a new proof inspired by Nisan's proof of Impagliazzo's hardcore set theorem. The reduction in our proof has polynomial complexity in the distinguishing probability and provides a new characterization of the notion of "pseudoentropy" of a distribution.

We also follow the connection between the two theorems and obtain a new proof of Impagliazzo's hardcore set theorem via iterative partitioning and energy increment. While our reduction has exponential complexity in some parameters, it has the advantage that the hardcore set is efficiently recognizable.

Keywords: pseudorandomness, additive combinatorics, regularity lemmas, pseudoentropy

*Faculty of Mathematics and Computer Science, Weizmann Institute of Science, Rehovot 76100, Israel. omer.reingold@weizmann.ac.il Research supported by US-Israel Binational Science Foundation Grant 2006060.

†Computer Science Division, U.C. Berkeley. luca@cs.berkeley.edu Work partly done while visiting Princeton University and the IAS. This material is based upon work supported by the National Science Foundation under grants CCF-0515231 and CCF-0729137 and by the US-Israel Binational Science Foundation grant 2006060.

‡Computer Science Division, U.C. Berkeley. madhurt@cs.berkeley.edu Work partly done while visiting Princeton University. This material is based upon work supported by the National Science Foundation under grants CCF-0515231 and CCF-0729137 and by the US-Israel Binational Science Foundation grant 2006060.

§School of Engineering and Applied Sciences, Harvard University. salil@eecs.harvard.edu. Work done during a visit to U.C. Berkeley, supported by the Miller Foundation for Basic Research in Science and a Guggenheim Fellowship. This materials is also based on work supported by US-Israel Binational Science Foundation under grant 2006060, and the Office of Naval Research under grant N00014-04-1-0478.

ISSN 1433-8092

1 Introduction

Green and Tao [GT], in one of the great mathematical breakthroughs of this decade, have proved that there exist arbitrarily long arithmetic progressions of primes. Somewhat imprecisely, their proof proceeds by establishing the following two claims:

- Let R be a "pseudorandom" set of integers, and D be a subset of R of constant density. Then D contains arbitrarily long arithmetic progressions.
- There is a set R of integers that is pseudorandom and such that the primes have constant density inside R.

The first claim is the hardest to establish, and its proof is the most innovative part of the paper, blending combinatorial, analytic and ergodic-theoretic techniques. In turn (and, again, this account is slightly imprecise), the proof of the first claim proceeds by combining the following three results.

- **Dense Model Theorem:** Let R be pseudorandom and D a subset of R of constant density within R (both R and D may be very sparse within the integers). Then there is a set M that has constant density within the integers and is "indistinguishable" from D. (We think of M as a dense "model" of D.)
- **Szemerédi's Theorem [Sze]:** If M is a set of integers of constant density, then it contains a constant fraction of all arithmetic progressions of any fixed length.
- **Lemma:** A set that contains a constant fraction of all arithmetic progressions of some fixed length k is "distinguishable" from a set with no arithmetic progressions of length k

The key new step of the Green–Tao proof is their Dense Model Theorem. This theorem about dense subsets of pseudorandom sets was originally stated in the specific setting of sets of integers and for certain specific notions of pseudorandomness and indistinguishability. It is natural to ask if a similar statement holds when we consider other domains, like $\{0,1\}^n$, and for other notions of pseudorandomness and indistinguishability such as those used in complexity theory and foundations of cryptography. A very general Dense Model Theorem, which has a complexity-theoretic version as a special case, was in fact proven by Tao and Ziegler [TZ]. However, the "reduction" implicit in their proof has exponential complexity in the distinguishing probability, making it inapplicable for common complexity-theoretic or cryptographic settings of parameters.

In this paper, we provide a new proof of the Dense Model Theorem, in which the reduction has polynomial complexity in the distinguishing probability. Our proof is inspired by Nisan's proof of the Impagliazzo Hardcore Theorem [Imp], and is simpler than the proofs of Green, Tao, and Ziegler. A complexity-theoretic interpretation of our result yields a new characterization of the "pseudoentropy" of a distribution. We also exploit the connection between the two theorems in the reverse direction to obtain a new proof of the Hardcore Theorem based on iterative partitioning and energy increments. While the reduction in this proof has exponential complexity in some parameters, it has the advantage that the hardcore set is efficiently recognizable.

We find it intriguing that there is such an intimate connection between ideas in the additive combinatorics literature and such central complexity-theoretic concepts as pseudorandomness and indistinguishability. The fact that we can translate the proofs in both directions, obtaining some new properties in each case, suggests that both complexity theory and additive combinatorics are likely to benefit from this connection in the future.

1.1 Dense Model Theorems

Let us briefly recall how we define pseudorandomness and indistinguishability in complexity theory (in the non-uniform setting). We have a finite domain X, for example $\{0,1\}^n$, and a collection $\mathcal{F}$ of "efficiently computable" functions $f : X \to \{0,1\}$, for example all the functions computed by circuits of size at most $s(n)$ for some complexity bound $s(\cdot)$. We say that a distribution R on X is ϵ-pseudorandom for $\mathcal{F}$ if for every function $f \in \mathcal{F}$ we have

$$|\mathbb{P}[f(R) = 1] - \mathbb{P}[f(U_X) = 1]| \leq \epsilon$$

where U_X is the uniform distribution over X.[1] More generally, we say that two distributions A and B are ϵ-indistinguishable by $\mathcal{F}$ if for every $f \in \mathcal{F}$

$$|\mathbb{P}[f(A) = 1] - \mathbb{P}[f(B) = 1]| \leq \epsilon$$

We also need to specify what "density" means when we refer to distributions rather than to sets. We say that a distribution A is δ-dense in B if, informally, it is possible to describe the process of sampling from B as "with probability δ, sample from A, with probability $1-\delta$, $(\dots)$" which is equivalent to the condition

$$\forall x \in X,\ \mathbb{P}[A = x] \leq \frac{1}{\delta} \cdot \mathbb{P}[B = x]$$

Given these definitions, an general Dense Model Theorem would have the following form: Let X be a finite domain, $\mathcal{F}$ a collection of boolean (or bounded) functions on X, and $\epsilon, \delta > 0$ be real parameters. Then there exists an $\varepsilon' > 0$ and a collection $\mathcal{F}'$ of boolean functions on X such that if R is ϵ'-pseudorandom for $\mathcal{F}'$ and D is δ-dense in R, then there is a *model distribution* M that is δ-dense in U_X and that is ϵ-indistinguishable from D for $\mathcal{F}$. Ideally, ϵ' should not be too much smaller than ϵ, and the functions in $\mathcal{F}'$ should not be too much more "complex" than functions in $\mathcal{F}$. Indeed, in a complexity-theoretic setting, we'd like both of these relations to be polynomial so that the distinctions disappear when we consider asymptotic formulations with $1/\mathrm{poly}(n)$ distinguishing probabilities and functions computed by polynomial-size circuits.

Tao and Ziegler [TZ] have proved such a result in broad generality, albeit with an exponential loss in the distinguishing probability. Formally, their theorem can be restated as as follows.

Theorem 1.1 (Tao and Ziegler) *Let X be a finite universe, $\mathcal{F}$ a collection of bounded functions $f : X \to [0,1]$, $\epsilon > 0$ an accuracy parameter and $\delta > 0$ a density parameter. Let R be a distribution over X and D a δ-dense distribution in R.*

Suppose that D is distinguishable from all dense models. That is, suppose that for every model distribution M that is $\delta/2$-dense in U_X, there is a function $f \in F$ such that

$$|\mathbb{E}[f(D)] - \mathbb{E}[f(M)]| \geq \epsilon$$

Then R is not pseudorandom. That is, there are functions $f_1, \dots, f_k$ in $\mathcal{F}$, with $k = \mathrm{poly}(1/\epsilon, 1/\delta)$ such that

$$\left| \mathbb{E}\left[\prod_i f_i(R)\right] - \mathbb{E}\left[\prod_i f_i(U_X)\right] \right| \geq \exp(-\mathrm{poly}(1/\epsilon, 1/\delta))$$

[1] In the above expression, and in the rest of the paper, we use the same notation for a distribution D over a sample space X, and for a random variable ranging over X and taking on values of X according to D.

Theorem 1.1 is a restatement of Theorem 7.1 in [TZ].[2] To match it with the discussion above, take $\mathcal{F}'$ to be the set of functions that are k-fold products of functions in $\mathcal{F}$, and $\epsilon' = \exp(-\text{poly}(1/\epsilon, 1/\delta))$.

Theorem 1.1 can be applied to a computational setting where $\mathcal{F}$ contains only Boolean functions, hence $\mathbb{E}[f(A)] = \mathbb{P}[f(A) = 1]$ for every distribution A. In such a setting the theorem does imply that if a distribution D is δ-dense in a distribution R that is ϵ'-pseudorandom for circuits of size s', then D is ϵ-indistinguishable for circuits of size s from some distribution M that is $\delta/2$-dense in the uniform distribution, where $\epsilon' = \exp(-\text{poly}(1/\epsilon, 1/\delta))$ and $s' = s \cdot \text{poly}(1/\epsilon, 1/\delta)$. The exponentially small bound on the distinguishing probability ϵ' for R, however, is unsuitable for typical complexity-theoretic and cryptographic settings that consider distinguishing probabilities of $1/\text{poly}(n)$ (where $X = \{0,1\}^n$). Reading into the Tao-Ziegler proof and specializing it to the Boolean setting, it is possible to improve the bound on ϵ' to polynomial and derive the following statement.

Theorem 1.2 (Tao and Ziegler – Boolean case) *Let X be a finite universe, $\mathcal{F}$ a collection of Boolean functions $f : X \to \{0,1\}$, $\epsilon \in (0, 1/2)$ an accuracy parameter and $\delta \in (0, 1/2)$ a density parameter. Let R be a distribution over X and D a δ-dense distribution in R.*

Suppose that D is distinguishable from all dense models. That is, suppose that for every model distribution M that is $\delta/4$-dense in U_X there is a function $f \in F$ such that

$$|\,\mathbb{E}[f(D)] - \mathbb{E}[f(M)]| \geq \epsilon$$

Then R is not pseudorandom. That is, there are functions $f_1, \ldots, f_k$ in $\mathcal{F}$, with $k = \text{poly}(1/\epsilon, 1/\delta)$, and $g : \{0,1\}^k \to \{0,1\}$ such that if we define $h(x) := g(f_1(x), \ldots, f_k(x))$ we have

$$|\,\mathbb{E}[h(R)] - \mathbb{E}[h(U_X)]| \geq (\epsilon\delta)^{O(1)}$$

It seems that, in such a statement, everything has polynomial efficiency as required, but unfortunately the function g in the conclusion can be arbitrary. In particular, its circuit complexity cannot be bounded any better than by an exponential in k, and hence exponential in $1/\epsilon$ and $1/\delta$. The conclusion that we can derive is that if a distribution D is δ-dense in a distribution R that is ϵ'-pseudorandom for circuits of size s', then D is ϵ-indistinguishable from a distribution $\delta/4$-dense in the uniform distribution by circuits of size s, where $\epsilon' = (\epsilon\delta)^{O(1)}$ and $s' = s \cdot \text{poly}(1/\epsilon, 1/\delta) + \exp(\text{poly}(1/\epsilon, 1/\delta))$.

In this paper we present a new proof of a Dense Model Theorem, in the spirit of Nisan's proof of the Impagliazzo Hardcore Theorem [Imp], where all parameters are polynomially bounded. The key change will be that the combining function g will be a linear threshold function, and hence can be implemented by a circuit of size $O(k)$.

Theorem 1.3 (Main) *Let X be a finite universe, $\mathcal{F}$ a collection of Boolean functions $f : X \to \{0,1\}$, $\epsilon > 0$ an accuracy parameter and $\delta > 0$ a density parameter. Let R be a distribution over X and D a δ-dense distribution in R.*

[2] The two statements of the theorem are completely equivalent, with the following translation. Our functions f are called *dual functions* in [TZ], where they are allowed to range over a bounded interval instead of $[0,1]$, but one can restrict to $[0,1]$ with no loss of generality after scaling. Our distribution R plays the role of the measure ν in the Tao–Ziegler formulation, under the normalization $\mathbb{P}[R = a] = \nu(a)/\sum_z \nu(z)$. Our distribution D is their measure $g()$ after the normalization $\mathbb{P}[D = a] = g(a)/\sum_z g(z)$. Our distribution M is their function g_1, after similar normalization, and their g_2 equals $g - g_1$. This translation applies if $\mathbb{E}[g(U_X)] \geq \delta$, but the general case reduces to the case of g having sufficiently large average; otherwise, we can simply set their g_1 and g_2 to be identically zero.

Suppose that D is distinguishable from all dense models. That is, suppose that for every model distribution M that is δ-dense in U_X there is a function $f \in F$ such that

$$|\mathbb{E}[f(D)] - \mathbb{E}[f(M)]| \geq \epsilon$$

Then R is not pseudorandom. That is, there are functions $f_1, \ldots, f_k$ in $\mathcal{F}$, with $k = \mathrm{poly}(1/\epsilon, \log 1/\delta)$, and a linear threshold function *$g : \{0,1\}^k \to \{0,1\}$ such that if we define $h(x) := g(f_1(x), \ldots, f_k(x))$ we have*

$$|\,\mathbb{E}[h(R)] - \mathbb{E}[h(U_X)]| \geq \Omega(\epsilon\delta)$$

Our proof can also recover Theorem 1.1 in full generality. When we apply our proof to the setting of Theorem 1.1 (where we require the distinguishing function to be a product of f_i rather than a low-complexity combination of f_i) we too incur an exponential loss in the distinguishing probability, but our proof is simpler than the original proof of Tao and Ziegler.

Gowers [Gow] independently discovered a simplified proof of Theorem 1.1 that is similar to ours.

1.2 Applications

The min-entropy of a distribution D is defined as $\mathrm{H}_\infty(D) := \min_a \log(1/\mathbb{P}[D = a])$, and it can be seen that a distribution D ranging over $\{0,1\}^n$ has min-entropy at least $n - t$ if and only if it is 2^{-t}-dense in the uniform distribution. Following Håstad et al. [HILL], we say that a distribution has pseudoentropy at least k if it is computationally indistinguishable from some distribution of min-entropy at least k.[3] It follows from our main theorem that if a distribution is 2^{-t}-dense inside a pseudorandom distribution then it has pseudoentropy at least $n - t$, provided $\delta = 2^{-t}$ is nonnegligible (i.e. $t = O(\log n)$ when considering $1/\mathrm{poly}(n)$ distinguishing probabilities). The converse can also be easily seen to be true, and thus our main result characterizes pseudoentropy in terms of density in pseudorandom distributions.

An example of this application is the following. Suppose that $G : \{0,1\}^m \to \{0,1\}^n$ is a good pseudorandom generator, and that B is a *biased*, adversarially chosen, distribution over seeds, about which we do not know anything except that its min-entropy is at least $m - t$. Then it is not possible any more to guarantee that the output of $G(B)$ is pseudorandom. In fact, if 2^{-t} is negligible (i.e. smaller than the distinguishing probability) then it is possible that $G(B)$ is constant.Our main result, however, implies that if 2^{-t} is nonnegligible then there is a distribution M of min-entropy at least $n-t$ such that $G(B)$ and M are indistinguishable. This application works most naturally in the non-uniform setting, where we take $\mathcal{F}$ to be the set of functions computable by bounded size circuits, but using ideas of Barak, Shaltiel, and Wigderson [BSW] we can show that a distribution dense inside a pseudorandom distribution must have large pseudoentropy even in the uniform setting.

The versatility of the Tao-Ziegler result and ours seems to go even beyond number theory and complexity theory, and it seems likely that more applications will be found. As an illustration, we describe a corollary in graph theory. Consider the case where X is the set of edges of the complete graph K_n; we think of a distribution over X as a (scaled) weighted graph, and we let $\mathcal{F}$ be the set of predicates that check whether a given edge belongs to a particular cut. In this set-up, two graphs

[3] Håstad et al. actually only require that the distribution is computationally indistinguishable from some distribution with *Shannon entropy* at least k, but it is common to work with min-entropy instead. Indeed, even the constructions of Håstad et al. work by first converting Shannon entropy into min-entropy by taking many independent copies of the distribution.

are "indistinguishable" if every cut is crossed by approximately the same fraction of edges, and a graph is "pseudorandom" if it obeys an expansion-like property. The Tao-Ziegler result thus shows that a dense subgraph of an expander is "modeled" by a truly dense graph. This is interesting because, for example, by applying the Szemerédi Regularity Lemma to the model one can recover known Regularity Lemmas for dense subgraphs of expanders [KR].[4]

1.3 The Green–Tao–Ziegler Proof, and a New Construction of Hardcore Sets

The original proofs by Green, Tao and Ziegler [GT, TZ] are based iteratively constructing a partition of X so that D is "regular" with respect to the partition. (Very roughly speaking, the condition is that for most blocks, D conditioned on the block is indistinguishable from the uniform distribution on the block.) As in the proof of the Szemerédi Regularity Lemma, one starts from the trivial one-block partition and then, as long as the partition is not regular, one uses a "counter-example" to the regularity condition to refine the partition. A potential function (or "energy increment" in the finitary ergodic-theoretic language used by Green, Tao and Ziegler) argument is used to bound the number of steps that such a process can take until it terminates.

It is intriguing that such a technique can prove a result like Theorem 1.2, which is genuinely complexity-theoretic, and we believe it could be useful in other settings as well. As a proof of concept, we provide a new proof the Impagliazzo Hardcore [Imp] using these techniques. While our proof incurs an exponential loss in terms of one of the parameters, the proof gives a "constructive" statement that does not seem to follow from other approaches.

Informally, the Hardcore Theorem says that if a function g is mildly hard to compute in the sense that every efficient algorithm errs on a noticeable fraction of inputs, then there is a relatively large 'hardcore' set H of inputs on which g is very hard to compute. We prove the following version of this theorem. Suppose that every efficient algorithm fails in computing g on at least a δ fraction of inputs. Then there is an efficiently recognizable set H of density at least δ such that $\delta/2 \leq \mathbb{P}[g(U_H) = 1] \leq 1 - \delta/2$, and it is intractable to have advantage ϵ over a constant predictor in computing g on H. This is true for every ϵ and δ, but the relation between the notions of "efficiency" in the premise and the conclusion depends exponentially on $1/\epsilon$ and $1/\delta$.

In Impagliazzo's proof, the relation is polynomial in $1/\epsilon$ and $1/\delta$ and g is nearly balanced on H, meaning that is intractable to compute g any more reliably than by making a uniform random guess. The efficient recognizability of the set H, however, is new, and it is a property that is incompatible with the requirement of being balanced.

2 Proof of the Main Theorem

In this section we prove the following result, which is a common generalization of our Main Theorem 1.3 and of the Tao-Ziegler Theorem 1.1

[4] We discuss this application purely as an illustration of the generality of the principle that "dense subsets of pseudorandom objects have a dense model," but we make no new claim. As mentioned, Regularity Lemmas for dense subsets of pseudorandom graphs were known, due to Kohayakawa and Rödl (see [KR]); furthermore the Tao-Ziegler result is used in this application, and our stronger version does not seem to offer an improvement; finally, the connection between Green-Tao-Ziegler style arguments and Regularity Lemmas is well known in the additive combinatorics community.

Theorem 2.1 *Let X be a finite universe, $\mathcal{F}$ a collection of bounded functions $f : X \to [0,1]$, $\epsilon > 0$ an accuracy parameter and $\delta > 0$ a density parameter. Let R be a distribution over X and D be a δ-dense distribution in R.*

Suppose that D is distinguishable from all dense models. That is, suppose that for every model distribution M that is δ-dense in U_X there is a function $f \in \mathcal{F}$ such that

$$|\mathbb{E}[f(D)] - \mathbb{E}[f(M)]| \geq \epsilon \tag{1}$$

Then R is not pseudorandom. That is,

1. *There are functions $f_1, \ldots, f_k \in \mathcal{F}$, with $k = O((1/\epsilon^2) \cdot \log(1/\epsilon\delta))$, and parameters $a_1, \ldots, a_k \in \{-1, +1\}$ and $t \in \mathbb{R}$ such that if we define $h : X \to \{0,1\}$ by*

$$h(x) = 1 \Leftrightarrow \sum_i a_i f(x_i) \geq t,$$

   then we have

$$|\mathbb{E}[h(R)] - \mathbb{E}[h(U_X)]| \geq \Omega(\epsilon\delta)$$

2. *There are functions $f_1, \ldots, f_k \in \mathcal{F}$, with $k = \mathrm{poly}(1/\epsilon, 1/\delta)$, such that if we define $h : X \to \{0,1\}$ by $h(x) := \Pi_i f_i(x)$ we have*

$$|\mathbb{E}[h(R)] - \mathbb{E}[h(U_X)]| \geq \exp(-\mathrm{poly}(1/\epsilon, 1/\delta))$$

Proof: For a function $f : X \to [0,1]$, we define its complement to be the function $1 - f$ and its negation to be the function $-f$, and we let $1 - \mathcal{F}$ (resp., $-\mathcal{F}$) be the set of complements (resp., negations) of functions in $\mathcal{F}$. Observe that if allow f to range over $\mathcal{F} \cup (1 - \mathcal{F})$, we may remove the absolute value in (1).

Intuition. Consider any distribution M that is δ-dense in U_X. By hypothesis, there is a function $f \in \mathcal{F} \cup (1 - \mathcal{F})$ such that $\mathbb{E}[f(D)] - \mathbb{E}[f(M)] \geq \epsilon$. One may hope that f also distinguishes between R and U_X. Since D and M are δ-dense, we can write R as a convex combination $\delta D + (1-\delta)\hat{D}$ and U_X as a convex combination $\delta M + (1-\delta)\hat{M}$. Then

$$\begin{aligned} \mathbb{E}[f(R)] - \mathbb{E}[f(U_X)] &= \delta \cdot (\mathbb{E}[f(D)] - \mathbb{E}[f(M)]) + (1-\delta) \cdot \left(\mathbb{E}[f(\hat{D})] - \mathbb{E}[f(\hat{M})]\right) \\ &\geq \delta\epsilon - (1-\delta) \cdot \left(\mathbb{E}[f(\hat{M})] - \mathbb{E}[f(\hat{D})]\right). \end{aligned}$$

We would be done if we could ensure that $\mathbb{E}[f(\hat{M})] - \mathbb{E}[f(\hat{D})]$ is small, e.g. smaller than $\epsilon\delta$. Ideally, we would like to choose M to consist of the inputs on which f is largest; this would guarantee that the average of f on $\hat{M}$ is small. However, this approach is circular — according to our hypothesis, f may depend on the choice of M.

Thus, the first step in our proof is to "switch quantifiers" in our hypothesis, and to exhibit a single function $\bar{f}$ that distinguishes *every* M from D. The price that we pay is that $\bar{f}$ is no longer (guaranteed to be) an element of $\mathcal{F} \cup (1 - \mathcal{F})$, but is rather a *convex combination* of elements of $\mathcal{F} \cup (1 - \mathcal{F})$.

Claim 2.2 *There exists a function $\bar{f} : X \to [0,1]$ that is a convex combination of functions in $\mathcal{F} \cup (1 - \mathcal{F})$ and such that for every distribution M that is δ-dense in U_X we have*

$$\mathbb{E}\left[\bar{f}(D)\right] - \mathbb{E}\left[\bar{f}(M)\right] \geq \epsilon$$

Proof of claim: This is an application of duality of linear programming or, equivalently, of the min-max theorem in game theory. In the latter language, we think of a zero-sum game where the first player picks a function $f \in \mathcal{F}$, the second player picks a distribution M that is δ-dense in U_X, and the payoff is $\mathbb{E}[f(D)] - \mathbb{E}[f(M)]$ for the first player, and $-(\mathbb{E}[f(D)] - \mathbb{E}[f(M)])$ for the second player.

By the min-max theorem, the game has a "value" α for which the first player has an optimal mixed strategy (a distribution over strategies) $\bar{f}$, and the second player has an optimal mixed strategy $\bar{M}$, such that

$$\forall M \ \delta\text{-dense in } U_X, \quad \mathbb{E}[\bar{f}(D)] - \mathbb{E}[\bar{f}(M)] \geq \alpha \tag{2}$$

and

$$\forall f \in \mathcal{F} \cup (1 - \mathcal{F}), \quad \mathbb{E}[f(D)] - \mathbb{E}[f(\bar{M})] \leq \alpha \tag{3}$$

Since $\bar{M}$ is a distribution over δ-dense distributions, $\bar{M}$ is δ-dense as well. The hypothesis of the theorem tells us that there exists a function f distinguishing D from $\bar{M}$ with advantage at least ϵ. Taking this f in Inequality Inequality (3), we get that $\alpha \geq \epsilon$. The claim now follows from Equation (2). □

Now, following the earlier intuition, we consider the set S consisting of the $\delta \cdot |X|$ elements of X with the largest value of $\bar{f}()$, and take the uniform distribution over S, denoted U_S, as our model distribution. Since U_S is δ-dense in U_X, we have that $\mathbb{E}[\bar{f}(D)] \geq \mathbb{E}[\bar{f}(U_S)] + \epsilon$. In other words, the function $\bar{f}$ "distinguishes" D from U_S in the sense that $\bar{f}$ is a bounded function and its average is noticeably larger over D versus over U_S. Now we would like to use $\bar{f}$ in order to distinguish R from U_X.

First, however, we show that D and U_S can also be distinguished via a Boolean function, which is in fact a thresholded version of $\bar{f}$). This will follows from the next claim.

Claim 2.3 *Let $F : X \to [0,1]$ be a bounded function, let Z and W be distributions such that $\mathbb{E}[F(Z)] \geq \mathbb{E}[F(W)] + \epsilon$. Then there is a real number $t \in [\epsilon/2, 1]$ such that*

$$\mathbb{P}[F(Z) \geq t] \geq \mathbb{P}[F(W) \geq t - \epsilon/2] + \epsilon/2$$

Proof of claim: We use the fact that, for a nonnegative random variable Y, we have $\mathbb{E}[Y] = \int_0^\infty \mathbb{P}[Y \geq t]\, dt$. Suppose towards a contradiction that for every $\epsilon/2 \leq t \leq 1$ we have

$$\mathbb{P}[F(Z) \geq t] < \mathbb{P}[F(W) \geq t - \epsilon/2] + \epsilon/2$$

Then

$$\begin{aligned}
\mathbb{E}[F(Z)] &= \int_0^1 \mathbb{P}[F(Z) \geq t]\, dt \\
&< \int_0^{\epsilon/2} \mathbb{P}[F(Z) \geq t]\, dt + \int_{\epsilon/2}^1 (\mathbb{P}[F(W) \geq t - \epsilon/2] + \epsilon/2) dt \\
&< \epsilon/2 + \mathbb{E}[F(W)] + \epsilon/2
\end{aligned}$$

□

By applying the claim with $F = \bar{f}$, $Z = D$ and $W = U_S$, we obtain a probability q and a threshold t such that

$$\mathbb{P}\left[\bar{f}(U_S) \geq t - \epsilon/2\right] = q$$
$$\mathbb{P}\left[\bar{f}(D) \geq t\right] \geq q + \epsilon/2$$

In particular, these conditions imply that the event that $\bar{f}$ is above the threshold t distinguishes between U_S and D. We will now show that this event also distinguishes U_X from R. For this, we will use the fact that U_S is the δ-dense distribution that maximizes $\bar{f}$.

Since $q < 1$ (as $q + \epsilon/2 \leq 1$), we have that the condition $\bar{f}(x) \geq t - \epsilon/2$ fails for some elements of S. By the definition of S, *this condition also fails everywhere outside of* S. Recalling that S was chosen to contain a δ fraction of the elements of X, we have

$$\mathbb{P}\left[\bar{f}(U_X) \geq t - \epsilon/2\right] = \delta q. \tag{4}$$

Recalling also that X has density δ in R, we have

$$\mathbb{P}\left[\bar{f}(R) \geq t\right] \geq \delta \cdot \mathbb{P}\left[\bar{f}(D) \geq t\right] \geq \delta q + \delta\epsilon/2. \tag{5}$$

We have just shown that the event that $\bar{f}$ is above the threshold t distinguishes between R and U_X, with some additional slackness (in the sense that for $f(U_X)$ we consider the smaller threshold $t - \epsilon/2$). This slackness will allow us to replace the threshold version of $\bar{f}$ with low-complexity approximations, thus establishing the theorem. We will use different approximations for Parts 1 and 2 of the theorem. In both cases, it will be useful to assume that $\bar{f}$ is a convex combination of (i.e. distribution on) functions in $\mathcal{F} \cup -\mathcal{F}$ rather than $\mathcal{F} \cup (1 - \mathcal{F})$; this can be achieved by reducing the threshold t by $\mathbb{P}[\bar{f} \notin \mathcal{F}]$.

Proof of Part (1). Viewing $\bar{f}$ as a distribution over functions $f \in \mathcal{F} \cup -\mathcal{F}$, Chernoff bounds imply that it will be well-approximated by the average of a few functions sampled randomly from the distribution. Formally, we have:

Claim 2.4 *Let $F : \Omega \to [-1, 1]$ be a convex combination of bounded functions from a class $\mathcal{G}$, let Z, W be two distributions on Ω, and let $\alpha, \beta > 0$.*

Then there are functions $f_1, \ldots, f_k$ in $\mathcal{G}$ (not necessarily distinct) where $k = O((1/\alpha^2) \cdot \log(1/\beta))$, such that

$$\mathbb{P}\left[\left|F(Z) - \frac{f_1(Z) + \cdots + f_k(Z)}{k}\right| > \alpha\right] \leq \beta$$

and

$$\mathbb{P}\left[\left|F(W) - \frac{f_1(W) + \cdots + f_k(W)}{k}\right| > \alpha\right] \leq \beta$$

Proof of claim: This is an immediate consequence of the Chernoff bound. Fix an input x. Pick randomly and independently k functions $f_1, \ldots, f_k$ from $\mathcal{G}$ with the probability distribution given by F. For $k = O((1/\alpha^2) \cdot \log(1/\beta))$, we have that for every fixed element x

$$\mathbb{P}_{f_1,\ldots,f_k}\left[\left|F(x) - \frac{f_1(x) + \cdots + f_k(x)}{k}\right| > \alpha\right] \leq \beta/10.$$

Therefore, for every probability distribution Y,

$$\mathop{\mathbb{E}}_{f_1,\ldots,f_k}\left(\mathbb{P}_{x\in Y}\left[\left|F(x)-\frac{f_1(x)+\cdots+f_k(x)}{k}\right|>\alpha\right]\right)\leq\beta/10.$$

By Markov's inequality,

$$\mathbb{P}_{f_1,\ldots,f_k}\left[\mathbb{P}_{x\in Y}\left[\left|F(x)-\frac{f_1(x)+\cdots+f_k(x)}{k}\right|>\alpha\right]>\beta\right]\leq 1/10$$

for every probability distribution Y. Therefore, there exists a choice of functions $f_1,\ldots,f_k$ such that $\mathbb{P}\left[|F(x)-(f_1(x)+\cdots+f_k(x))/k|>\alpha\right]\leq\beta$, both when x is selected according to W, and when it is selected according to Z. □

Apply Claim 2.4 with family $\mathcal{G}=\mathcal{F}\cup-\mathcal{F}$, function $F=\bar{f}$, distributions $Z=R$, $W=U_X$, and parameters $\beta=\epsilon\delta/10$, $\alpha=\epsilon/10$. We then have $k=O((1/\epsilon^2)\cdot\log(1/\epsilon\delta))$ functions $f_1,\ldots,f_k\in\mathcal{F}\cup-\mathcal{F}$ such that $\sum_i f_i(x)/k$ is a good approximation of $\bar{f}(x)$ in the sense that both

$$\mathbb{P}\left[\sum_i f_i(U_X)\geq kt-.4k\epsilon\right]\leq\delta q+.1\epsilon\delta,$$

and also

$$\mathbb{P}\left[\sum_i f_i(R)\geq kt-.1k\epsilon\right]\geq\delta q+.4\delta\epsilon$$

This means that if we define Boolean h by

$$h(x)=1\Leftrightarrow\left(\sum_i f_i(x)\geq kt-.4k\epsilon\right)$$

we will have that h satisfies $|\,\mathbb{E}[h(R)]-\mathbb{E}[h(U_X)]|\geq\Omega(\epsilon\delta)$ as required by the theorem.

Proof of Part (2). For Part (2), we want to turn our distinguisher into a product of f_i's rather than a threshold function. Similarly to [TZ], we do this by approximating our distinguisher from Inequalities (4) and (5), which is a linear threshold function in $\bar{f}(x)$, by a low-degree polynomial in $\bar{f}(x)$. From this we deduce that some power of $\bar{f}(x)$ must be a distinguisher (albeit with exponentially smaller advantage), and then we observe that a power of $\bar{f}$ is a convex combination of products of elements of $\mathcal{F}\cup-\mathcal{F}$ to obtain a distinguisher of the latter type.

The approximation of the threshold function $[z\geq t]$ is given by the following special case of the Weierstrass Approximation Theorem.

Claim 2.5 *For every $\alpha,\beta\in[0,1]$, $t\in[\alpha,1]$, there exists a polynomial p of degree* $\mathrm{poly}(1/\alpha,1/\beta)$ *and with coefficients bounded in absolute value by* $\exp(\mathrm{poly}(1/\alpha,1/\beta))$ *such that*

1. *For all $z\in[0,1]$, we have $p(z)\in[0,1]$.*
2. *For all $z\in[0,t-\alpha]$, we have $p(z)\in[0,\beta]$.*
3. *For all $z\in[t,1]$, we have $p(z)\in[1-\beta,1]$.*

We set $\alpha = \epsilon/2$ and $\beta = \epsilon\delta/8$ in the claim to obtain a polynomial $p(z) = \sum_{i=0}^{d} c_i z^i$ of degree $d = \text{poly}(1/\epsilon, 1/\delta)$ with coefficients satisfying $|c_i| \leq \exp(\text{poly}(1/\epsilon, 1/\delta))$. Combining the properties of the polynomial p with Inequalities (4) and (5), we get:

$$\begin{aligned} \mathbb{E}\left[p\left(\bar{f}(U_X)\right)\right] &\leq \delta q + \epsilon\delta/8 \\ \mathbb{E}\left[p\left(\bar{f}(R)\right)\right] &\geq \delta q + 3\epsilon\delta/8 \end{aligned}$$

and so $p(\bar{f}(x)) = \sum_i c_i \bar{f}(x)^i$ distinguishes between U_X and R with probability at least $\epsilon\delta/4$. Thus, there must exist a single term $c_k \bar{f}(x)^k$ that distinguishes U_X and R with probability at least $\epsilon\delta/(4(d+1))$, which in turn implies that $\bar{f}(x)^k$ distinguishes U_X and R with probability at least $\epsilon' = \epsilon\delta/(4(d+1)|c_k|) = \exp(-\text{poly}(1/\epsilon, 1/\delta))$. That is,

$$\mathbb{E}\left[\bar{f}(R)^k\right] - \mathbb{E}\left[\bar{f}(U_X)^k\right] \geq \epsilon'.$$

Note that for every x, $\bar{f}(x) = \mathbb{E}_{f_1,\ldots,f_k}[\prod f_k(x)]$, where $f_1, \ldots, f_k$ are sampled according to $\bar{f}$, when viewed as a distribution on $\mathcal{F} \cup -\mathcal{F}$. By averaging, there exist $f_1, \ldots, f_k \in \mathcal{F} \cup -\mathcal{F}$, such that

$$\mathbb{E}\left[\prod_i f_i(R)\right] - \mathbb{E}\left[\prod_i f_i(U_X)\right] \geq \epsilon'.$$

If we take the absolute value of the left-hand-side, then we may remove the negations from any functions in $-\mathcal{F}$, completing the proof of Part (2). ■

3 Hardcore Theorems via Iterative Partitioning

Impagliazzo's Hardcore Theorem [Imp] and its variants say that if a function $g : \{0,1\}^n \to \{0,1\}$ is "mildly hard", meaning that every "efficient" algorithm f errs in computing g on at least some δ fraction of inputs in X, then there is a "hardcore" set $H \subset \{0,1\}^n$ of inputs, of density roughly δ, on which g is "very hard" to compute. In Impagliazzo's original formulation, "very hard" means that no efficient f can compute g on a random input in H much better than random guessing, i.e. $f(x) = g(x)$ with probability at most $1/2 + \epsilon$ on a random $x \in H$. This conclusion implies the following three properties:

1. g is nearly balanced on H, i.e. $\mathbb{P}_{x \in H}[g(x) = 1] \in [1/2 - \epsilon, 1/2 + \epsilon]$. (Otherwise, a trivial constant predictor f would compute g with probability larger than $1/2 + \epsilon$.)

2. No efficient f can compute g on a random input in H much better than a constant predictor, i.e. $\mathbb{P}_{x \in H}[f(x) = g(x)] \leq \max\{\mathbb{P}_{x \in H}[g(x) = 0], \mathbb{P}_{x \in H}[g(x) = 1]\} + \epsilon$. (Indeed, the right-hand side is always at least $1/2 + \epsilon$.)

3. No efficient f can distinguish a random element of $H \cap g^{-1}(0)$ from a random element of $H \cap g^{-1}(1)$, except with probability $O(\epsilon)$. That is, for every efficient f,

$$|\mathbb{P}_{x \in H \cap g^{-1}(0)}[f(x) = 1] - \mathbb{P}_{x \in H \cap g^{-1}(1)}[f(x) = 1]| \leq O(\epsilon).$$

   (Using the fact that g is nearly balanced on H, it can be shown that if f distinguishes the two distributions with probability greater than 4ϵ, then either f or its negation computes g correctly with probability greater than $1/2 + \epsilon$ on a random element of H.)

When g is nearly balanced on H (as in Property 1), then Properties 2 and 3 are actually equivalent to the original conclusion of the Hardcore Theorem (up to a constant factor change in ϵ). However, when g is not balanced on H, then they are weaker. Indeed, in the extreme case that g is constant on H, then Property 2 trivially holds (because a constant predictor succeeds with probability 1) and Property 3 is not even well-defined. But as long as we require that g is not extremely biased on H, then both Properties 2 and 3 are already nontrivial and interesting (even if weaker than the conclusion original Hardcore Theorem).

In this section, we will show how iterative partitioning arguments, in the spirit of the proofs of the Szemerédi's Regularity Lemma [Sze] and the Green–Tao–Ziegler of the Dense Model Theorems [GT, TZ], can be used to prove Hardcore Theorems (albeit with a loss in efficiency that is exponential in ϵ and δ). These will include one with a conclusion of the same type as in Impagliazzo's original result [Imp], as well as ones establishing Properties 2 and 3 where we do not guarantee g is nearly balanced (but only that it is not extremely biased). The novel feature of our results establishing Properties 2 and 3 is that the hardcore set H is efficiently recognizable. This feature is impossible to achieve in general if we require that g be nearly balanced on H. Indeed, if we select a random function g in which each input is set to 1 independently with probability $1-\delta$, then with high probability, g will be mildly hard to compute, but will be biased on every efficiently recognizable set of noticeable density.[5]

We begin with our version of the Hardcore Theorem where it is hard to compute g on H better than a constant predictor. Let $\mathcal{F}$ be a class of functions and k be an integer parameter. Then we denote by $C(\mathcal{F},k)$ the class of functions of the form $h(f_1(x),\ldots,f_k(x))$, where $f_i \in \mathcal{F}$ and $h:\{0,1\}^k \to \{0,1\}$ is arbitrary.

Theorem 3.1 *Let $\mathcal{F}$ be a class of boolean functions, $g:\{0,1\}^n \to \{0,1\}$ be a function, $\epsilon,\delta>0$ be given parameters. Then there is a $k \leq 1/\delta\epsilon^2$ such that the following holds.*

Suppose that for every $f \in C(\mathcal{F},k)$ we have $\mathbb{P}[f(x) \neq g(x)] > \delta$.

Then there is a set $H \subseteq \{0,1\}^n$ of density at least δ — indeed, with both $|H \cap g^{-1}(0)|$ and $|H \cap g^{-1}(1)|$ being of density at least $\delta/2$ — such that, for every $f \in F$,

$$\mathbb{P}_{x\in H}[f(x)=g(x)] \leq \max\{\mathbb{P}_{x\in H}[g(x)=1], \mathbb{P}_{x\in H}[g(x)=0]\} + \epsilon \tag{6}$$

Furthermore, the characteristic function of H is in $C(\mathcal{F},k)$.

Note that since $|H \cap g^{-1}(b)| \geq (\delta/2)\cdot 2^n$ for $b=0,1$, it follows that g is not too biased on H. Specifically, $\mathbb{P}_{x\in H}[g(x)=b] \geq \delta/2$ for both $b=0,1$. The main inefficiency in the theorem is that in order to derive the conclusion, the function g must be mildly hard for all of $C(\mathcal{F},k)$, which contains functions of circuit complexity exponential in k.

Proof: Let $\mathcal{P}=(P_1,\ldots,P_m)$ be a partition of $\{0,1\}^n$. Then we let $Y(\mathcal{P})$ be the union of the sets P_i where $g()$ equals 1 on a majority of elements, and $N(\mathcal{P})$ be the union of the remaining sets. We say that a partition $\mathcal{P}=(P_1,\ldots,P_m)$ *satisfies the stopping condition* if at least one of the following conditions holds

[5] Alternatively, we can set $g(x)$ to be the first bit of x with probability $1-\delta$, independently for each x. Then g will be nearly balanced globally, and can be computed with probability nearly $1-\delta$ on every efficiently recognizable set. Thus, additionally requiring that g be globally balanced does not help in strengthening the conclusion of the theorem.

- $Y(\mathcal{P}) \cap g^{-1}(0)$ has density at least $\delta/2$ in $\{0,1\}^n$ and for every $f \in \mathcal{F}$ we have

$$\mathbb{P}_{x \in Y(\mathcal{P})}[g(x) = f(x)] \leq \mathbb{P}_{x \in Y(\mathcal{P})}[g(x) = 1] + \epsilon \ .$$

- $N(\mathcal{P}) \cap g^{-1}(1)$ has density at least $\delta/2$ in $\{0,1\}^n$ and for every $f \in \mathcal{F}$ we have

$$\mathbb{P}_{x \in N(\mathcal{P})}[g(x) = f(x)] \leq \mathbb{P}_{x \in N(\mathcal{P})}[g(x) = 0] + \epsilon \ .$$

Note that if $\mathcal{P}$ satisfies the stopping condition, then either $Y(\mathcal{P})$ or $N(\mathcal{P})$ have all the properties we require of the set H in the statement of the theorem, except the efficient computability. We will now show that we can find a partition that satisfies the stopping condition and where $Y(\mathcal{P})$ and $N(\mathcal{P})$ are efficiently computable.

First we introduce some terminology: the minority inputs in $Y(\mathcal{P})$ are the inputs $x \in Y(\mathcal{P})$ such that $g(x) = 0$; similarly, the minority inputs in $N(\mathcal{P})$ are the inputs $x \in N(\mathcal{P})$ such that $g(x) = 1$

We construct the partition iteratively. We begin at the 0-th step with the trivial partition $\mathcal{P} := (\{0,1\}^n)$. We maintain the invariant that, at step i, there are functions $f_1, \ldots, f_i$ in $\mathcal{F}$ such that the partition at step i is generated by $f_1, \ldots, f_i$ in the following sense: the partition has a set $P_{b_1,\ldots,b_i}$ for each bit string $b_1, \ldots, b_i$, defined as

$$P_{b_1,\ldots,b_i} := \{x : f_1(x) = b_1 \ldots f_i(x) = b_i\} \tag{7}$$

Note that the union of any subset of the sets in the partition is computable in $C(\mathcal{F}, i)$. So we are done if, at some step $i \leq 1/\epsilon^2\delta$, the partition satisfies the stopping condition.

Suppose then that the partition at step i does not satisfy the stopping condition. Provided $i \leq k$, we claim that the total number of minority inputs in $Y(\mathcal{P})$ and $N(\mathcal{P})$ must be at least $\delta \cdot 2^n$. If not, consider the function that, on input x, computes $f_1(x), \ldots, f_i(x)$, and then outputs the majority answer in $P_{f_1(x),\ldots,f_i(x)}$; such a function is in $C(\mathcal{F}, i)$ and it would compute g correctly on greater than a $1 - \delta$ fraction of inputs.

This means that, if the partition does not satisfy the stopping condition and $i \leq k$, then there is a set H, which is either $Y(\mathcal{P})$ and $N(\mathcal{P})$, that contains at least $(\delta/2) \cdot 2^n$ minority elements and such that there is a function $f \in \mathcal{F}$ that has advantage ϵ over the constant predictor. We then refine our partition according to f. That is, at step $i+1$ our partition is the one generated by $f_1, \ldots, f_i, f$.

We want to show that this process terminates after no more than $k \leq 1/\epsilon^2\delta$ steps. To this end, we associate a potential function to a partition, observe that the value of the potential function is at most 1 and at least 0, and show that at each step of the argument the value of the potential function increases by at least $\epsilon^2\delta$.

For a partition $\mathcal{P} = (P_1, \ldots, P_t)$, we define its potential function as

$$\mathcal{E}(\mathcal{P}) := \sum_{P \in \mathcal{P}} \frac{|P|}{2^n} \cdot \mathbb{P}_{x \in P}[g(x) = 1]^2 = \mathop{\mathbb{E}}_{P}\left[\mathop{\mathbb{E}}_{x \in P}[g(x)]^2\right],$$

where the latter expectation is taken over a random block $P \in \mathcal{P}$ chosen with probability $|P|/2^n$. That is, we compute the average over the blocks of the partition of the square of the density of the YES instances of g inside each blocks. Up to an additive term, this is the variance of the density of YES instances of g across blocks. It is clear by definition that this quantity is positive and at most 1.

Now we show that if P is a set in the partition and we further partition it according to a function f that has some α advantage over the constant predictor for g, then the contribution of the elements

of P to the potential function increases by at least $\alpha^2|P|/2^n$. To prove this, we first observe that the change in the potential function equals $|P|/2^n$ times the quantity

$$\Delta(P) := \mathbb{P}[f=0]\cdot\mathbb{P}[g=1|f=0]^2 + \mathbb{P}[f=1]\cdot\mathbb{P}[g=1|f=1]^2 - \mathbb{P}[g=1]^2,$$

where for readability we drop the input x to f and g and all probabilities are taken over a random $x \in P$. Note that $\Delta(P)$ is equal to the variance of the random variable W where we choose a bit $b \in \{0,1\}$ with probability $\mathbb{P}[f=b]$ and set $W = \mathbb{P}[g=1|f=b]$. (To see this, observe that $\mathbb{E}[W] = \mathbb{P}[g=1]$ and the first two terms in the expression for $\Delta\mathcal{E}(P)$ are equal to $\mathbb{E}[W^2]$.) In particular $\Delta(P)$ is always nonnegative, so the potential function cannot decrease when we refine P. In addition, we have:

$$\Delta(P) = \mathrm{Var}[W] = \mathbb{E}[(W-\mathbb{E}[W])^2] \geq \mathbb{E}[|W-\mathbb{E}[W]|]^2 \tag{8}$$

Now,

$$\begin{aligned}
&\mathbb{E}\left[|W-\mathbb{E}[W]|\right] \\
&\quad= \mathbb{P}[f=0]\cdot|\mathbb{P}[g=1|f=0]-\mathbb{P}[g=1]| + \mathbb{P}[f=1]\cdot|\mathbb{P}[g=1|f=1]-\mathbb{P}[g=1]| \\
&\quad= \mathbb{P}[f=0]\cdot|\mathbb{P}[g=0|f=0]-\mathbb{P}[g=0]| + \mathbb{P}[f=1]\cdot|\mathbb{P}[g=1|f=1]-\mathbb{P}[g=1]| \\
&\quad= |\mathbb{P}[g=0\wedge f=0]-\mathbb{P}[g=0]\cdot\mathbb{P}[f=0]| + |\mathbb{P}[g=1\wedge f=1]-\mathbb{P}[g=1]\cdot\mathbb{P}[f=1]| \qquad (9)\\
&\quad\geq |\mathbb{P}[g=0\wedge f=0]-\mathbb{P}[g=0]\cdot\mathbb{P}[f=0]+\mathbb{P}[g=1\wedge f=1]-\mathbb{P}[g=1]\cdot\mathbb{P}[f=1]| \\
&\quad\geq \mathbb{P}[f=g]-\max\{\mathbb{P}[g=0],\mathbb{P}[g=1]\}
\end{aligned}$$

To summarize, the change in potential function contributed by elements of P is given by $|P|/2^n$ times

$$\Delta(P) \geq \left(\mathbb{P}_{x\in P}[f(x)=g(x)] - \max\{\mathbb{P}_{x\in P}[g(x)=0],\mathbb{P}_{x\in P}[g(x)=1]\}\right)^2.$$

Now, recall that, if we haven't satisfied the stopping condition, then there is a set $H \in \{Y(\mathcal{P}), N(\mathcal{P})\}$ that contains at least $(\delta/2)\cdot 2^n$ minority elements and a function f that has advantage ϵ over the constant predictor on H. By symmetry, we may assume that $H = Y(\mathcal{P})$. Then overall change in potential function is given by:

$$\begin{aligned}
\Delta\mathcal{E}(\mathcal{P}) &= \sum_P \frac{|P|}{2^n}\cdot\Delta(P) \\
&\geq \sum_P \frac{|P|}{2^n}\cdot\left(\mathbb{P}_{x\in P}[f(x)=g(x)] - \max\{\mathbb{P}_{x\in P}[g(x)=0],\mathbb{P}_{x\in P}[g(x)=1]\}\right)^2 \\
&\geq \sum_{P\subseteq H} \frac{|P|}{2^n}\cdot\left(\mathbb{P}_{x\in P}[f(x)=g(x)] - \mathbb{P}_{x\in P}[g(x)=1]\right)^2 \\
&= \frac{|H|}{2^n}\cdot\mathop{\mathbb{E}}_{P\subseteq H}\left[\left(\mathbb{P}_{x\in P}[f(x)=g(x)] - \mathbb{P}_{x\in P}[g(x)=1]\right)^2\right] \\
&\geq \delta\cdot\mathop{\mathbb{E}}_{P\subseteq H}\left[\mathbb{P}_{x\in P}[f(x)=g(x)] - \mathbb{P}_{x\in P}[g(x)=1]\right]^2 \\
&= \delta\cdot\left(\mathbb{P}_{x\in H}[f(x)=g(x)] - \mathbb{P}_{x\in H}[g(x)=1]\right)^2 \\
&\geq \delta\epsilon^2.
\end{aligned}$$

This means that the process we described above finds a partition that satisfies the stopping condition after no more than $1/\delta\epsilon^2$ steps. The theorem follows by setting $k = \lfloor 1/\delta\epsilon^2 \rfloor$. ■

We now prove a more general result that implies the above Hardcore Theorem, one of the original flavor (where g cannot be computed on H with probability more than $1/2 + \epsilon$), as well as one achieving Property 3 mentioned above (where $H \cap g^{-1}(0)$ and $H \cap g^{-1}(1)$ are indistinguishable from each other). We will state this more general theorem in terms of the partition we construct, and then deduce the Hardcore Theorems as corollaries. For a fixed function $g : \{0,1\}^n \to \{0,1\}$, and $P \subseteq \{0,1\}^n$, we write $\mathrm{maj}(P)$ to denote the majority value of g on P, $\min(P)$ for the minority value, P^{maj} to be the set of elements of P on which g takes on value $\mathrm{maj}(P)$, and similarly $P^{\min}$.

Theorem 3.2 *Let $\mathcal{F}$ be a class of boolean functions, $g : \{0,1\}^n \to \{0,1\}$ be a function, $\epsilon, \delta > 0$ be given parameters. Then there is a $k \leq 1/\delta^2\epsilon^2$ such that the following holds.*

Suppose that for every $f \in C(\mathcal{F}, k)$ we have $\mathbb{P}[f(x) \neq g(x)] > \delta$.

Then there is a partition $\mathcal{P} = (P_1, \ldots, P_m)$ of $\{0,1\}^n$ such that such that

1. *$\bigcup_{P \in \mathcal{P}} P^{\min}$ is of size at least $\delta \cdot 2^n$, and*
2. *For every $f \in \mathcal{F}$,*

$$\mathop{\mathbb{E}}_{P \in D_{\min}} [|\mathbb{P}_{x \in P^{\mathrm{maj}}}[f(x) = 1] - \mathbb{P}_{x \in P^{\min}}[f(x) = 1]|] \leq \epsilon,$$

where $D_{\min}$ is the distribution that selects $P \in \mathcal{P}$ with probability proportional to $|P^{\min}|$.

Moreover, the partition $\mathcal{P}$ is defined by k functions $f_1, \ldots, f_k \in \mathcal{F}$ (in the sense of Equation (7)).

Proof: As before, we will proceed by iteratively constructing the partition $\mathcal{P} = (P_1, \ldots, P_m)$, and argue that at each stage, either it satisfies the required properties or we can refine the partition in such a way that a potential function grows noticeably.

Like in the previous proof, the set of all minority inputs, $\bigcup_{P \in \mathcal{P}} P^{\min}$, must be of size at least $\delta \cdot 2^n$ (provided $i \leq k$); otherwise there is a function $f \in C(\mathcal{F}, k)$ that computes g on more than a $1 - \delta$ fraction of inputs. Thus, if Condition 2 in the theorem holds (and $i \leq k$), we are finished.

Suppose that Condition 2 fails. That is, there is a function $f \in \mathcal{F}$ such that

$$\mathop{\mathbb{E}}_{P \in D_{\min}} [|\mathbb{P}_{x \in P^{\mathrm{maj}}}[f(x) = 1] - \mathbb{P}_{x \in P^{\min}}[f(x) = 1]|] > \epsilon.$$

We want to use this to argue that the potential function increases, where we use the same potential function as in the previous proof, namely

$$\mathcal{E}(\mathcal{P}) := \sum_{P \in \mathcal{P}} \frac{|P|}{2^n} \cdot \mathbb{P}_{x \in P}[g(x) = 1]^2.$$

As before, we first reason about what happens in a single set P of the partition. Like in the previous proof (specifically Inequalities (8) and (9)), the increase in the potential function corresponding to elements of P is $|P|/2^n$ times $\Delta(P)$, where

$$\Delta(P)^{1/2} \geq |\mathbb{P}[g = 0 \wedge f = 0] - \mathbb{P}[g = 0] \cdot \mathbb{P}[f = 0]| + |\mathbb{P}[g = 1 \wedge f = 1] - \mathbb{P}[g = 1] \cdot \mathbb{P}[f = 1]|.$$

We bound each of the terms on the RHS as follows:

$$\begin{aligned}
&|\mathbb{P}[g=1 \wedge f=1] - \mathbb{P}[g=1]\cdot\mathbb{P}[f=1]| \\
&= |\mathbb{P}[g=1 \wedge f=1] - \mathbb{P}[g=1]\cdot\mathbb{P}[f=1 \wedge g=1] - \mathbb{P}[g=1]\cdot\mathbb{P}[f=1 \wedge g=0]| \\
&= |\mathbb{P}[g=0]\cdot\mathbb{P}[g=1 \wedge f=1] - \mathbb{P}[g=1]\cdot\mathbb{P}[f=1 \wedge g=0]| \\
&= \mathbb{P}[g=0]\cdot\mathbb{P}[g=1]\cdot|\mathbb{P}[f=1|g=1] - \mathbb{P}[f=1|g=0]| \\
&\geq \frac{\min\{\mathbb{P}[g=0],\mathbb{P}[g=1]\}}{2}\cdot|\mathbb{P}[f=1|g=1] - \mathbb{P}[f=1|g=0]| \\
&= \frac{|P^{\min}|}{2|P|}\cdot|\mathbb{P}_{x\in P^{\mathrm{maj}}}[f(x)=1] - \mathbb{P}_{x\in P^{\min}}[f(x)=1]|\,.
\end{aligned}$$

Similarly, we have:

$$\begin{aligned}
&|\mathbb{P}[g=0 \wedge f=0] - \mathbb{P}[g=0]\cdot\mathbb{P}[f=0]| \\
&\geq \frac{|P^{\min}|}{2|P|}\cdot|\mathbb{P}_{x\in P^{\mathrm{maj}}}[f(x)=0] - \mathbb{P}_{x\in P^{\min}}[f(x)=0]| \\
&= \frac{|P^{\min}|}{2|P|}\cdot|\mathbb{P}_{x\in P^{\mathrm{maj}}}[f(x)=1] - \mathbb{P}_{x\in P^{\min}}[f(x)=1]|\,.
\end{aligned}$$

Thus

$$\begin{aligned}
\Delta\mathcal{E}(\mathcal{P})^{1/2} &= \left(\sum_P \frac{|P|}{2^n}\cdot\Delta(P)\right)^{1/2} \\
&\geq \sum_P \frac{|P|}{2^n}\cdot\Delta(P)^{1/2} \\
&\geq \sum_P \frac{|P^{\min}|}{2^n}\cdot|\mathbb{P}_{x\in P^{\mathrm{maj}}}[f(x)=1] - \mathbb{P}_{x\in P^{\min}}[f(x)=1]| \\
&\geq \delta\cdot\mathop{\mathbb{E}}_{P\in D_{\min}}[|\mathbb{P}_{x\in P^{\mathrm{maj}}}[f(x)=1] - \mathbb{P}_{x\in P^{\min}}[f(x)=1]|] \\
&\geq \delta\epsilon.
\end{aligned}$$

That is, the potential function increases by at least $\delta^2\epsilon^2$ at each step, so we finish with $k \leq 1/\delta^2\epsilon^2$. ■

Now we show how this theorem implies a Hardcore Theorem of the original form, giving a δ-dense H on which g is hard to predict with probability greater than $1/2+\epsilon$. Like many proofs of the Hardcore Theorem, we obtain a δ-dense *distribution* rather than a δ-dense *set*. But Impagliazzo [Imp] shows how to go from the former to the latter, with a small loss in parameters.

Corollary 3.3 *Let $\mathcal{F}$ be a class of boolean functions, $g:\{0,1\}^n \to \{0,1\}$ be a function, $\epsilon,\delta>0$ be given parameters. Then there is a $k \leq 1/\delta^2\epsilon^2$ such that the following holds.*

Suppose that for every $f \in C(\mathcal{F},k)$ we have $\mathbb{P}[f(x)\neq g(x)] > \delta$.

Then there is a distribution *H that is 2δ-dense in $\{0,1\}^n$ such that for every $f\in\mathcal{F}$, we have $\mathbb{P}_{x\in H}[f(x)=g(x)] < (1+\epsilon)/2$.*

Proof: Let $\mathcal{P}$ be the partition given by Theorem 3.2. The distribution H is defined as follows: Select $P \in D_{\min}$, and with probability $1/2$, output $x \in P^{\min}$ and with probability $1/2$, output $x \in P^{\maj}$. To see that this distribution is 2δ-dense, observe that with probability $1/2$, H produces a uniformly random element of $\bigcup_{P\in\mathcal{P}} P^{\min}$, which is a δ-dense set, and with probability $1/2$ H samples from a disjoint distribution of even higher density (since $|P^{\maj}| \geq |P^{\min}$ for every P).)

Now we argue that g is very hard to compute on H.

$$\begin{aligned}
\mathbb{P}_{x\in H}[f(x) = g(x)] &= \mathop{\mathbb{E}}_{P\in D_{\min}}\left[\frac{1}{2}\cdot\mathbb{P}_{x\in P^{\min}}[f(x)=\min(P)] + \frac{1}{2}\cdot(1-\mathbb{P}_{x\in P^{\maj}}[f(x)=\min(P)])\right] \\
&\leq \frac{1}{2}+\frac{1}{2}\cdot\mathop{\mathbb{E}}_{P\in D_{\min}}[|\mathbb{P}_{x\in P^{\min}}[f(x)=\min(P)] - \mathbb{P}_{x\in P^{\maj}}[f(x)=\min(P)]|] \\
&= \frac{1}{2}+\frac{1}{2}\cdot\mathop{\mathbb{E}}_{P\in D_{\min}}[|\mathbb{P}_{x\in P^{\min}}[f(x)=1] - \mathbb{P}_{x\in P^{\maj}}[f(x)=1]|] \\
&\leq \frac{1}{2}+\frac{1}{2}\cdot\epsilon
\end{aligned}$$

■

Next we show how Theorem 3.2 implies Theorem 3.1, with a slightly worse bound on the number k of functions we need to combine.

Proof (Thm. 3.2⇒Thm. 3.1 with $k \leq 1/\delta^2\epsilon^2$): Let $\mathcal{P}$ be the partition of $\{0,1\}^n$ guaranteed by Theorem 3.2. As in the proof of Theorem 3.1, let $Y(\mathcal{P})$ be the union of the sets P such that $\maj(P) = 1$, and $N(\mathcal{P})$ be the union of the sets such that $\maj(P) = 0$. At least one set $H \in \{Y(\mathcal{P}), N(\mathcal{P})\}$, contains at least $(\delta/2)\cdot 2^n$ minority elements, i.e. $|H^{\min}| \geq (\delta/2)\cdot 2^n$. We now argue that no $f \in \mathcal{F}$ can predict g on H much better than a constant predictor:

$$\begin{aligned}
&\mathbb{P}_{x\in H}[f(x) = g(x)] \\
&= \frac{1}{|H|}\cdot\sum_{P\in\mathcal{P},P\subseteq H}\left(|P^{\min}|\cdot\mathbb{P}_{x\in P^{\min}}[f(x)=\min(H)] + |P^{\maj}|\cdot\mathbb{P}_{x\in P^{\maj}}[f(x)=\maj(H)]\right) \\
&= \frac{1}{|H|}\cdot\sum_{P\in\mathcal{P},P\subseteq H}\left(|P^{\min}|\cdot\mathbb{P}_{x\in P^{\min}}[f(x)=\min(H)] + |P^{\maj}|\cdot(1-\mathbb{P}_{x\in P^{\maj}}[f(x)=\min(H)])\right) \\
&\leq \frac{1}{|H|}\cdot\sum_{P\in\mathcal{P},P\subseteq H}\left(|P^{\maj}| + |P^{\min}|\cdot(\mathbb{P}_{x\in P^{\min}}[f(x)=\min(H)] - \mathbb{P}_{x\in P^{\maj}}[f(x)=\min(H)])\right) \\
&\leq \frac{|H^{\maj}|}{|H|} + \frac{|H^{\min}|}{|H|}\cdot\mathop{\mathbb{E}}_{P\overset{R}{\leftarrow}D_{\min}}[|\mathbb{P}_{x\in P^{\min}}[f(x)=1] - \mathbb{P}_{x\in P^{\maj}}[f(x)=1]|] \\
&\leq \max\{\mathbb{P}_{x\in H}[g(x)=0], \mathbb{P}_{x\in H}[g(x)=1]\} + \epsilon/2.
\end{aligned}$$

■

Finally, we deduce a Hardcore Theorem establishing Property 3 with an efficiently recognizable hardcore set H, with the price that the hardcore set is only of density $\Omega(\epsilon\delta^2)$ rather than $\Omega(\delta)$ as in most Hardcore Theorems.

Theorem 3.4 *Let $\mathcal{F}$ be a class of boolean functions, $g : \{0,1\}^n \to \{0,1\}$ be a function, $\epsilon, \delta > 0$ be given parameters. Then there is a $k = O(1/\delta^2\epsilon^4)$ such that the following holds.*

Suppose that for every $f \in C(\mathcal{F}, k)$ we have $\mathbb{P}[f(x) \neq g(x)] > \delta$.
Then there is a set $H \subseteq \{0,1\}^n$ such that such that

1. *$H^{\min}$ is of size at least $\Omega(\epsilon\delta^2) \cdot 2^n$, and*
2. *For every $f \in \mathcal{F}$,*
$$|\mathbb{P}_{x \in H^{\mathrm{maj}}}[f(x) = 1] - \mathbb{P}_{x \in H^{\min}}[f(x) = 1]| \leq \epsilon.$$

Moreover, the characteristic function of H is in $C(\mathcal{F}, k)$.

Proof: Set $t = \lceil 8/\epsilon\delta \rceil$, and $\epsilon' = \epsilon/4t$. Apply Theorem 3.2 with ϵ' rather than ϵ to obtain an integer $k \leq 1/((\epsilon')^2\delta) = O(1/\delta^2\epsilon^4)$ and a partition $\mathcal{P}$. For $i = 1, \ldots, t$, let H_i be the union of all sets $P \in \mathcal{P}$ for which $\mathbb{P}_{x \in P}[g(x) = 1]$ is in the interval $[(i-1)/t, i/t)$. (We also include in H_t the sets P on which g is always 1.) By taking t to be even, we can ensure that all sets P in H_i satisfy $\mathrm{maj}(P) = \mathrm{maj}(H_i)$.

We will show that at least one of sets H_i satisfies the conclusions of Theorem 3.4. Note that the characteristic function of each H_i is computable in $C(\mathcal{F}, k)$, since the partition $\mathcal{P}$ is defined in terms of k functions from $\mathcal{F}$.

Let $D'_{\min}$ be the distribution on $\{1, \ldots, t\}$ that selects i with probability proportional to $|H_i^{\min}|$. Note that:
$$\mathbb{P}_{i \in D'_{\min}}\left[|H_i^{\min}| < \frac{4}{\epsilon t} \cdot |H_i|\right] \leq \frac{\sum_i (4/\epsilon t) \cdot |H_i|}{\sum_i |H_i^{\min}|} \leq \frac{(4/\epsilon t) \cdot 2^n}{\delta \cdot 2^n} \leq \frac{1}{2},$$
because $t \geq 8/\epsilon\delta$. Thus, there exists an $H = H_j$ such that $|H^{\min}| \geq (4/\epsilon t) \cdot |H|$ and $\mathbb{P}_{i \in D'_{\min}}[H_i = H] \geq 1/2t$. Observe that the former condition implies that all sets P in H have the same number of minority elements up to a *relative error* of $(1/t)/(4/\epsilon t) = \epsilon/4$, and the latter condition implies that $|\mathrm{H}_\infty| \geq (1/2t) \cdot \sum_i |H_i| \geq (\delta/2t) \cdot 2^n = \Omega(\epsilon\delta^2)$.

Suppose, for sake of contradiction, that
$$\left|\mathbb{P}_{x \in H_i^{\mathrm{maj}}}[f(x) = 1] - \mathbb{P}_{x \in H_i^{\min}}[f(x) = 1]\right| > \epsilon.$$

Then:
$$\begin{aligned}
&\mathop{\mathbb{E}}_{P \in D_{\min}}\left[|\mathbb{P}_{x \in P^{\mathrm{maj}}}[f(x) = 1] - \mathbb{P}_{x \in P^{\min}}[f(x) = 1]|\right] \\
&= \mathop{\mathbb{E}}_{i \in D'_{\min}} \mathop{\mathbb{E}}_{P \in D_{\min} | P \subseteq H_i}\left[|\mathbb{P}_{x \in P^{\mathrm{maj}}}[f(x) = 1] - \mathbb{P}_{x \in P^{\min}}[f(x) = 1]|\right] \\
&\geq \mathop{\mathbb{E}}_{i \in D'_{\min}}\left[\left|\mathop{\mathbb{E}}_{P \in D_{\min} | P \subseteq H_i}\left[\mathbb{P}_{x \in P^{\mathrm{maj}}}[f(x) = 1] - \mathbb{P}_{x \in P^{\min}}[f(x) = 1]\right]\right|\right] \\
&\geq \frac{1}{2t} \cdot \left|\mathop{\mathbb{E}}_{P \in D_{\min} | P \subseteq H}\left[\mathbb{P}_{x \in P^{\mathrm{maj}}}[f(x) = 1] - \mathbb{P}_{x \in P^{\min}}[f(x) = 1]\right]\right| \\
&\geq \frac{1}{2t} \cdot \left(\left|\mathbb{P}_{x \in H_i^{\mathrm{maj}}}[f(x) = 1] - \mathbb{P}_{x \in H_i^{\min}}[f(x) = 1]\right| - \epsilon/2\right) \\
&> \frac{\epsilon}{4t} = \epsilon',
\end{aligned}$$
contradicting the property of $\mathcal{P}$. To see the second-to-last inequality, observe that selecting $P \subseteq H$ with probability proportional to $|P^{\min}|$ and then selecting $x \in P^{\min}$ is equivalent to selecting

$x \in H^{\min}$, so the second terms are equal to each other. However, the first terms, which refer to P^{maj} and H^{maj} do not correspond exactly, since in the former case, we still select P with probability proportional to its number of minority elements, and not the number of majority elements. But since all sets in H have the same fraction of minority elements up to a relative error of $\epsilon/4$, these two distributions will have statistical distance at most $\epsilon/2$. ■

4 Characterizing Pseudoentropy

Now we discuss some additional applications to of the Dense Model Theorems proved earlier.

In this section, we show that if a distribution is dense inside a pseudorandom distribution, then it has large "pseudoentropy" in the sense of Håstad, Impagliazzo, Levin and Luby [HILL]. In the nonuniform setting (i.e. when the distinguishers are circuits), the implication is an easy consequence of our main result. Using the techniques of Barak, Shaltiel, and Wigderson [BSW], a form of this implication can also be proved in the uniform case, when the distinguishers are probabilistic algorithms instead of circuits.

We then apply a Dense Model Theorem to graphs, taking our universe X to be the set of all edges in the complete graph on n vertices and distinguishers to be cuts in the graph. In this setting, a graph is "pseudorandom" if the probability that a random edge in the graph crosses a given cut is approximately the same as in case of the complete graph. Sparse expanders do satisfy this property, and actually a slightly weaker condition suffices for our purposes. We show that the Tao-Ziegler Dense Model Theorem directly implies an analogue of Szemerédi's Regularity Lemma for dense subgraphs of such pseudorandom graphs. Regularity Lemmas in the context of sparse graphs were first proved by Kohayakawa and Rödl (for a survey, see [KR]).

4.1 Nonuniform Distinguishers

Recall that two distributions X and Y are ϵ-indistinguishable for a class of (boolean) distinguishers $\mathcal{F}$ if

$$\forall f \in \mathcal{F} \quad |\mathbb{P}[f(Y) = 1] - \mathbb{P}[f(X) = 1]| < \epsilon.$$

(In this section, we are interested in the class $\mathcal{F}_s$ consisting of all boolean circuits of size at most s.) We say that a distribution X on $\{0,1\}^n$ is ϵ-pseudorandom for $\mathcal{F}$ if X is ϵ-indistinguishable from U_n, the uniform distribution on $\{0,1\}^n$. Håstad, Impagliazzo, Levin, and Luby [HILL] generalized the concept of pseudorandomness to the following more general notion of pseudoentropy.

Definition 4.1 ([HILL]) [6] *A distribution D on $\{0,1\}^n$ has ϵ-pseudoentropy k for $\mathcal{F}$ if there exists a distribution M on $\{0,1\}^n$ such that*

1. $\mathrm{H}_\infty(M) \geq k$. *That is,* $\mathbb{P}[M = x] \leq 2^{-k}$ *for all* x.
2. M *and* D *are* ϵ*-indistinguishable for* $\mathcal{F}$.

Since U_n is the unique distribution on $\{0,1\}^n$ with min-entropy at least n, having ϵ-pseudoentropy n is equivalent to being ϵ-pseudorandom. Now, to see the connection of this notion with Dense

[6] Håstad et al. actually only require that the distribution is computationally indistinguishable from some distribution with *Shannon entropy* at least k, but it is common to work with min-entropy instead. Indeed, even the constructions of Håstad et al. work by first converting Shannon entropy into min-entropy by taking many independent copies of the distribution.

Model Theorems, observe that a distribution M is δ-dense in U_n iff $\mathrm{H}_\infty(M) \geq n - \log(1/\delta)$. Thus we have:

Proposition 4.2 *Fix a class of distinguishers $\mathcal{F}$. A distribution D is ϵ-indistinguishable from a δ-dense distribution M if and only if D has ϵ-pseudoentropy $n - \log(1/\delta)$.*

Thus, Dense Model Theorems say that if a distribution D is δ-dense in a pseudorandom set, then D has pseudoentropy $n - \log(1/\delta)$. Specifically, using Theorem 2.1, we get:

Corollary 4.3 *Let R be a distribution on $\{0,1\}^n$ that is ϵ-indistinguishable from U_n by circuits of size s, and let D be δ-dense in R. Then D has $\Omega(\epsilon/\delta)$-pseudoentropy $n - \log(1/\delta)$ for circuits of size $\Omega(s \cdot \epsilon^2/\log(1/\epsilon\delta))$.*

We observe that the reverse implication also holds:

Proposition 4.4 *If a distribution D on $\{0,1\}^n$ has ϵ-pseudoentropy $n - \log\frac{1}{\delta}$ for circuits of size s, then D is δ-dense in some distribution R that is ϵ/δ-pseudorandom for circuits of size s.*

Proof: By hypothesis, D is ϵ-indistinguishable from some distribution M that is δ-dense in the uniform distribution U_n. We can write $U_n = \delta M + (1-\delta)N$ for some distribution N. Define $R = \delta D + (1-\delta)N$; note that D is δ-dense in R. Since any distinguisher that ϵ/δ-distinguishes R from U_n must ϵ-distinguish D from M, we can conclude that R is ϵ/δ-pseudorandom. ■

Thus, we have an *equivalence* between being δ-dense in a pseudorandom set and having pseudoentropy $n - \log(1/\delta)$. One important comment, however, is that both directions only give nontrivial results when $\delta \gg \epsilon$. Typically, $\epsilon > 1/\mathrm{poly}(s) \gg 1/2^n$, so the equivalence only characterizes the case when discussing pseudoentropy $n - \log(1/\delta)$ that is very high (and says nothing about, say, pseudoentropy $n/2$).

4.2 Uniform Distinguishers

In this section, we extend our treatment to uniform distinguishers, by using an idea of Barak, Shaltiel, and Wigderson [BSW].

First, we provide recall standard definitions of indistinguishability and pseudorandomness in the uniform setting. These definitions are necessarily asymptotic. We refer to a sequence of distributions $\{Y_n\}$ where Y_n is distributed over $\{0,1\}^n$, as an ensemble. We say that an ensemble $Y = \{Y_n\}$ is δ-dense in $Z = \{Z_n\}$ if for every n, Y_n is δ-dense in Z_n. For functions $\epsilon : \mathbb{N} \to [0,1]$, $t : \mathbb{N} \to \mathbb{N}$, we say that two ensembles X and Y are ϵ-indistinguishable for time t if for every (uniform) probabilistic algorithm A running in time at most $t(n)$, we have

$$|\mathbb{P}[A(X_n) = 1] - \mathbb{P}[A(Y_n) = 1]| \leq \epsilon(n)$$

for all sufficiently large n. We say that an ensemble X is ϵ-pseudorandom for time t if it is ϵ-indistinguishable from the uniform ensemble $U = \{U_n\}$.

Now, we consider how to define pseudoentropy in the uniform setting. A first attempt is to say that an ensemble D has pseudoentropy $k(n)$ if it is indistinguishable from an ensemble $M = \{M_n\}$ where $\mathrm{H}_\infty(M_n) \geq k(n)$ for all n. However, Barak et al. [BSW] realized that a weaker definition is more appropriate for the uniform setting. In this definition, we allow the model distribution M to depend on the *description size* of the distinguisher A.

Definition 4.5 ([BSW]) *For functions $\epsilon : \mathbb{N} \to [0,1]$, $k : \mathbb{N} \to \mathbb{N}$, $t : \mathbb{N} \to \mathbb{N}$, an ensemble of distributions $D = \{D_n\}$ is said to have ϵ-*pseudoentropy *k for time t if there is a growing function h such that for every n, there is a distribution M_n with entropy at least $k(n)$ that is indistinguishable from D_n by all machines A with description size up to $h(n)$ and running time at most $t(n)$. Formally,*

$$\exists h \in \omega(1)\ \forall n\ \exists M_n, \mathrm{H}_\infty(M_n) \geq k(n)\ \forall A, |A| \leq h(n) \quad |\mathbb{P}\left[A(D_n) = 1\right] - \mathbb{P}\left[A(M_n) = 1\right]| \leq \epsilon$$

Even for this relaxed definition, it can be shown that ϵ-pseudorandomness is equivalent to having ϵ-pseudoentropy n.

We now want to show that if an ensemble $\{D_n\}$ is δ-dense inside a pseudorandom distribution $\{R_n\}$, then D_n has pseudoentropy at least $k(n) = n - \log 1/\delta$. We state the result below in the contrapositive.

Theorem 4.6 *Let $k, t : \mathbb{N} \to \mathbb{N}$, $\delta, \epsilon : \mathbb{N} \to [0,1]$ be any functions. Let $D = \{D_n\}$ and $R = \{R_n\}$ be two ensembles of distributions such that D is δ-dense in R.*

Suppose that D does not have ϵ-pseudoentropy $n - \log(1/\delta)$. That is, for every $h \in \omega(1)$, there is some n such that for every M_n which is δ-dense in U_n, one can find a machine A running in time $t(n)$ with $|A| < h(n)$ and

$$\mathbb{P}\left[A(D_n) = 1\right] - \mathbb{P}\left[A(M_n) = 1\right] | \geq \epsilon \tag{10}$$

Then R is not pseudorandom. Specifically, for the same n as above, one can find a machine $|A'|$ running in time at most $O((1/\epsilon(n)) \cdot t(n) \log t(n) \cdot 2^{h(n)})$ with $|A'| \leq 2^{O(h(n))}/\epsilon(n)$ such that

$$|\mathbb{P}\left[A'(R_n) = 1\right] - \mathbb{P}\left[A'(U_n) = 1\right]| \geq \epsilon'(n)$$

where $\epsilon' = \Omega(\epsilon\delta)$.

Observe that by taking ϵ to be *constant* and $h(n)$ to be a sufficiently slowly growing function of n, the description size of A' can be bounded by any growing function of n (and the running time of A' made to be smaller than $t(n) \log^2 t(n)$), violating uniform $\epsilon'(n)$-pseudorandomness of R.

Proof: We can assume without loss of generality that the encoding of A and $\bar{A}$ (the machine with output as complement of A) have the same description size and hence the class of machines with $|A| \leq h(n)$ is closed under complementation. By arguments identical to Claim 2.2, there is a distribution $\mathcal{A}$ over machines of description size at most $h(n)$ such that for all M_n δ-dense in U_n,

$$\mathbb{E}[\mathcal{A}(D_n)] - \mathbb{E}[\mathcal{A}(M_n)] \geq \epsilon$$

$\mathcal{A}$ is a distribution over at most $2^{h(n)}$ machines. We now round all the probabilities in $\mathcal{A}$ to the nearest multiple of $2^{-(h(n)+s)}$ to obtain a new distribution $\tilde{\mathcal{A}}$. For $s = \log(8/\epsilon)$, this cause a loss of at most $\epsilon/2$ in the distinguishing probability. Thus,

$$\mathbb{E}[\tilde{\mathcal{A}}(D_n)] - \mathbb{E}[\tilde{\mathcal{A}}(M_n)] \geq \epsilon/2$$

By Claim 2.3, there exists a threshold t such that

$$\mathbb{P}_{x \in D_n}\left[\mathbb{E}\,\tilde{\mathcal{A}}(x) \geq t + \epsilon/4\right] > \mathbb{P}_{x \in M_n}\left[\mathbb{E}\,\tilde{A}(x) > t\right] + \epsilon/4$$

Where the probabilities are over the distributions D_n and M_n and the expectations over $\tilde{A}$.

Taking S to be the uniform distribution over the set of $\delta \cdot 2^n$ inputs where $\tilde{A}$ values are the highest and applying the above with $M_n = S$, we get

$$\mathbb{P}\left[\mathbb{E}\,\tilde{\mathcal{A}}(U_n) > t\right] = \delta \mathbb{P}\left[\mathbb{E}\,\tilde{\mathcal{A}}(S) > t\right]$$

while

$$\mathbb{P}_{x \in R_n}\left[\mathbb{E}\,\tilde{\mathcal{A}}(x) > t + \epsilon/4\right] > \delta \mathbb{P}_{x \in S}\left[\mathbb{E}\,\tilde{\mathcal{A}}(x) > t\right] + \epsilon\delta/4$$

We now construct the machine A' as follows. The description of A' contains the descriptions and probabilities in $\tilde{\mathcal{A}}$ of all the $2^{h(n)}$ machines in the support of $\tilde{\mathcal{A}}$. We thus have $|A'| = O(h(n)2^{2h(n)+\log(1/\epsilon)})$. A' then runs all the machines and computes $\mathbb{E}\,\tilde{\mathcal{A}}(x)$ for a given input x. It then outputs 1 if this quantity is greater than t' and 0 otherwise, where t' is the smallest multiple of $\epsilon/8$ greater than t. By the above, A' distinguishes R_n and U_n with probability at least $\epsilon\delta/4$.

Finally, note that the running time of A' can be controlled by choosing $h(n)$ to be small enough. Specifically, if A runs in time $t(n)$, and $h(n)$ is chosen such that $2^{3h(n)} < w(n)$, then the running time of A' is $O(\frac{1}{\epsilon}t(n)\log(t(n))w(n))$. ■

We note that one can consider an even weaker notion of pseudoentropy than Definition 4.5, where we allow the 'model distribution' M_n to depend on the machine A (and not only its description size). For such a definition, the restriction to constant ϵ is no longer necessary.

5 A Dense Model Theorem for Graphs

In this section, we apply our Dense Model Theorem to graphs, taking our universe X to be the set of all edges in the complete graph on n vertices and distinguishers to be cuts in the graph. In this setting, a graph is "pseudorandom" if the probability that a random edge in the graph crosses a given cut is approximately the same as in case of the complete graph. Sparse expanders do satisfy this property, and actually a slightly weaker condition suffices for our purposes. We show that the Tao-Ziegler Dense Model Theorem directly implies an analogue of Szemerédi's Regularity Lemma for dense subgraphs of such pseudorandom graphs. Regularity Lemmas in the context of sparse graphs were first proved by Kohayakawa and Rödl (for a survey, see [KR]); in Section 5.3, we show that our sparse regularity lemmas are equivalent to the known ones (despite being formulated using slightly different notions of pseudorandomness and regularity).

5.1 The Dense Model Theorem

We start by defining our family of distinguishers and what it means to be pseudorandom with respect to those distinguishers. We view an undirected graph $G = (V, E)$ as a subset of the universe $X = V \times V$. An edge $\{u, v\}$ in the graph is counted as *both* pairs (u, v) and (v, u). We refer to the uniform distribution over all the ordered pairs corresponding to edges in G as U_G. Our family of distinguishers $\mathcal{F}$ will consist of functions $f_{S,T} : V \times V \to \{0, 1\}$ for $S, T \subseteq V$, $S \cap T = \emptyset$ defined as

$$f_{S,T}(u, v) = 1 \Leftrightarrow u \in S \text{ and } v \in T$$

Note that the class of distinguishers is closed under products since $f_{S_1,T_1} \cdot f_{S_2,T_2} = f_{S_1 \cap S_2, T_1 \cap T_2}$. Thus, a distribution that fools all distinguishers in $\mathcal{F}$ also fools products of functions from $\mathcal{F}$.

Intuitively, the distinguishers check how often a pair (u, v) selected according to some distribution crosses a cut from S and T. Hence, for a graph G to be pseudorandom, this probability must be the same whether we draw the pairs from the distribution U_G defined by the edges of the graph or from the uniform distribution over X. When the probability differs by at most η, we call the graph η-pseudorandom.

Definition 5.1 *We say a graph G is η-*pseudorandom *if for every pair of disjoint*[7] *sets S and T*

$$\left|\frac{e_G(S,T)}{2|E(G)|} - \frac{|S||T|}{n^2}\right| < \eta$$

where $e(S,T)$ denotes the number of edges in G with one endpoint in S and the other in T and $E(G)$ denotes the set of edges in G.

Note that the quantity on the left in the definition of pseudorandomness is exactly the probability with which $f_{S,T}$ distinguishes U_G and U_X and hence η-pseudorandomness is equivalent to being η-indistinguishable by functions in $\mathcal{F}$.

We remark that expanders (with sufficient eigenvalue gap) are a special case of the above definition. If G is a d-regular expander with $|V| = n$ and second eigenvalue $\lambda \leq \eta d$, then by the Expander Mixing Lemma

$$\left|\frac{e_G(S,T)}{2|E(G|} - \frac{|S||T|}{n^2}\right| \;\leq\; \frac{\lambda}{d} \cdot \sqrt{\frac{|S||T|}{n^2}} \;\leq\; \eta$$

We now prove a Dense Model Theorem for dense subgraphs of pseudorandom graphs. For proving this, it is easy to obtain a *distribution* over dense *directed* graphs as a model using Theorem 2.1. We then convert this model to a single dense undirected graph using standard symmetrization and sampling arguments.

Theorem 5.2 (Dense Model Theorem for Graphs) *Let G be an η-pseudorandom graph and let H be a subgraph of G with $\delta|E(G)|$ edges. Then there exists a graph H' with at least $\delta n^2/2$ edges such that for all pairs of disjoint sets $S, T \subseteq V$*

$$\left|\frac{e_H(S,T)}{2|E(H)|} - \frac{e_{H'}(S,T)}{2|E(H')|}\right| < \epsilon$$

provided $\eta = \exp(-\text{poly}(1/\epsilon, 1/\delta))$.

Proof: We first notice that for any $f_{S,T} \in \mathcal{F}$, the probability of distinguishing U_G from U_X is very small.

$$\left|\mathop{\mathbb{E}}_{(u,v)\in U_G}[f_{S,T}(u,v)] - \mathop{\mathbb{E}}_{(u,v)\in U_X}[f_{S,T}(u,v)]\right| = \left|\frac{e_G(S,T)}{2|E(G)|} - \frac{|S||T|}{n^2}\right| < \eta \leq \exp(-poly(1/\epsilon, 1/\delta))$$

By closure of $\mathcal{F}$ under products, this is also true for products of functions from $\mathcal{F}$. By Theorem 2.1, there exists a a δ-dense distribution M over pairs $(u, v) \in V \times V$. such that for all $f_{S,T} \in \mathcal{F}$

$$\left|\frac{e_H(S,T)}{2|E(H)|} - \mathop{\mathbb{E}}_{(u,v)\in M}[f_{S,T}(u,v)]\right| < \epsilon/2$$

[7] It would be more natural to define pseudorandomness for any pair of sets rather than only for disjoint sets. However, the definition in terms of disjoint sets corresponds more directly to the formulations used in the literature on regularity lemmas.

We think of M as specifying a *weighted* and *directed* graph. We first convert the graph to an undirected one. Consider the distribution M^T with $\mathbb{P}\left[M^T = (u,v)\right] = \mathbb{P}\left[M = (v,u)\right]$. Note that $\mathbb{E}_{(u,v)\in M^T}[f_{S,T}(u,v)] = \mathbb{E}_{(u,v)\in M}[f_{T,S}(u,v)]$, so M^T is an equally good dense model of H. Indeed,

$$\left|\frac{e_H(S,T)}{2|E(H)|} - \mathop{\mathbb{E}}_{(u,v)\in M^T}[f_{S,T}(u,v)]\right| = \left|\frac{e_H(T,S)}{2|E(H)|} - \mathop{\mathbb{E}}_{(u,v)\in M}[f_{T,S}(u,v)]\right| < \epsilon/2$$

for every $f_{S,T} \in \mathcal{F}$. We then consider the distribution $M' = \frac{1}{2}M + \frac{1}{2}M^T$, which by the above argument, is also a good dense model while having the additional property that $\mathbb{P}\left[M' = (u,v)\right] = \mathbb{P}\left[M' = (v,u)\right]$.

Following [Imp], we go from the weighted graph to an unweighted by a sampling argument. Let H' be constructed by picking each edge $\{u,v\}$ with probability $(\mathbb{P}\left[M' = (u,v)\right] + \mathbb{P}\left[M' = (v,u)\right]) \cdot \delta n^2/2$. The expected number of edges in H' is then $\delta n^2/2$. By Chernoff bounds, the number of edges is very close to $\delta n^2/2$ with high probability. In particular

$$\mathbb{P}\left[\left||E(H')| - \frac{1}{2}\delta n^2\right| > \frac{\epsilon\delta}{8}n^2\right] \leq 2\exp(-\epsilon^2\delta n^2/64)$$

Consider any two disjoint sets S and T. We have the distinguishing probability of $f_{S,T}$ on the distribution M' (using the fact that it is symmetric) and the sample H' as

$$\left|\mathop{\mathbb{E}}_{(u,v)\in M'}[f_{S,T}(u,v)] - \frac{e_{H'}(S,T)}{2|E(H')|}\right| = \left|\frac{\mathbb{E}\, e_{H'}(S,T)}{\delta n^2} - \frac{e_{H'}(S,T)}{2|E(H')|}\right|$$

where the second expectation is over the sampling process. To bound the probability that the above quantity is large, we can restrict ourselves to the case when $|E(H')|$ is about $\delta n^2/2$ by using the previous Chernoff bound.

$$\begin{aligned}
&\mathbb{P}\left[\left|\frac{\mathbb{E}\, e_{H'}(S,T)}{\delta n^2} - \frac{e_{H'}(S,T)}{2|E(H')|}\right| > \frac{\epsilon}{2}\right] \\
\leq\ &\mathbb{P}\left[\frac{\mathbb{E}\, e_{H'}(S,T)}{\delta n^2} - \frac{e_{H'}(S,T)}{(1+\epsilon/4)\delta n^2} > \frac{\epsilon}{2}\right] + \mathbb{P}\left[\frac{e_{H'}(S,T)}{(1-\epsilon/4)\delta n^2} - \frac{\mathbb{E}\, e_{H'}(S,T)}{\delta n^2} > \frac{\epsilon}{2}\right] \\
&+\ \mathbb{P}\left[\left||E(H')| - \frac{1}{2}\delta n^2\right| > \frac{\epsilon\delta}{8}n^2\right] \\
\leq\ &\exp(-\epsilon^2\delta^2 n^4/32n^2) + \exp\left(-\epsilon^2\delta^2 n^4/32n^2\right) + 2\exp\left(-\epsilon^2\delta n^2/64\right) \\
\leq\ &4\exp\left(-\epsilon^2\delta^2 n^2/64\right)
\end{aligned}$$

where each term is individually bounded by Chernoff bounds. By a union bound, with probability at least $1 - 2^{2n}\cdot 4\exp(-\frac{\epsilon^2\delta^2 n^2}{64}) - \exp(-\frac{\epsilon^2\delta n^2}{64}) = 1 - \exp(-\Omega(n))$,[8] we have for all disjoint $S, T \subseteq V$

$$\begin{aligned}
\left|\frac{e_H(S,T)}{2|E(H)|} - \frac{e_{H'}(S,T)}{2|E(H')|}\right| &\leq \left|\frac{e_H(S,T)}{2|E(H)|} - \mathop{\mathbb{E}}_{(u,v)\in M'}[f_{S,T}(u,v)]\right| + \left|\mathop{\mathbb{E}}_{(u,v)\in M'}[f_{S,T}(u,v)] - \frac{e_{H'}(S,T)}{2|E(H')|}\right| \\
&< \epsilon/2 + \epsilon/2 = \epsilon.
\end{aligned}$$

[8] We may assume that $\epsilon, \delta > 1/\log n$, for otherwise $\eta < 1/n^2$ and the only η-pseudorandom graph is the complete graph with self-loops.

Finally, we would like H' above to have more than $\delta n^2/2$ edges and not just more than $(1-\epsilon/4)\delta n^2/2$ which is what the Chernoff bound gives. However, there must be at least one "good" H' with more than $\delta n^2/2$ edges, since otherwise the expected number of edges is at most

$$(\delta n^2/2 - 1)(1 - \exp(-\Omega(n))) + n^2 \exp(-\Omega(n)) < \delta n^2/2$$

which is a contradiction. ■

5.2 Regularity lemmas for subgraphs of pseudorandom graphs

We now proceed to variants of the regularity lemma for subgraphs of pseudorandom graphs described earlier. Roughly speaking, regularity lemmas state that a graph can be divided into a constant number of "pieces" such that the edges between most pairs of pieces are very uniformly distributed. This uniformity is measured by the concept of regularity.

Definition 5.3 *Given a graph H and $\epsilon > 0$, we say that a pair of disjoint subsets $A, B \subseteq V$ is* ϵ-regular *in H if for every $S \subseteq A, |S| \geq \epsilon|A|$ and $T \subseteq B, |T| \geq \epsilon|B|$, we have*

$$\left| \frac{e_H(A,B)}{|A||B|} - \frac{e_H(S,T)}{|S||T|} \right| \leq \frac{\epsilon |E(H)|}{n^2}$$

When G is the complete graph and H is any δ-dense subgraph (i.e. any graph with $\delta n^2/2$ edges), we are in the setting of Szemerédi's regularity lemma, which says that H can be partitioned into a constant number of subsets such that most pairs of subsets are regular.

Theorem 5.4 (Regularity lemma for dense graphs) *Let $\epsilon, \delta > 0$ and $k_0 \geq 1$ be given. Then there exists a constant $K = K(\epsilon, \delta, k_0) \geq k_0$ such that if H is graph $|E_H| \geq \delta n^2$, then there exists a partition of V into disjoint sets $A_0, \ldots, A_k$ for $k_0 \leq k \leq K$ with the following properties*

1. *$|A_0| \leq \epsilon n$*
2. *$|A_1| = \ldots = |A_k|$*
3. *At most $\epsilon \binom{k}{2}$ pairs (A_i, A_j) for $1 \leq i < j \leq k$ are not ϵ-regular (w.r.t the complete graph)*

We now state and prove a version of the regularity lemma for sparse graphs. For simplicity, we only state the non-bipartite version.

Theorem 5.5 (Regularity lemma for sparse graphs) *Let $\epsilon, \delta > 0$ and $k_0 \geq 1$ be given. Then there exist constants $\eta = \eta(\epsilon, \delta, k_0) > 0$ and $K = K(\epsilon, \delta, k_0) \geq k_0$ such that if G is η-pseudorandom and H is any subgraph of G with $|E_H| \geq \delta |E_G|$, then there exists a partition of V into disjoint sets $A_0, \ldots, A_k$ for $k_0 \leq k \leq K$ with the following properties*

1. *$|A_0| \leq \epsilon_1 n$*
2. *$|A_1| = \ldots = |A_k|$*
3. *At most $\epsilon \binom{k}{2}$ pairs (A_i, A_j) for $1 \leq i < j \leq k$ are not ϵ-regular with respect to G.*

Proof: We begin by applying the Dense Model Theorem for graphs (Theorem 5.2) to H to obtain a δ-dense subgraph H' of the complete graph which is γ-indistinguishable from H by cuts. We can then apply Szemerédi's regularity lemma to H' to find a partition $A_0, \ldots, A_k$ of V that satisfies the required conditions in H' with pairs being β-regular, for $\beta < \epsilon$ to be chosen later.

We now show that the above partition, which is β-regular for H', also satisfies the regularity conditions in H. The proof is essentially by triangle inequality. We show that for $S \subseteq A$ and $T \subseteq B$, the fractions of total edges of the graph that go between A and B (and also between S and T) are close in H and H' since they are indistinguishable by cuts. However, in H' the "density" of edges between the pair (A, B) must be close to that for the pair (S, T) by the regularity of the partition. Combining these two arguments gives that the partition must be regular in H also.

Let A, B be an β-regular pair for H' in the parition and let S, T be subsets of A and B respectively such that $|S| \geq \beta|A|$ and $|T| \geq \beta|B|$. Using the indinstinguishability of H and H', we get that

$$\left|\frac{e_H(A,B)}{2|E(H)|} - \frac{e_{H'}(A,B)}{2|E(H')|}\right| \leq \gamma \implies \left|\frac{e_H(A,B)}{|A||B|} \cdot \frac{|E(H')|}{|E(H)|} - \frac{e_{H'}(A,B)}{|A||B|}\right| \leq \frac{2\gamma|E(H')|}{|A||B|} \leq 4\gamma K^2$$

Here K is an upper bound on the number of partitions appearing in the (dense) regularity lemma and we use the fact that $|A|, |B| \leq (1-\beta)n/K \leq n/2K$ and $|E(H')| \leq n^2/2$. A similar relation must also hold for S and T.

$$\left|\frac{e_H(S,T)}{|S||T|} \cdot \frac{|E(H')|}{|E(H)|} - \frac{e_{H'}(S,T)}{|S||T|}\right| \leq \frac{2\gamma|E(H')|}{|S||T|} \leq 4\frac{\gamma K^2}{\beta^2}$$

Also, by regularity of the parition for H', we get

$$\left|\frac{e_{H'}(S,T)}{|S||T|} - \frac{e_{H'}(A,B)}{|A||B|}\right| \leq \frac{\beta|E(H')|}{n^2}$$

Choosing $\gamma = \epsilon\delta \cdot \beta^2/12K^2$, $\beta < \epsilon\delta/3$ and $\eta = \exp(-poly(\frac{1}{\delta}, \frac{1}{\gamma}))$, we get by triangle inequality that

$$\left|\frac{e_H(S,T)}{|S||T|} - \frac{e_H(A,B)}{|A||B|}\right| \leq \epsilon\delta\frac{|E(H)|}{|E(H')|} \leq \frac{\epsilon|E(H)|}{n^2}$$

This shows that any β-regular pair in the partition for H' must also be an ϵ-regular pair in the parition for H. Since $\beta < \epsilon$, this gives an ϵ-regular partition of H. ■

We can also show sparse analogues of the weak regularity lemma. As opposed to Theorem 5.5 which requires most pairs in the partition to be regular, weak regularity only requires that pairs be regular on average.

Definition 5.6 *For a graph H, we say that a parition $A_0, \ldots A_k$ is* weakly ϵ-regular *if for all disjoint sets S and T, with $|S|, |T| > \epsilon n$*

$$\left|\frac{e_H(S,T)}{|S||T|} - \sum_{i,j}\frac{|S \cap A_i||T \cap A_j|}{|S||T|}\frac{e_H(A_i,A_j)}{|A_i||A_j|}\right| < \frac{\epsilon|E(H)|}{n^2}$$

While the number of parts in the regularity lemma can be a tower of exponentials in $poly(1/\epsilon, 1/\delta)$, every dense graph admits a weakly regular partition with relatively fewer parts.

Lemma 5.7 (Weak regularity lemma for dense graphs) *Let $\epsilon, \delta > 0$ be given and let H be a graph with at least δn^2 edges. Then, there exists a weakly ϵ-regular partition $A_0, \ldots, A_k$ of the vertices of H, with $k \leq \exp(poly(1/\epsilon, 1/\delta))$.*

Using the same arguments as in the previous section, we can show that for sufficiently small η, any δ-dense subgraph of an η-pseudorandom graph admits a weakly regular partition. We state the lemma without a proof.

Lemma 5.8 (Weak regularity lemma for sparse graphs) *Let $\epsilon, \delta > 0$ be given. Then there exists a constant $\eta = \eta(\epsilon, \delta) > 0$ such that if G is η-pseudorandom and H is any subgraph of G with $|E_H| \geq \delta |E_G|$, then there exists a weakly ϵ-regular partition $A_0, \ldots, A_k$ of the vertices of H, with $k \leq exp(exp(poly(1/\epsilon, 1/\delta)))$.*

5.3 Comparison to Kohayakawa's version

The known versions of the sparse regularity lemma in the literature are different in the sense that they use slightly different notions of pseudorandomness and regularity than the ones in our lemmas. In this section, we show that the notions are equivalent and hence our arguments provide a proof of the version of the regularity lemma first stated by [Koh].

The following concept of uniformity defined by Kohayakawa is the notion of pseudorandomness used in literature on regularity lemmas.

Definition 5.9 *We say a graph $G = (V, E)$ is η-*uniform *if there exists some $p \in [0, 1]$ such that for all $S, T \subseteq V$ with $S \cap T = \emptyset$ $|S|, |T| \geq \eta |V|$, we have*

$$|e_G(S, T) - p|S||T|| \;\leq\; \eta p |S||T|$$

In fact, it is easy to see that p must always be (roughly) $2|E(G)|/n^2$. It would be immediate, if we could take $|S| = |T| = |V|$ in the above definition. However, since S and T must be disjoint, we produce sets S and $V \setminus S$, each of which contain about half the vertices, with about half the total number of edges crossing the cut $(S, V \setminus S)$.

Fact 5.10 *Let $G = (V, E)$ be a graph on n vertices. Then for all $\alpha > 0$ (and sufficiently large $n, |E|$) , there is a cut $(S, V \setminus S)$ such that $|\,|S| - n/2| \leq \alpha n$ and $|e(S, V \setminus S) - |E|/2| \leq \alpha |E|$.*

Using S and T from the above fact, and sufficiently small α gives that for an η-uniform graph $\left(\frac{1-\eta}{1-4\alpha^2}\right) \frac{2|E(G)|}{n^2} < p < \left(\frac{1+\eta}{1-4\alpha^2}\right) \frac{2|E(G)|}{n^2}$. Adjusting η if necessary, we use $p = 2|E(G)|/n^2$ in the rest of this section.

We now show that η-uniformity implies $O(\eta)$-pseudorandomness and is hence sufficient to conclude our version of the regularity lemma.

Claim 5.11 *If $G = (V, E)$ is η-uniform then it is also 10η-pseudorandom.*

Proof: We need to show that for all disjoint $S, T \subseteq V$

$$\left| \frac{e(S, T)}{2|E(G)|} - \frac{|S||T|}{n^2} \right| < 10\eta$$

If $|S|, |T| > \eta n$, then the statement follows easily from the η-uniformity of the graph (taking $p = 2|E(G)|/n^2$).

$$\left|\frac{e(S,T)}{2|E(G)|} - \frac{|S||T|}{n^2}\right| = \left|e(S,T) - \frac{2|E(G)|}{n^2}|S||T|\right| \frac{1}{2|E(G)|} \leq \eta \frac{2|E(G)|}{n^2}|S||T| \cdot \frac{1}{2|E(G)|} \leq \eta$$

If S or T have size less than ηn, then $|S||T|/n^2 < \eta$ and it suffices to show that $e(S,T) < 20\eta|E(G)|$. This can be separately analyzed in the following cases:

- If $|V \backslash (S \cup T)| \geq 3\eta n$, we can add (different) new vertices to both S and T, getting disjoint sets S' and T' of size ηn each. Then by uniformity, $e(S,T) \leq e(S',T') \leq 2\eta|E(G)|$.
- In the case when $|V \backslash (S \cup T)| < 3\eta n$, one of the sets must be very small (of size less than ηn) and the other must be very large. Let S be the smaller set. Include vertices in S (possibly moving some from T to S), to get sets S', T' with $|S'| \geq 3\eta n$. By uniformity we have that $e(S',T') < 8\eta|E(G)|$. This implies that $e(S',S') > e(S,T) - 8\eta|E(G|)$ as the edges from S, T must either cross from S' to T' or stay within S'.

  We can now use Fact 5.10 to parition S' into two parts S_1 and S_2 such that $|S_1|, |S_2| > (1-\alpha)3\eta n/2$ and $e(S_1,S_2) > (1-\alpha)(e(S,T)/2 - 4\eta|E(G)|)$. Applying the uniformity assumption to S_1 and S_2 gives (for sufficiently small α)

  $$(1-\alpha)(e(S,T)/2 - 4\eta|E(G)|) < e(S_1,S_2) < 4\eta^2|E(G)| + 2\eta|E(G)| \implies e(S,T) < 20\eta|E(G)|$$

  ■

Finally, we make a small remark about the notion of regularity in Kohayakawa's version of the regularity lemma. In his version, regularity is defined with respect to the graph G of which H is a subgraph. A pair of sets $A, B \subseteq V$ is said to be ϵ-regular if for all $S \subseteq A, |S| \geq \epsilon|A|$ and $T \subseteq B, |T| \geq \epsilon|B|$

$$\left|\frac{e_H(A,B)}{e_G(A,B)} - \frac{e_H(S,T)}{e_G(S,T)}\right| \leq \epsilon$$

However the difference is only superficial as by the η-uniformity of G, $(1-\eta)\frac{2|E(G)|}{n^2}|S||T| < e_G(S,T) < (1+\eta)\frac{2|E(G)|}{n^2}|S||T|$. Hence, the variants of the regularity lemma stated earlier are equivalent to the ones in [Koh] and [KR].

Acknowledgments

We thank Terence Tao for pointing us to [TZ] and translating the theorems from the paper. We also thank Avi Wigderson and Noga Alon for references and discussions on regularity lemmas, Russell Impagliazzo and Yishay Masour for insightful comments, and Timothy Gowers for referring us to [Gow].

References

[BSW] Boaz Barak, Ronen Shaltiel, and Avi Wigderson. Computational analogues of entropy. In *Proceedings of RANDOM'03*, pages 200–215, 2003. 4, 18, 19, 20

[Gow] Timothy Gowers. Decompositions, approximate structure, transference, and the Hahn-Banach theorem. Preprint, 2008. 4, 27

[GT] Ben Green and Terence Tao. The primes contain arbitrarily long arithmetic progressions. To appear in Annals of Mathematics. math.NT/0404188, 2004. 1, 5, 11

[HILL] Johan Håstad, Russell Impagliazzo, Leonid Levin, and Michael Luby. A pseudorandom generator from any one-way function. *SIAM Journal on Computing*, 28(4):1364–1396, 1999. 4, 18

[Imp] Russell Impagliazzo. Hard-core distributions for somewhat hard problems. In *36th Annual Symposium on Foundations of Computer Science*, pages 538–545, Milwaukee, Wisconsin, 23–25 October 1995. IEEE. 1, 3, 5, 10, 11, 15, 23

[Koh] Y. Kohayakawa. Szemerédi's regularity lemma for sparse graphs. In *FoCM '97: Selected papers of a conference on Foundations of computational mathematics*, pages 216–230, New York, NY, USA, 1997. Springer-Verlag New York, Inc. 26, 27

[KR] Y. Kohayakawa and V. Rödl. Szemerédi's regularity lemma and quasi-randomness. In *Recent advances in algorithms and combinatorics*. Springer, Berlin, 2002. 5, 18, 21, 27

[Sze] Endre Szemerédi. On sets of integers containing no k elements in arithmetic progression. *Acta Arithmetica*, 27:199–245, 1975. 1, 11

[TZ] Terence Tao and Tamar Ziegler. The primes contain arbitrarily long polynomial progressions. arXiv:math/0610050, 2006. 1, 2, 3, 5, 9, 11, 27, 28

A An Overview of the Green-Tao-Ziegler Proof

In this section we show how the proof of [TZ, Theorem 7.1], when translated into our notation, and specialized to the case of Boolean functions, gives a proof of Theorem 1.2. Here we restate the result in a more quantitative way.

Theorem A.1 *Let $\mathcal{F}$ be a class of functions $f : X \to \{0,1\}$, R be a distribution over X, and D be another distribution over $\{0,1\}^n$ that is δ-dense in R.*

Suppose that for every M that is $\delta/4$-dense in U_X there is an $f \in \mathcal{F}$ such that

$$|\mathbb{P}[f(D) = 1] - \mathbb{P}[f(M) = 1]| \geq \epsilon \tag{11}$$

Then there is a boolean function $h : X \to \{0,1\}$ of the form $h(x) = g(f_1(x), \ldots, f_k(x))$ where $k = O(\epsilon^{-3}\delta^{-1})$, $f_i \in \mathcal{F}$, such that

$$|\mathbb{P}[h(R) = 1] - \mathbb{P}[h(U_X) = 1]| \geq \gamma$$

where $\gamma = \epsilon^6\delta^4/2^{16}$.

Proof: We assume the conclusion fails and we construct a distribution M that contradicts the assumption. We construct the distribution M by partitioning "most of" the space X into disjoint sets $P_1, \ldots, P_k$. The distribution M is then constructed by choosing a P_i with probability $\mathbb{P}[D \in P_i]$ and then sampling uniformly from P_i. We formalize the notion of the partition we need in the following definition:

Definition A.2 *A partition $\mathcal{P} = (P_1, \ldots, P_k, \Omega)$ of X is called α-noisy, if*

- $\mathbb{P}[R \in \Omega] + \mathbb{P}[U_X \in \Omega] < \alpha$
- $\forall i\ \mathbb{P}[R \in P_i] \leq (1+\alpha)\mathbb{P}[U_X \in P_i]$

We then define the distribution $M = M(\mathcal{P})$ as

$$\mathbb{P}[M = a] = \begin{cases} 0 & \text{if } a \in \Omega \\ \frac{1}{1-\mathbb{P}[D\in\Omega]} \cdot \mathbb{P}[D \in P_i] \cdot \frac{1}{|P_i|} & \text{if } a \in P_i \end{cases}$$

Note that if the partition $\mathcal{P}$ is α-noisy, then

$$\mathbb{P}[M = a] = \frac{1}{1-\mathbb{P}[D\in\Omega]} \cdot \mathbb{P}[D \in P_i] \cdot \frac{1}{|P_i|} \leq \frac{1}{1-\frac{\alpha}{\delta}} \cdot \frac{1}{\delta}(1+\alpha) \cdot \frac{1}{2^n} \tag{12}$$

We will later require that $\alpha < \epsilon\delta/4$, which implies that M is $\delta/(1+\epsilon)$-dense in the uniform distribution. We then need to show that there exists an α-noisy partition $\mathcal{P}$ (for $\alpha < \epsilon\delta/4$) such that $M(\mathcal{P})$ is ϵ-indistinguishable from D.

We construct the required partition iteratively starting with the trivial partition $(X, \emptyset)$ which is also 0-noisy. At an intermediate step, suppose we have the partition $\mathcal{P} = (P_1, \ldots, P_k, \Omega)$ which is α-noisy for some small enough $\alpha < \epsilon\delta/4$. If $M(\mathcal{P})$ is ϵ-indistinguishable from X then we are done. Else, there must be a distinguisher f such that

$$\begin{aligned} & \mathbb{E} f(D) - \frac{1}{1-\mathbb{P}[D\in\Omega]} \cdot \left(\sum_i \mathbb{P}[D \in P_i] \cdot \mathbb{E} f(U_X|P_i) \right) > \epsilon \\ \implies & \sum_i \mathbb{P}[D \in P_i] \left(\mathbb{E} f(D|P_i) - \mathbb{E} f(U_X|P_i) \right) > \epsilon - \mathbb{P}[D \in \Omega] \left(\epsilon + \mathbb{E} f(D|\Omega) - \mathbb{E} f(D) \right) \\ \implies & \sum_i \mathbb{P}[D \in P_i] \left(\mathbb{E} f(D|P_i) - \mathbb{E} f(U_X|P_i) \right) > \frac{\epsilon}{2} \end{aligned} \tag{13}$$

where the last inequality follows from the assumption that $\alpha < \delta\epsilon/4$ which implies $\mathbb{P}[D \in \Omega] < \alpha/\delta < \frac{\epsilon}{4}$.

A.1 Defining the energy

We now define the *energy* of the partition as

$$\mathcal{E} = \sum_i \mathbb{P}[U_X \in P_i] \left(\frac{\mathbb{P}[D \in P_i]}{\mathbb{P}[U_X \in P_i]} \right)^2$$

Consider a random variable Z defined on the partition as $Z(P_i) = \frac{\mathbb{P}[D\in P_i]}{\mathbb{P}[U_X \in P_i]}$ and $Z(\Omega) = 0$. Then the energy can be thought of as $\mathbb{E} Z^2$ where we choose each atom P_i with probability $\mathbb{P}[U_X \in P_i]$. Observe that the energy defined in this manner is monotone with respect to refinements. Let each P_i be split into parts P_i^j for $0 \leq j \leq k_i$ and let the new energy be $\mathcal{E}'$. Then

$$\mathcal{E}' \;=\; \mathop{\mathbb{E}}_{i,j} (Z(P_i^j))^2 \;\geq\; \mathop{\mathbb{E}}_{i} (\mathop{\mathbb{E}}_{j} Z(P_i^j))^2 \;=\; \mathop{\mathbb{E}}_{i} (Z(P_i))^2 \;=\; \mathcal{E}$$

Claim A.3 *If $\mathcal{P}$ is an α-noisy partition, then $\mathcal{E}(\mathcal{P}) \leq (1+\alpha)/\delta$*

Proof:

$$\mathcal{E} = \sum_i \mathbb{P}\left[D \in P_i\right] \frac{\mathbb{P}\left[D \in P_i\right]}{\mathbb{P}\left[U_X \in P_i\right]} \leq \sum_i \mathbb{P}\left[D \in P_i\right] \cdot \frac{1}{\delta} \cdot \frac{\mathbb{P}\left[R \in P_i\right]}{\mathbb{P}\left[U_X \in P_i\right]} \leq \frac{1}{\delta} \cdot (1+\alpha)$$

■

A.2 Refining the partition

If the partition $(P_1, \ldots, P_k, \Omega)$ does not provide a distribution indistinguishable from D, we refine the partition according to the distinguisher f in (13). We split each partition P_i into P_i^0 and P_i^1 consisting of elements in P_i where output of f is 0 and 1 respectively. To ensure low noise, we need to add the "bad" sets to Ω. Define

$$B = \left\{(i,j) \mid \mathbb{P}\left[R \in P_i^j\right] > (1+\sqrt{\gamma})\mathbb{P}\left[U_X \in P_i^j\right]\right\}$$

We need to show that B constitutes only a small fraction of the space.

Claim A.4 $\sum_{(i,j)\in B} \mathbb{P}\left[R \in P_i^j\right] + \sum_{(i,j)\in B} \mathbb{P}\left[U_X \in P_i^j\right] < 3\sqrt{\gamma}$

Proof: From the definition of B

$$\sum_{(i,j)\in B} \mathbb{P}\left[R \in P_i^j\right] - \sum_{(i,j)\in B} \mathbb{P}\left[U_X \in P_i^j\right] > \sqrt{\gamma} \sum_{(i,j)\in B} \mathbb{P}\left[U_X \in P_i^j\right]$$

Hence, we must have $\sum_{(i,j)\in B} \mathbb{P}\left[U_X \in P_i^j\right] \leq \sqrt{\gamma}$, otherwise the recognizer for the set $\bigcup_{(i,j)\in B} P_i^j$ gives a distinguisher h for R and U_X as claimed in the Theorem.

The above can also be rewritten as

$$\sum_{(i,j)\in B} \mathbb{P}\left[R \in P_i^j\right] - \sum_{(i,j)\in B} \mathbb{P}\left[U_n \in P_i^j\right] > \frac{\sqrt{\gamma}}{1+\sqrt{\gamma}} \sum_{(i,j)\in B} \mathbb{P}\left[R \in P_i^j\right]$$

which implies that $\sum_{(i,j)\in B} \mathbb{P}\left[R \in P_i^j\right] \leq \sqrt{\gamma}(1+\sqrt{\gamma}) < 2\sqrt{\gamma}$ else the indicator function for $\bigcup_{(i,j)\in B} P_i^j$ will again be a distinguisher. Combining the two bounds proves the claim. ■

We now take $\Omega' = \Omega \bigcup \left(\cup_{(i,j)\in B} P_i^j\right)$. We renumber the sets in $[k] \times \{0,1\} \setminus B$ as $(Q_1, \ldots, Q_{k'})$. Then, by the above claim, $(Q_1, \ldots, Q_{k'}, \Omega')$ is an $(\alpha + 3\sqrt{\gamma})$-noisy partition.

A.3 Accounting for the energy increment

We now show that the existence of a distinguisher leads to an incresase in energy when the partition is refined according to the output of the distinguisher. Together with the upper bound on energy, this would imply that this iterative process must stop after a finite number of steps, leading to a distribution which is ϵ-indistinguishable from D.

We first calculate the increase in energy from splitting a single partition P_i into P_i^0 and P_i^1. Let $\mathbb{E}\, f(D|P_i) - \mathbb{E}\, f(U_X|P_i) = \delta_i$. Let $\mathbb{P}\left[D \in P_i\right] = a$ and $\mathbb{P}\left[U_X \in P_i\right] = b$. Taking $\mathbb{E}\, f(U_X|P_i) = p$, we get $\mathbb{E}\, f(D|P_i) = p + \delta_i$. Then, the contribution of this partition to the increase in energy is given by

$$\begin{aligned}
\Delta\mathcal{E}_i &= \frac{(\mathbb{P}\left[D \in P_i^0\right])^2}{\mathbb{P}\left[U_X \in P_i^0\right]} + \frac{(\mathbb{P}\left[D \in P_i^1\right])^2}{\mathbb{P}\left[U_X \in P_i^1\right]} - \frac{(\mathbb{P}\left[D \in P_i\right])^2}{\mathbb{P}\left[U_X \in P_i\right]} \\
&= \frac{a^2(p+\delta_i)^2}{bp} + \frac{a^2(1-p-\delta_i)^2}{b(1-p)} - \frac{a^2}{b} \\
&= \frac{a^2}{b} \cdot \delta_i^2 \left(\frac{1}{p} + \frac{1}{1-p}\right) \geq 4\delta_i^2 \cdot \frac{a^2}{b}
\end{aligned}$$

However, note that the above calculation only holds for sets P_i such that neither P_i^0 nor P_i^1 is included in Ω'. Let $G = \{i | P_i^0, P_i^1 \notin B\}$. Also, note that at most one of P_i^0 and P_i^1 can be included in Ω'. We assume (without loss of generality) that for every $i \notin G$, $P_i^0 \in B$. For all i, we use the following notation for the calculations below:

- $x_i = \mathbb{P}\left[D \in P_i\right]$
- $u_i = \mathbb{P}\left[U_X \in P_i\right]$
- $p_i = \mathbb{P}\left[f(U_X|P_i) = 1\right] = \mathbb{P}\left[U_X \in P_i^1\right] / \mathbb{P}\left[U_X \in P_i\right]$
- $\delta_i = \left|\mathbb{E}\, f(D|P_i) - \mathbb{E}\, f(U_X|P_i)\right| = \left|\frac{\mathbb{P}[D \in P_i^1]}{\mathbb{P}[D \in P_i]} - \frac{\mathbb{P}[U_X \in P_i^1]}{\mathbb{P}[U_X \in P_i]}\right| = \left|\frac{\mathbb{P}[D \in P_i^0]}{\mathbb{P}[D \in P_i]} - \frac{\mathbb{P}[U_X \in P_i^0]}{\mathbb{P}[U_X \in P_i]}\right|$

The total change in energy is

$$\begin{aligned}
\Delta\mathcal{E} \geq\ & \sum_{i \in G} \left(\frac{x_i^2}{u_i}\right) \cdot 4\delta_i^2 + \sum_{i \notin G} \frac{x_i^2}{u_i} \left(\left(\frac{(p_i \pm \delta_i)^2}{p_i}\right) - 1\right) \\
=\ & \sum_{i \in G} \left(\frac{x_i^2}{u_i}\right) \cdot 4\delta_i^2 + \sum_{i \notin G} \frac{x_i^2}{u_i} \left(-(1-p_i) \pm 2\delta_i + \frac{\delta_i^2}{p_i}\right) \\
\geq\ & \sum_{i} \left(\frac{x_i^2}{u_i}\right) \cdot \delta_i^2 + \sum_{i \notin G} \frac{x_i^2}{u_i} \left(-(1-p_i) \pm 2\delta_i\right)
\end{aligned}$$

Analyzing the terms separately, we note that

$$\sum_i \left(\frac{x_i^2}{u_i}\right) \cdot \delta_i^2 \geq \left(\sum_i u_i \cdot \frac{x_i}{u_i} \cdot \delta_i\right)^2 \geq \epsilon^2/4 \qquad \text{(using (13))}$$

Also,

$$\sum_{i \notin G} \frac{x_i^2}{u_i}(1-p_i) = \sum_{i \notin G} \frac{x_i^2}{u_i^2} \cdot u_i(1-p_i) \leq (\frac{1}{\delta}(1+\alpha))^2 \sum_{i \notin G} \mathbb{P}\left[U_n \in P_i^0\right] \leq (\frac{1}{\delta}(1+\alpha))^2 \sqrt{\gamma}$$

Finally,

$$\begin{aligned}
\sum_{i\notin G}\frac{x_i^2}{u_i}\cdot\delta_i &= \sum_{i\notin G}\frac{x_i}{u_i}\mathbb{P}\left[X\in P_i\right]\left|\frac{\mathbb{P}\left[X\in P_i^0\right]}{\mathbb{P}\left[X\in P_i\right]}-\frac{\mathbb{P}\left[U_n\in P_i^0\right]}{\mathbb{P}\left[U_n\in P_i\right]}\right| \\
&\leq \frac{1}{\delta}\cdot(1+\alpha)\left[\sum_{i\notin G}\mathbb{P}\left[X\in P_i^0\right]+\sum_{i\notin G}\frac{\mathbb{P}\left[X\in P_i\right]}{\mathbb{P}\left[U_n\in P_i\right]}\cdot\mathbb{P}\left[U_n\in P_i^0\right]\right] \\
&\leq \frac{1}{\delta}\cdot(1+\alpha)\left[\frac{1}{\delta}\sum_{i\notin G}\mathbb{P}\left[R\in P_i^0\right]+\frac{1}{\delta}(1+\alpha)\sum_{i\notin G}\mathbb{P}\left[U_n\in P_i^0\right]\right] \\
&\leq \left[\frac{1}{\delta}\cdot(1+\alpha)\right]^2\cdot 3\sqrt{\gamma} \qquad \text{(by Claim A.4)}
\end{aligned}$$

Thus, for $\epsilon^2 \geq 64[\delta^{-1}(1+\alpha)]^2\sqrt{\gamma}$, we have that $\Delta\mathcal{E} \geq \epsilon^2/8$. By Claim A.3 the process can continue for at most $8(1+\alpha)/\delta\epsilon^2 \leq 16/\delta\epsilon^2$ steps. Finally, we choose $\gamma = \epsilon^6\delta^4/2^{16}$ so that $\alpha \leq 3\sqrt{\gamma}\cdot\frac{16}{\delta\epsilon^2} < \epsilon\delta/4$. ■

ECCC ISSN 1433-8092
http://eccc.hpi-web.de/