2008 49th Annual IEEE Symposium on Foundations of Computer Science

Dense Subsets of Pseudorandom Sets*

[Extended Abstract]

Omer Reingold [†]
Weizmann Institute of Science
omer.reingold@weizmann.ac.il

Luca Trevisan [‡]
U.C. Berkeley
luca@cs.berkeley.edu

Madhur Tulsiani[‡]
U.C. Berkeley
madhurt@cs.berkeley.edu

Salil Vadhan [§]
Harvard University
salil@eecs.harvard.edu

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. A proof similar to ours has also been independently discovered by Gowers [2].

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.

1. Introduction

Green and Tao [3], 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 [11]:** 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

[*] A full version of this paper is available on ECCC [9] and a short exposition of the main theorem is also available on ArXiv [10].

[†] Research supported by US-Israel Binational Science Foundation Grant 2006060.

[‡] Work partly done while visiting Princeton University and the IAS. Research supported by the National Science Foundation under grants CCF-0515231 and CCF-0729137 and by the US-Israel Binational Science Foundation grant 2006060.

[§] 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.

0272-5428/08 $25.00 © 2008 IEEE
DOI 10.1109/FOCS.2008.38

IEEE computer society

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 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 [12]. 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 [6], 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. It was pointed out to us by Russell Impagliazzo [7] that the connection between Dense Model Theorems and Hardcore Theorems goes even further, and one can deduce the Dense Model Theorem directly from a sufficiently strong version of the Hardcore Theorem.

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

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

where U_X is the uniform distribution over X.[1] More generally, we say that two distributions A and B are ϵ-indistinguishable by a family $\mathcal{F}$ of bounded functions $f : X \to [0,1]$, if for every $f \in \mathcal{F}$

$$|\,\mathbb{E}[f(A)] - \mathbb{E}[f(B)]| \le \epsilon$$

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

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

Given these definitions, a 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 $1/\text{poly}(n)$ distinguishing probabilities and functions computed by polynomial-size circuits.

Tao and Ziegler [12] 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

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

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

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

[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 [12].[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 [6], 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.*

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 $k = \text{poly}(1/\epsilon, \log 1/\delta)$, 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 [2] 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. [4], 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

[2] The two statements of the theorem are completely equivalent, with the following translation. Our functions f are called *dual functions* in [12], 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.

[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.

has pseudoentropy at least $n - t$, provided $\delta = 2^{-t}$ is nonnegligible (i.e. $t = O(\log n)$ when considering $1/\text{poly}(n)$ 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 [1] 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 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 [8].[4]

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

The original proofs by Green, Tao and Ziegler [3, 12] 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 Theorem [6] 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 $\delta/2 \le \mathbb{P}[g(U_H) = 1] \le 1 - \delta/2$, 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 a result, which is a common generalization of our Main Theorem 1.3 and of the Tao-Ziegler Theorem 1.1. We state our result for a more general class of distributions than the ones which are dense in a pseudorandom distribution. We call these distributions (as defined below) pseudodense.

Definition 2.1 *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. We say that D has* pseudo-density *δ w.r.t. $\mathcal{F}$, with error ϵ, if for all $f \in \mathcal{F}$*

$$\begin{aligned} \delta \cdot \mathbb{E}[f(D)] &\le \mathbb{E}[f(X)] + \epsilon \\ \textit{and} \qquad \delta \cdot \mathbb{E}[(1-f)(D)] &\le \mathbb{E}[(1-f)(X)] + \epsilon \end{aligned}$$

[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 [8]); also, the connection between Green-Tao-Ziegler style arguments and Regularity Lemmas is well known in the additive combinatorics community.

It is easy to see that if a distribution D is δ-dense in a distribution R which ϵ-pseudorandom with respect $\mathcal{F}$, then D has pseudo-density δ w.r.t. $\mathcal{F}$.

We can now state the general result in terms of pseudo-dense distributions.

Theorem 2.2 *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 D be a distribution over X.*

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

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

Then D is not pseudo-dense. 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

   $$\delta \cdot \mathbb{E}[h(D)] \geq \mathbb{E}[h(U_X)] + \Omega(\epsilon\delta)$$

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

   $$\delta \cdot \mathbb{E}[h(D)] \geq \mathbb{E}[h(U_X)] + \exp(-\text{poly}(1/\epsilon, 1/\delta))$$

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 $\mathbb{E}[f(D)] - \mathbb{E}[f(M)] \geq \epsilon$. We would be done if we could replace $\mathbb{E}[f(M)]$ in this inequality by $\frac{1}{\delta}\mathbb{E}[f(U_X)]$.

But the fact that M is δ-dense in U_X only gives us the inequality in the other direction, i.e. $\mathbb{E}[f(M)] \leq \frac{1}{\delta}\mathbb{E}[f(U_X)]$, which is inadequate. 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 M is large. 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 2.3 *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: 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

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

and

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

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 (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 [5]. 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 $\bar{f}$ to "prove" that D is not pseudodense.

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 follow from the following claim (whose proof is omitted). A similar step, of using a threshold function (with a randomly chosen threshold) also appears in Holenstein's proof of the hardcore lemma [5].

[5] In case $\delta|X|$ is not an integer, we define U_S to be uniform on the $\lfloor \delta|X| \rfloor$ inputs that maximize $f()$, to have probability 0 on the $|X| - \lceil \delta|X| \rceil$ inputs that minimize $f()$, and to have an appropriate probability value on the remaining elements in order to make U_S δ-dense.

Claim 2.4 *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*

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

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

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

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

We have just shown that the indicator for the event that $\bar{f}$ is above the threshold t proves that D is not pseudodense, 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 $\mathcal{F} \cup (1 - \mathcal{F})$; for $\bar{f} = \sum_i c_i f_i + \sum_j d_j(1 - f_j)$, this can be achieved by reducing the threshold t by $\sum_j d_j$.

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. We can get $k = O((1/\epsilon^2) \cdot \log(1/\epsilon\delta))$ 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

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

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 $\delta \cdot \mathbb{E}[h(D)] - \mathbb{E}[h(U_X)] \geq \Omega(\epsilon\delta)$ as required by the theorem.

■

3 Hardcore Theorems via Iterative Partitioning

Impagliazzo's Hardcore Theorem [6] 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,

$$\left| \mathop{\mathbb{E}}_{x \in H \cap g^{-1}(0)}[f(x)] - \mathop{\mathbb{E}}_{x \in H \cap g^{-1}(1)}[f(x)] \right| \leq O(\epsilon)$$

(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 [11] and the Green–Tao–Ziegler of the Dense Model Theorems [3, 12], 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 [6], 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.[6]

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 $h(f_1(x),\dots,f_k(x))$, 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 $f \in F$,

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

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 $\mathcal{P} = (P_1,\dots,P_m)$ 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 $\mathcal{P} = (P_1,\dots,P_m)$ *satisfies the stopping condition* if at least one of the following conditions holds

- $Y(\mathcal{P}) \cap g^{-1}(0)$ 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 $f_1,\dots,f_i$ in $\mathcal{F}$ such that the partition at step i is generated by $f_1,\dots,f_i$ in the following sense: the partition has a set $P_{b_1,\dots,b_i}$ for each bit string $b_1,\dots,b_i$, defined as

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

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 $f_1(x),\dots,f_i(x)$, and then outputs the majority answer in $P_{f_1(x),\dots,f_i(x)}$; 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

[6] 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.

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.

The following claim (proof omitted) shows that if we further divide every block in the partition according to the value of f, then the energy increases.

Claim 3.2 *Refining $\mathcal{P}$ according to f increases the $\mathcal{E}(\mathcal{P})$ by at least $\delta\epsilon^2$.*

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 state 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). The proof of the theorem is deferred to the full version of the paper.

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 3.3 *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 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 $f \in \mathcal{F}$,*

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

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)).

From the above theorem implies a Hardcore Theorem of the original form, giving a 2δ-dense H on which g is hard to predict with probability greater than $1/2 + \epsilon$.

Corollary 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 $k \leq 1/\delta^2\epsilon^2$ 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 $\mathbb{P}_{x \in H}[f(x) = g(x)] < (1+\epsilon)/2$.*

4 Applications

In this section we discuss some additional applications to of the Dense Model Theorems proved earlier.

We first show that if a distribution is pseudo-dense, then it has large "pseudoentropy" in the sense of Håstad, Impagliazzo, Levin and Luby [4]. In the non-uniform 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 [1], 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 [8]).

4.1 Characterizing Pseudoentropy

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

if

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

(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 [4] generalized the concept of pseudorandomness to the following more general notion of pseudoentropy.

Definition 4.1 ([4]) [7] *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 Model Theorems, observe that a distribution M is δ-dense in U_n iff $\mathrm{H}_\infty(M) \geq n - \log(1/\delta)$. Dense Model Theorems say that if a distribution D has pseudodensity δ, then D has pseudoentropy $n - \log(1/\delta)$. Specifically, using Theorem 2.2, we get:

Corollary 4.2 *Let D be a distribution on $\{0,1\}^n$ that has pseudodensity δ with respect to circuits of size s, with error ϵ. 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 4.3 *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 (and hence D has pseudo-density δ).*

Thus, we have an *equivalence* between having pseudodensity δ 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 $n/2$). We defer the analogue the above characterization for uniform distinguishers to the full version of the paper.

4.2 Regularity Lemma for Sparse Graphs

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 4.4 *We say a graph G is η-pseudorandom if for every pair of disjoint 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 now prove a Dense Model Theorem for dense subgraphs of pseudorandom graphs.

Theorem 4.5 (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 $\eta = \exp(-\mathrm{poly}(1/\epsilon, 1/\delta))$.

[7] 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.

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 4.6 *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 4.7 (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 (without proof) our version of the regularity lemma for sparse graphs. For simplicity, we only state the non-bipartite version.

Theorem 4.8 (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.*

Using the Dense Model Theorem, we can also show sparse analogues of the weak regularity lemma. As opposed to Theorem 4.8 which requires most pairs in the partition to be regular, weak regularity only requires that pairs be regular on average.

A more complete discussion including the statements and proofs of different variants the sparse regularity lemma can be found in the full version of the paper.

Acknowledgments

We thank Terence Tao, Avi Wigderson, Noga Alon, Russell Impagliazzo, Yishay Mansour, and Timothy Gowers for comments, suggestions and references. We also thank the anonymous reviewers for their feedback.

References

[1] B. Barak, R. Shaltiel, and A. Wigderson. Computational analogues of entropy. In *Proceedings of RANDOM'03*, pages 200–215, 2003.

[2] T. Gowers. Decompositions, approximate structure, transference, and the Hahn-Banach theorem. Preprint, 2008.

[3] B. Green and T. Tao. The primes contain arbitrarily long arithmetic progressions. To appear in Annals of Mathematics. math.NT/0404188, 2004.

[4] J. Håstad, R. Impagliazzo, L. Levin, and M. Luby. A pseudorandom generator from any one-way function. *SIAM Journal on Computing*, 28(4):1364–1396, 1999.

[5] T. Holenstein. Key agreement from weak bit agreement. In *STOC '05: Proceedings of the thirty-seventh annual ACM symposium on Theory of computing*, pages 664–673, New York, NY, USA, 2005. ACM.

[6] R. Impagliazzo. Hard-core distributions for somewhat hard problems. In *36th Annual Symposium on Foundations of Computer Science*, pages 538–545, Milwaukee, Wisconsin, 23–25 Oct. 1995. IEEE.

[7] R. Impagliazzo. Personal Communication, November 2008.

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

[9] O. Reingold, L. Trevisan, M. Tulsiani, and S. Vadhan. Dense subsets of pseudorandom sets. Technical Report TR08-045, Electronic Colloquium on Computational Complexity, 2008.

[10] O. Reingold, L. Trevisan, M. Tulsiani, and S. Vadhan. New proofs of the Green-Tao-Ziegler dense model theorem: An exposition. arXiv:math/0806081, 2008.

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

[12] T. Tao and T. Ziegler. The primes contain arbitrarily long polynomial progressions. arXiv:math/0610050, 2006.