Regularity, Boosting, and Efficiently Simulating Every High-Entropy Distribution

Luca Trevisan* Madhur Tulsiani† Salil Vadhan‡

November 22, 2008

Abstract

We show that every high-entropy distribution is indistinguishable from an efficiently samplable distribution of the same entropy. Specifically, we prove that if D is a distribution over $\{0,1\}^n$ of min-entropy at least $n-k$, then for every S and ϵ there is a circuit C of size at most $S \cdot \mathrm{poly}(\epsilon^{-1}, 2^k)$ that samples a distribution of entropy at least $n-k$ that is ϵ-indistinguishable from D by circuits of size S.

Stated in a more abstract form (where we refer to indistinguishability by arbitrary families of distinguishers rather than bounded-size circuits), our result implies (a) the Weak Szemerédi Regularity Lemma of Frieze and Kannan (b) a constructive version of the Dense Model Theorem of Green, Tao and Ziegler with better quantitative parameters (polynomial rather than exponential in the distinguishing probability ϵ), and (c) the Impagliazzo Hardcore Set Lemma. It appears to be the general result underlying the known connections between "regularity" results in graph theory, "decomposition" results in additive combinatorics, and the Hardcore Lemma in complexity theory.

We present two proofs of our result, one in the spirit of Nisan's proof of the Hardcore Lemma via duality of linear programming, and one similar to Impagliazzo's "boosting" proof. A third proof by iterative partitioning, which gives the complexity of the sampler to be exponential in $1/\epsilon$ and 2^k, is also implicit in the Green-Tao-Ziegler proofs of the Dense Model Theorem.

*Computer Science Division, U.C. Berkeley. luca@cs.berkeley.edu. This material is based upon work supported by the National Science Foundation under grant CCF-0729137 and by the US-Israel Binational Science Foundation under grant 2006060.

†Computer Science Division, U.C. Berkeley. madhurt@cs.berkeley.edu This material is based upon work supported by the National Science Foundation under grant CCF-0729137 and by the US-Israel Binational Science Foundation under 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, a Guggenheim Fellowship, US-Israel Binational Science Foundation grant 2006060.

1 Introduction

We show that *every set of noticeable density is computationally indistinguishable (by adversaries of fixed polynomial complexity) from an efficiently samplable distribution of the same density.*

Suppose that D is a subset of $\{0,1\}^n$ of *arbitrary complexity* containing at least $\delta 2^n$ elements (or, more generally, suppose D is an *arbitrary* distribution of min-entropy at least $\log \delta 2^n$), and choose an arbitrary size parameter S and an approximation parameter $\epsilon > 0$. Then we show that there is a distribution M over $\{0,1\}^n$ of min-entropy at least $\log \delta 2^n$ (that is, a convex combination of distributions that are uniform over sets of size $\delta 2^n$) such that D and M are ϵ-indistinguishable by circuits of size $\leq S$, and such that M is samplable and computable by circuits of size $S \cdot \mathrm{poly}(\epsilon^{-1}, \delta^{-1}))$.[1]

Such a result is implicit in the "Constructive Dense Model Theorem" of Green, Tao, and Ziegler [GT, TZ] (c.f. Theorem 7.1 in [TZ]), but with the weaker consequence that M is samplable and computable in size $S \cdot 2^{\mathrm{poly}(\epsilon^{-1}, \delta^{-1})}$.

Our main theorem can be stated in a more abstract form (see Theorem 1.1 below), in which we refer to arbitrary families of distinguishers rather than just bounded-size circuits. In such a form, our main theorem gives the following three results as corollaries:

1. **The Weak Szemerédi Regularity Lemma** of Frieze and Kannan [FK], a result in graph theory, establishing that every graph is "approximated" by an object of "complexity" that depends only on the quality of the approximation and not on the size of the original graph.

   This can be proved via a variation of the proof of the original Szemerédi Regularity Lemma (which provided a stronger notion of approximation), which proceeds by iteratively partitioning the set of vertices. The iterative partitioning proof establishes the existence of an approximator of complexity exponential in ϵ^{-1}, where ϵ is the approximation parameter. The Frieze-Kannan proof gives, in addition, an approximating obejct of complexity polynomial in ϵ^{-1}.

   Our main theorem also provides an approximating object of complexity polynomial in ϵ^{-1}.

2. **The Dense Model Theorem** of Green, Tao and Ziegler [GT, TZ], a result in additive combinatorics stating that if R is a (possibly very sparse) pseudorandom subset of a set X, and $D \subseteq R$, $|D| \geq \delta |R|$ is a subset containing a large fraction of the elements of R, then there is a large model set[2] $M \subseteq X$ $|M| \geq \delta |X|$ that contains a large fraction of all the elements of X and that is "indistinguishable" from D.[3]

   The original proof used an iterative partition approach similar to known proofs of "regularity" results in graph theory. The model set M is explicitly defined in the proof, and it has complexity exponential in the approximation parameter $1/\epsilon$. The strength of the pseudorandomness condition required on R is also exponential in $1/\epsilon$.

[1] It would of course be preferable if the complexity of M were smaller than S, e.g. giving a sampler of fixed polynomial complexity that that generates a distribution that is indistinguishable from D by adversaries of arbitrary polynomial size. However, this is impossible to achieve in general; see Remark 1.6 below.

[2] Technically, both D and M are distributions rather than sets, and the statement of the Theorem refers to their min-entropy rather than theirs size. One could recover a statement about sets by "rounding" the distribution M to the uniform distribution over a large set.

[3] In the additive combinatorics literature, this result is referred to as a "transference" result, because it allow to transfer results that are known for dense sets of integers to dense subsets of pseudorandom sets of integers.

Independently, Gowers [Gow] and Reingold et al. [RTTV] provided another proof based on duality of linear programming. The proof is non-constructive in its definition of the model set M, but the strength of the pseudorandomness condition on R (as discussed in [RTTV]) only needs to be polynomial in $1/\epsilon$. Impagliazzo [Imp2] proved that such a non-constructive version of the Dense Model Theorem with polynomial parameters can be derived from a strong version of the Hardcore Lemma (see below), such as the one proved by Holenstein [Hol].

As a corollary of our main theorem, we prove a constructive version in which M is explicitly defined and has complexity polynomial in $1/\epsilon$, and the strength of the pseudorandomness requirement on R is also polynomial in $1/\epsilon$. Such a Constructive Dense Model Theorem with polynomial parameters is new.

3. **The Impagliazzo Hardcore Lemma** [Imp1], a result in complexity theory stating that if a problem is hard-on-average in a weak sense on uniformly distributed inputs, then there is a "hardcore" subset of inputs of noticeable density[4] such that the problem is hard-on-average in a must stronger sense on inputs randomly drawn from such set.

   There are two main approaches to proving the Hardcore Lemma: a proof due to Nisan, which uses duality of linear programming, and it was the inspiration for the [RTTV] proof the Dense Model Thereom, and a proof due to Impagliazzo, which proceeds in stages and uses a smoothed threshold function. Both proofs are reported in [Imp1]. Klivans and Servedio [KS] relate the problem of constructing hardcore sets to the problem of designing boosting algorithms in learning theory; they show that Impagliazzo's proof of the hardcore lemma can be seen as a boosting algorithm, and that any known boosting algorithm can be used to give a proof of the existence of hardcore sets. Holenstein [Hol] shows how both proofs in [Imp1] can be optimized to give tighter guarantees and require less non-uniformity. Reingold et al. [RTTV] give a proof of the hardcore lemma via iterative partitioning (with an exponential loss in the distinguishing probability), inspired by the Green-Tao-Ziegler proof of the Dense Model Theorem.

As the above discussion implies, connections were known between the above results: the iterative partitioning technique could prove all of them, with exponential loss in some parameters, and the linear programming duality and the boosting technique could prove the Hardcore lemma and the non-constructive version of the Dense Model Theorem, with polynomial parameters.

In this paper we show that the linear programming duality and the boosting techniques can be used to prove our main theorem, and hence all three results (including the constructive version of the Dense Model Theorem), all with polynomial parameters.

We thus enrich the set of known connections between the Regularity Lemma, Dense Model Theorem and Hardcore Lemma by showing that a common generalization of the three of them (our main result) is provable via each of the proof techniques known for each of them.

1.1 Our Main Theorem

We now state our main result, in abstract form.

If $\mathcal{F}$ is a family of real-valued functions, we say that a function h has *complexity at most C relative to $\mathcal{F}$* if there are functions $f_1, \ldots, f_k \in \mathcal{F}$, $k \leq C$ such h can be defined by combining them using at

[4] As in the case of the Dense Model Theorem, the result is formally stated and proved in terms of distributions and min-entropy rather than sets and set size, although a "rounding" argument could prove that a statement about sets would also be true.

most C of the following operations: (a) multiplication by a constant, (b) application of a boolean threshold function, (c) sum, (d) product.

Theorem 1.1 (Main) *Let X be a finite set, μ a probability distribution over X, $\mathcal{F}$ be a collection of functions $f : X \to [0,1]$, $\epsilon > 0$ an approximation parameter, and $g : X \to [0,1]$ an arbitrary bounded function.*

Then there is a function $h : X \to [0,1]$ satisfying $\mathbb{E}_\mu[h] = \mathbb{E}_\mu[g]$ that is

1. **Efficient relative to $\mathcal{F}$:** *h has complexity $\epsilon^{-O(1)}$ relative to $\mathcal{F}$;*
2. **Indistinguishable from g:** *for every $f \in \mathcal{F}$, we have*

$$\left| \mathop{\mathbb{E}}_{x\sim\mu}[g(x)f(x)] - \mathop{\mathbb{E}}_{x\sim\mu}[h(x)f(x)] \right| \leq \epsilon$$

Remark 1.2 We stress that the theorem applies to *arbitrary* functions g, including random functions and functions of very high average-case complexity. If $\mathcal{F}$ is defined to be the set of functions computable by circuits of size S, then h is computable in size $S \cdot \epsilon^{-O(1)}$. Thus, the indistinguishability property does not imply that h is a good approximation of g in the sense of the two functions agreeing on many inputs, which would be impossible if g has high average-case complexity. Rather, the indistinguishability means that, roughly speaking, although h may make many mistakes in computing g, inputs on which h is wrong are indistinguishable from inputs on which h is right. (This will become apparent in our proof that the Main Theorem implies the Impagliazzo Hardcore Lemma.)

Remark 1.3 We remark that, in almost all the applications of the above theorem, μ is the uniform measure, and this should be assumed whenever μ is not specified.

Remark 1.4 In additive combinatorics, results like our Main Theorem are stated as *decomposition* results (cf. Theorem 7.1 in [TZ], the "decomposition" statements in [Gow], or the examples given in Tao's FOCS 2007 tutorial [Tao]). In a "decomposition" statement of our main theorem, the conclusion would be that there are two functions $h_1 : X \to [0,1]$, $h_2 : X \to [-1,1]$ such that: (1) we can write $g = h_1 + h_2$, (2) h_1 has low complexity, and (3) h_2 is nearly orthogonal to all the functions in $\mathcal{F}$, that is, $|\langle h_2, f\rangle| \leq \epsilon$ for every $f \in \mathcal{F}$, where the inner product $\langle \cdot, \cdot \rangle$ is defined as $\langle f, g\rangle := \mathbb{E}_{x\sim\mu}[f(x)g(x)]$. The near-orthogonality condition of h_2 can be made cleaner by introducing the norm $||g||_{\mathcal{F}} = \min_{f\in\mathcal{F}} |\mathbb{E}_{x\sim\mu}[f(x)g(x)]|$. Then the condition on h_2 is simply $||h_2||_{\mathcal{F}} \leq \epsilon$. We could state our Main Theorem as a decomposition theorem by defining $h_1 := h$ and $h_2 := g - h$, but the form stated above is easier to use in our applications.

Remark 1.5 The choice of $[0,1]$ as a range for g, for h, and for the functions in $\mathcal{F}$, is not essential, and it would be equivalent to consider functions ranging in $[-1,1]$; the reason is that one can move from one setting to the other and back via the transformations $f \leftarrow \frac{1}{2} + \frac{1}{2}f$ and $f \leftarrow 1 - 2f$ which preserve complexity and indistinguishability. We shall use the $[-1,1]$ setting in our proofs of the main theorem.

Remark 1.6 If we take $\mathcal{F}$ to be the family of functions computable by circuits of size S, then h has circuit complexity $S \cdot \mathrm{poly}(\epsilon^{-1})$, which is higher than the complexity S that we allow for

the distinguishers. It would be great if we could have a function h of fixed polynomial complexity $C(n) = \text{poly}(n)$ that is indistinguishable from g by functions f of any polynomial complexity, or, in the non-asymptotic setting, say by functions of complexity $(C(n))^{\log n}$. This would imply that every high-entropy distribution D can be simulated by a distribution M samplable in fixed polynomial size and such that D and M are indistinguishable against adversaries of any polynomial size. (We get, instead, the weaker implication that D and M are indistinguishable by adversaries of size smaller than the size of the sampler for M.) Unfortunately, this stronger simulation is provably impossible. This can be seen by noting that in [TV] it was proven that it is possible to extract one (in fact, many) nearly unbiased bit from every high min-entropy distribution samplable by a circuit of size at most C, provided $C \leq 2^{o(n)}$; the extractor $\text{Ext} : \{0,1\}^n \to \{0,1\}$ has complexity $\text{poly}(C)$. But there is a high min-entropy distribution D that makes the output of Ext biased (the uniform distribution on either $\text{Ext}^{-1}(0)$ or $\text{Ext}^{-1}(1)$), and so for every distribution M samplable in size C, Ext acts as size-$\text{poly}(C)$ distinguisher between D and M. One can get a counterexample more directly by picking $g : \{0,1\}^n \to \{-1,1\}$ from a family of $O(S \log S)$-wise independent hash functions; with positive probability, such a g is such that $\mathbb{E}_x[g(x)f(x)] \leq .1$ for every function $f : \{0,1\}^n \to \{-1,1\}$ computable by a circuit of size S.[5] Furthermore, using an efficient construction of hash functions we can let g have circuit complexity $\tilde{O}(nS)$. Suppose now that our theorem could be strengthened so that we could find, for every g, a function h of circuit complexity S that is indistinguishable from g by all functions f of circuit complexity $\tilde{O}(nS)$; then by taking $f = g$ we get a contradiction because $\mathbb{E}_x[g(x)g(x)] = 1$ and $\mathbb{E}_x[h(x)g(x)] \leq .1$.

We give two proofs of our main theorem.

One proof (see Section 2) uses duality of linear programming and employs the following argument: either there is a function $\bar{h}$ that is a convex combination of functions of complexity $\tilde{O}(\epsilon^{-2})$ and that is $\epsilon/2$-indistinguishable from g by $\mathcal{F}$, or there is a universal distinguisher $\bar{f}$ that is a convex combination of functions from $\mathcal{F}$ and that $\epsilon/2$-distinguishes g from every function h of complexity $\tilde{O}(\epsilon^{-2})$. The latter case can be shown to be impossible, and so the former must hold; one then shows that $\bar{h}$ can be approximated by a function h of complexity $\tilde{O}(\epsilon^{-4})$ that is ϵ-indistinguishable from g by $\mathcal{F}$.

The second proof (see Section 3) uses a boosting-like argument to directly construct a function h as required of complexity $O(\epsilon^{-2})$.

In Section 4 we show how to abstract the argument of Frieze and Kannan [FK] to prove a weaker version of our Main Theorem, in which the approximating function h is not required to be bounded.

1.2 Efficiently Simulating High-Entropy Distributions

As a first application of our main result, we discuss how to efficiently simulate any high-entropy distribution. This is simply a matter of instantiating $\mathcal{F}$ to be the set of functions computed by small circuits, and of seeing bounded functions as describing probability distributions.

If D is a distribution of min-entropy $n - k$ that we wish to simulate, S is a circuit size, and δ is an indistinguishability parameter, then we define $g(x) := 2^{n-k} \cdot D(x)$ (notice that we have $0 \leq g(x) \leq 1$ because of the assumption on the min-entropy of D), and apply the main theorem with $\mathcal{F}$ being the class of functions computable by circuits of size $\leq S$ and $\epsilon = \delta 2^{-k}$.

[5] One can use the Chernoff bound for random variables with boundend independence to deduce that for a fixed function f there is a probability at most $2^{-O(S \log S)}$ that $\mathbb{E}_x[g(x)f(x)] > .1$, and then one can take a union bound over all $2^{O(S \log S)}$ functions computable by circuits of size S.

If $h()$ is the function that we get from the main theorem, then define $M(x) := 2^{k-n}h(x)$ and notice that M is a probability distribution of min-entropy $\geq n-k$, and that $M(x)$, like $h(x)$, is computable by a circuit of size $O(S \cdot \text{poly}(\epsilon^{-1}2^k))$. Also, M is samplable by a circuit of size $O(S \cdot \text{poly}(\epsilon^{-1}2^k))$ via rejection sampling. To see that M is indistinguishable from D, observe that for every function f computable by a circuit of size $\leq S$ we have

$$\begin{aligned} |\mathbb{P}_{x\sim D}[f(x)=1] - \mathbb{P}_{x\sim M}[f(x)=1]| &= \left|\sum_x D(x)f(x) - \sum_x M(x)f(x)\right| \\ &= \left|\mathop{\mathbb{E}}_x[g(x)2^k f(x)] - \mathop{\mathbb{E}}_x[h(x)2^k f(x)]\right| \\ &\leq 2^k\epsilon \end{aligned}$$

1.3 Deriving the Weak Regularity Lemma, the Impagliazzo Hardcore Lemma, and the Dense Model Theorem

The Weak Regularity Lemma, the Impagliazzo Hardcore Lemma and the Dense Model Theorem all follow relatively easily from our Main Theorem. In each case, one has to find a proper way to define the space X and the function g, to instantiate the family $\mathcal{F}$, and to interpret the efficiency and pseudorandomness properties of h.

1. Proving the **Weak Regularity Lemma** is mostly a matter of translating notation. Given a graph $G = (V, E)$, we define X to be the set of edges in a complete graph over V, so that we may see G as defining a boolean function $g : X \to \{0, 1\}$; we define $\mathcal{F}$ to contain, for every two disjoint sets of vertices S, T, a function $f_{S,T} : X \to \{0, 1\}$, defined to be the characteristic function of the set of edges having one endpoint in S and one in T.

   Applying the main theorem, we find a function $h : X \to [0, 1]$ which we may see as being a weighted graph H of "bounded complexity" that approximates G in the sense required by the Weak Regularity Lemma. The description of h as a function of at most $k = \text{poly}(1/\varepsilon)$ functions f_{S_i,T_i} induces a natural partition of the vertex set into at most 2^{2k} sets by taking all possible intersections of the sets S_i, T_i and their complements, such that h is constant on the edges between each pair of parts.

2. In the **Hardcore Lemma**, we are given a boolean function $g : \{0, 1\}^n \to \{0, 1\}$ that is hard-on-average in a weak sense, meaning that every small circuit errs in computing g on at least a δ fraction of inputs. We apply the main theorem with $\mathcal{F}$ being the set of all functions computed by small circuits, and we obtain h.

   Consider now the distribution in which element x has probability proportional to $|g(x)-h(x)|$. On the one hand, since h is an efficient function, $\sum_x |g(x) - h(x)| \geq \delta 2^n$, and so the above distribution is very dense, having min-entropy at least $\log \delta 2^n$. One the other hand, if f is any efficient function, the indistinguishability condition can easily be used to conclude that f has almost no correlation with g over the above distribution, which is thus a hardcore distribution in Impagliazzo's sense.

3. To prove the **Dense Model Theorem**, let $R \subseteq X$ be a pseudorandom set and $D \subseteq R$

a dense subset of R, $|D| = \delta|R|$.[6] Let g be the characteristic function of D, and let h be the efficient approximation given by the Main Theorem, *using U_R as the measure μ*. Now, $h : X \to [0,1]$ is defined over all of X, and because of the pseudorandomness of R we have

$$\mathop{\mathbb{E}}_{x\sim X}[h(x)] \approx \mathop{\mathbb{E}}_{x\sim R}[h(x)] = \mathop{\mathbb{E}}_{x\sim R}[g(x)] = \delta$$

Suppose now, for simplicity, that h is the characteristic function of a set M: then M has size $\approx \delta|X|$, and the indistinguishability condition between g and h can be used to argue that M is indistinguishable from D. In general, given h we can define a probability distribution M such that $M(x) = h(x)/\sum_x h(x)$ which has min-entropy $\approx \log \delta|X|$ and is indistinguishable from D. Note that M is samplable and computable in low complexity.

We can also use our Main Theorem to give a new proof of the Yao XOR Lemma(see Section 5.3, and, in general, it seems that the existence of our approximating functions is a useful tool to prove results about pseudorandomness and average-case complexity.

2 The Proof via Duality of Linear Programming

Our first proof of Theorem 1.1 uses duality of linear programming (or, equivalently, the finite-dimensional Hahn-Banach Theorem) in the form of the min-max theorem for two-player zero-sum games.

In our proof we shall use twice the following result.

Lemma 2.1 *Let X be a finite domain and let μ be a distribution on X. Let $\mathcal{G}$ be a set of bounded functions $g : X \to [-1,1]$, and let $\bar{g}$ be a convex combination of functions from $\mathcal{G}$. Then there are functions $g_1, \ldots, g_k \in \mathcal{G}$, for $k = O((1/\epsilon^2) \cdot \log(1/\epsilon))$ such that*

$$\mathop{\mathbb{E}}_{x\sim\mu}\left[\left|\bar{g}(x) - \left(\frac{1}{k}\sum_i g_i(x)\right)\right|\right] \leq \epsilon$$

Proof: The convex combination $\bar{g}$ defines a distribution on the functions in $\mathcal{G}$. Pick k random functions picked from $\mathcal{G}$ according to this distribution and let $\tilde{g}$ denote their average. Then for any fixed $x \in X$, $\mathbb{P}_{g_1,\ldots,g_k}\left[|\bar{g}(x) - \tilde{g}(x)| > \epsilon/2\right] \leq \epsilon/2$ for $k = O((1/\epsilon^2)\log(1/\epsilon))$. Thus, by linearity of expectation $\mathbb{E}_{g_1,\ldots,g_k}\left[\mathbb{P}_{x\sim\mu}\left[|\bar{g}(x) - \tilde{g}(x)| > \epsilon/2\right]\right] \leq \epsilon/2$. In particular, there exist some $g_1, \ldots, g_k$ such that $\mathbb{P}_{x\sim\mu}\left[|\bar{g}(x) - \tilde{g}(x)| > \epsilon/2\right] \leq \epsilon/2$ and hence $\mathbb{E}_{x\sim\mu}\left[|\bar{g}(x) - \tilde{g}(x)|\right] \leq \epsilon$. ■

Theorem 2.2 *Let X be a finite domain, μ a probability distribution over X, $g : X \to [-1,1]$ a bounded function, $\mathcal{F}$ a family of bounded functions $f : X \to [-1,1]$, and $\epsilon > 0$. Then there is a bounded function $h : X \to [-1,1]$ such that $\mathbb{E}[h] = \mathbb{E}[g]$ and*

1. *h has complexity at most $(1/\epsilon)^{O(1)}$ with respect to $\mathcal{F}$;*
2. *For all $f \in \mathcal{F}$,*

$$\left|\mathop{\mathbb{E}}_{x\sim\mu}[f(x)g(x)] - \mathop{\mathbb{E}}_{x\sim\mu}[f(x)h(x)]\right| \leq \epsilon$$

[6] This is a simplified setting; in general we shall be interested in the case in which D and R are allowed to be measures rather than sets.

Proof: Let $\mathcal{F}'$ be the closure of $\mathcal{F}$ under "negation," that is, $\mathcal{F}' := \{f, -f : f \in \mathcal{F}\}$.

Let $t = O((1/\epsilon^2) \cdot \log(1/\epsilon))$ be a parameter that we shall fix later, and let $\mathcal{H}$ be the set of all bounded functions $h : X \to [-1, 1]$ that have complexity at most t with respect to $\mathcal{F}'$ and such that $\mathbb{E}[h] = \mathbb{E}[g]$. Also, for a set $\mathcal{S}$ of functions, let $CH(\mathcal{S})$ denote the set of convex combinations of functions in $\mathcal{S}$.

We next use the min-max theorem of two-player zero sum games, which follows from duality of linear programming. Consider a two player zero-sum game, in which one player picks a function h from $\mathcal{H}$, the other picks f from $\mathcal{F}'$, and the payoff is $\mathbb{E}_{x\sim\mu}[f(x)g(x) - f(x)h(x)]$. By the min-max theorem, one of the two cases must hold:

$$\exists \bar{h} \in CH(\mathcal{H}).\ \forall f \in \mathcal{F}'.\quad \mathop{\mathbb{E}}_{x\sim\mu}[f(x)g(x) - f(x)\bar{h}(x)] \le \frac{\epsilon}{2} \tag{1}$$

$$\exists \bar{f} \in CH(\mathcal{F}').\ \forall h \in \mathcal{H}.\quad \mathop{\mathbb{E}}_{x\sim\mu}[\bar{f}(x)g(x) - \bar{f}(x)h(x)] > \frac{\epsilon}{2} \tag{2}$$

We argue that, for a proper choice of t, Case (2) is impossible, and then we use the function $\bar{h}$ from Case (1) to construct the function h as required.

Suppose Case (2) holds. Then, by Lemma 2.1, we know that there are functions $f_1, \ldots, f_k$, $k = O(\epsilon^{-2} \cdot \log \epsilon^{-1})$ such that if we define

$$\tilde{f}(x) := \frac{1}{k}\sum_{i=1}^{k} f_i(x)$$

we have

$$\left|\mathop{\mathbb{E}}_{x\sim\mu}[f(x)h(x)] - \mathop{\mathbb{E}}_{x\sim\mu}[\tilde{f}(x)h(x)]\right| \ \le\ \mathop{\mathbb{E}}_{x\sim\mu}\left[\left|f(x) - \tilde{f}(x)\right|\right] \ \le\ \frac{\epsilon}{10}$$

for every bounded function h. Then, by (2) and the triangle inequality, we have

$$\forall h \in \mathcal{H}.\ \mathop{\mathbb{E}}_{x\sim\mu}\left[\tilde{f}(x)g(x) - \tilde{f}(x)h(x)\right] \ > \frac{\epsilon}{2} - \frac{2\epsilon}{10} = \frac{3\epsilon}{10}$$

Define now $\hat{f}$ to be equal to $\tilde{f}$ rounded down to the next multiple of $\epsilon/10$. Then for every x, $|\tilde{f}(x) - \hat{f}(x)| \le \epsilon/10$ and so

$$\forall h \in \mathcal{H}.\ \mathop{\mathbb{E}}_{x\sim\mu}\left[\hat{f}(x)g(x) - \hat{f}(x)h(x)\right] > \frac{\epsilon}{10} \tag{3}$$

and $\hat{f}$ takes only the values $0, \epsilon/10, 2\epsilon/10, \ldots, 1 - \epsilon/10, 1$. For $0 \le i \le 10/\epsilon$, let $S_i = \{x : \hat{f}(x) = i\epsilon/10\}$ be the ith level set of $\hat{f}$ and let 1_{S_i} be the indicator function for this set. We define

$$h(x) := \sum_{i=0}^{10/\epsilon} c_i \cdot 1_{S_i}(x) \qquad \text{for } c_i = \left(\frac{1}{|S_i|}\sum_{z\in S_i} g(z)\right)$$

Notice that h has complexity at most $\max\{10/\epsilon, k\} = O((1/\epsilon^2)\log(1/\epsilon))$ with respect to $\mathcal{F}'$, and

that $\mathbb{E}[h] = \mathbb{E}[g]$, so that $h \in \mathcal{H}$ for a sufficiently large choice of t. Now we see that

$$\begin{aligned}
\mathop{\mathbb{E}}_{x\sim\mu}\left[\hat{f}(x)h(x)\right] &= \mathop{\mathbb{E}}_{x\sim\mu}\left[\sum_{i=0}^{10/\epsilon}\hat{f}(x)\cdot c_i\cdot 1_{S_i}(x)\right] \\
&= \mathop{\mathbb{E}}_{x\sim\mu}\left[\sum_{i=0}^{10/\epsilon}\hat{f}(x)\cdot 1_{S_i}(x)\cdot\frac{1}{|S_i|}\sum_{z\in S_i}g(z)\right] \\
&= \sum_i \mathbb{P}\left[\hat{f}(x)=i\frac{\epsilon}{10}\right]\cdot\mathop{\mathbb{E}}_{x\sim\mu}\left[\hat{f}(x)g(x)\,\middle|\,f(x)=i\frac{\epsilon}{10}\right] \\
&= \mathop{\mathbb{E}}_{x\sim\mu}\left[\hat{f}(x)g(x)\right]
\end{aligned}$$

This is in contradiction to (3), and so Case (2) above is impossible.

Thus we must be in case 1. That is, there must exist a function $\bar{h}$, which is a convex combination of functions of complexity at most $t = O((1/\epsilon^2)\log(1/\epsilon))$ and satisfies

$$\forall f\in\mathcal{F}'.\ \mathop{\mathbb{E}}_{x\sim\mu}\left[f(x)g(x)-f(x)\bar{h}(x)\right]\le\frac{\epsilon}{2}$$

It follows from Lemma 2.1 that there are functions $h_1,\ldots,h_k$, $k = O((1/\epsilon^2)\log(1/\epsilon))$ such that if we define $\tilde{h}(x) := \frac{1}{k}\sum_i h_i(x)$, then for all founded functions f we have $\left|\mathbb{E}_{x\sim\mu}\left[\bar{h}(x)f(x)-\tilde{h}(x)f(x)\right]\right| \le \epsilon/10$. (Note that we also have $\mathbb{E}[\tilde{h}] = \mathbb{E}[g]$ because each h_i has the same expectation as g.) So we get

$$\forall f\in\mathcal{F}'.\ \mathop{\mathbb{E}}_{x\sim\mu}\left[f(x)g(x)-f(x)\tilde{h}(x)\right]\le\frac{\epsilon}{2}+\frac{\epsilon}{10}=\frac{3\epsilon}{5}$$

and the theorem follows by noting that $\tilde{h}$ has complexity at most $O((1/\epsilon^4)\cdot(\log(1/\epsilon))^2)$. ■

3 The Proof via Boosting

In this section we give a proof of Theorem 1.1 simular to the proof via "boosting" of the Impagliazzo Hardcore Lemma. We obtain a complexity bound of $O(\epsilon^{-2})$ for h. As explained in Remark 1.5, considering functions ranging over $[-1,1]$, as we shall do below, is equivalent to considering functions ranging over $[0,1]$.

Theorem 3.1 *Let X be a finite domain, μ a probability distribution over X, $g : X \to [-1,1]$ a bounded function, $\mathcal{F}$ a family of boolean functions $f : X \to \{-1,1\}$, and $\epsilon > 0$. Then there is a bounded function $h : X \to [-1,1]$ such that:*

1. *h has complexity $O(1/\epsilon^2)$ with respect to $\mathcal{F}$.*
2. *For all $f \in \mathcal{F}$,*

$$\left|\mathop{\mathbb{E}}_{x\sim\mu}[f(x)g(x)]-\mathop{\mathbb{E}}_{x\sim\mu}[f(x)h(x)]\right| \le \epsilon$$

Remark 3.2 Note that the above version of the theorem only applies to boolean distinguishers. However, this is not a severe limitation as distinguisher taking values in $[-1,1]$ can be converted

to a distinguisher only taking values $\{-1,1\}$, with equal distinguishing advantage and only a small increase in complexity. Specifically, for a function $f : X \to [-1,1]$ such that $\mathbb{E}_{x\sim\mu}[f(x)(g(x) - h(x))] = \delta$, and some $t \in [-1,1]$ consider the function $f_t(x)$ which is 1 if $f(x) \leq t$ and -1 otherwise. Then

$$\mathop{\mathbb{E}}_{t\in[-1,1]}\left[\mathop{\mathbb{E}}_{x\sim\mu}[f_t(x)(g(x)-h(x))]\right] = \mathop{\mathbb{E}}_{x\sim\mu}[f(x)(g(x)-h(x))] = \delta$$

and hence there is some value of t for which f_t is the required boolean distinguisher. Also, f_t has complexity 1 with respect to the class containing f.

Remark 3.3 Theorem 1.1 requires $\mathbb{E}[h] = \mathbb{E}[g]$, which is not guaranteed by the above statement. By adding to $\mathcal{F}$ the function $\mathbf{1}$ which is identically equal to 1, the indistinguishability condition gives us $|\mathbb{E}[g] - \mathbb{E}[h]| \leq \epsilon$; we can then construct a new function $h' : X \to [-1,1]$ whose complexity is only an additive constant term larges than h and such that $\mathbb{E}[g] = \mathbb{E}[h']$ and that h and h' (and thus g and h') are $O(\epsilon)$-indistinguishable. If $|\mathbb{E}[h]| > |\mathbb{E}[g]|$, then we can simply define $h'(x) = h(x)\cdot(|\mathbb{E}\, g|/|\mathbb{E}\, h|)$. Otherwise, we can define $h'(x) := \gamma + (1-\gamma)h(x)$ or $h'(x) := \gamma\cdot(-1) + (1-\gamma)h(x)$ (depending on the sign of $\mathbb{E}[g]$) for an appropriate constant γ. In each case, we have $\mathbb{E}_x\,|h(x) - h'(x)| = O(\epsilon)$, and so h and h' are indistinguishable by arbitrary bounded functions.

Proof: Let $\mathcal{F}' = \{\mathcal{F}\} \cup \{-\mathcal{F}\}$ as before. We start with $h = h_0 \equiv 0$ as the identically zero function and iteratively modify the function h until it satisfies the second property. We will need to argue that the complexity of the function constructed at the end is $O(1/\epsilon^2)$.

Let h_{t-1} be the function obtained after $t-1$ steps of the iteration. If it fails to satisfy the second property, then there exists $f_t \in \mathcal{F}'$ such that $\mathbb{E}_{x\sim\mu}\,[f_t(x)(g(x) - h_{t-1}(x))] > \epsilon$. We then define the modified function h_t (with some parameter $0 < \gamma < 1$ to be chosen later) as

$$h_t(x) = \begin{cases} \gamma\cdot(\sum_{i=1}^t f_t(x)) & \text{if } \gamma\cdot(\sum_{i=1}^t f_t(x)) \in [-1,1] \\ -1 & \text{if } \gamma\cdot(\sum_{i=1}^t f_t(x)) < -1 \\ 1 & \text{if } \gamma\cdot(\sum_{i=1}^t f_t(x)) > 1 \end{cases}$$

Note that at different time steps, $h_t(x)$ only changes in steps of size γ and is bounded in $[-1,1]$. Hence, $h_t(x)$ can be described as a linear combination of $2/\gamma$ functions, each of which checks if $\sum_t f_t$ is in some interval of size γ. Also, if the process stops at time T, then the function h_T satisfies the second condition by definition. It only remains to prove that $T = O(1/\epsilon^2)$ (and choose γ appropriately). The proof will follow from the following claim

Claim 3.4 *For all $x \in X$ and for all $T \geq 1$*

$$\sum_{t=1}^{T} f_t(x)(g(x) - h_{t-1}(x)) \;\leq\; \frac{4}{\gamma} + \frac{\gamma T}{2}$$

We first show how Claim 3.4 implies the theorem. The condition that at every time step the f_t distinguishes between g and h_{t-1} implies that $\mathbb{E}_{x\sim\mu}\,[f_t(x)(g(x) - h_{t-1}(x))] > \epsilon$. Using this and Claim 3.4, we have

$$\epsilon T \;<\; \sum_{t=1}^{T}\mathop{\mathbb{E}}_{x\sim\mu}[f_t(x)(g(x)-h_{t-1}(x))] \;=\; \mathop{\mathbb{E}}_{x\sim\mu}\left[\sum_{t=1}^{T} f_t(x)(g(x)-h_{t-1}(x))\right] \;\leq\; \left(\frac{4}{\gamma} + \frac{\gamma T}{2}\right)$$

Choosing $\gamma = \epsilon$, this gives $T < 8/\epsilon^2$, which proves of the theorem.

We now prove the claim.

Proof (of Claim 3.4): Let $\Delta_t(x)$ denote $(g(x)-h_t(x))$. We intend to show that $\sum_{t=1}^{T} f_t(x)\Delta_{t-1}(x) \leq 4/\gamma + \gamma T/2$. We first note that $\Delta_t(x)$ and $\Delta_{t+1}(x)$ can differ only by γ, $-\gamma$ or 0 (they differ by 0 in case $h_t(x) = h_{t+1}(x) = \pm 1$). We now break the time steps between 1 and T into "level sets" based on the value of Δ_t. For $-1/\gamma < r \leq 1/\gamma$, we define

$$U_r = \{t \mid \Delta_{t-1} = g(x) + (r-1)\gamma \text{ and } \Delta_t = g(x) + r\gamma\}$$
$$L_r = \{t \mid \Delta_{t-1} = g(x) + r\gamma \text{ and } \Delta_t = g(x) + (r-1)\gamma\}$$

In other words, U_r is the set of times at which the value of $\Delta_t(x)$ *increases to* $g(x) + r\gamma$ and L_r is the of times when it *decreases from* $g(x) + r\gamma$. Thus ranging over all r, the sets U_r and L_r contain all times t except the ones for which $\Delta_t(x) = \Delta_{t+1}(x)$.

Since the values of Δ_t change only in steps of γ, for all $\forall r$. $|U_r - L_r| \leq 1$. Also, if $t_1 \in U_r$ and $t_2 \in L_r$, then the signs of $f_{t_1}(x)$ and $f_{t_2}(x)$ must differ and hence $f_{t_1}\Delta_{t_1-1}(x) + f_{t_2}\Delta_{t_2-1}(x) = \gamma$. Hence, we can pair up times in U_r with those in L_r and note that

$$\forall r \quad \sum_{t \in U_r} f_t(x)\Delta_{t-1}(x) + \sum_{t \in L_r} f_t(x)\Delta_{t-1}(x) \leq \gamma \cdot \frac{(|U_r| + |L_r|)}{2} + 2$$

where the 2 is an upper bound on $f_t(x)\Delta_{t-1}(x)$ for the at most one value of t that is left unpaired. For $L = \cup_r L$ and $U = \cup_r U$, we have

$$\sum_{t \in L \cup U} f_t(x)\Delta_{t-1}(x) \;=\; \sum_{r} \sum_{t \in L_r \cup U_r} f_t(x)\Delta_{t-1}(x) \;\leq\; \frac{4}{\gamma} + \gamma \cdot \frac{|L \cup U|}{2}$$

Finally, it remains to bound the contribution for time $t \notin L \cup U$. For such a t, $h_{t-1}(x) = h_t(x) = \pm 1$. We give the argument for the case when the value is fixed at 1. The argument for the -1 case is analogous.

Let $h_t(x)$ be fixed at 1 for $t_1 \leq t \leq t_2$. Then all times from $t_1 + 1$ to t_2 are not in $L \cup U$. The contribution due to this interval is

$$\sum_{t_1 < t \leq t_2} f_t(x)\Delta_{t-1}(x) = \sum_{t_1 < t \leq t_2} f_t(x)(g(x) - 1)$$

However, $g(x) - 1 \leq 0$ and $f_t(x) = 1$ for majority of the times $t \in \{t_1 + 1, \ldots, t_2\}$, (since h_t stays at its upper limit in this interval). Hence,

$$\sum_{t_1 < t \leq t_2} f_t(x)(g(x) - 1) \leq 0$$

Similarly, the contribution of the intervals in which $h_t(x)$ is identically -1 is also non-positive. This proves the claim since

$$\sum_{t=1}^{T} f_t(x)\Delta_{t-1}(x) \;\leq\; \frac{4}{\gamma} + \gamma \cdot \frac{|L \cup U|}{2} \;\leq\; \frac{4}{\gamma} + \gamma \cdot \frac{T}{2}$$

■

■

4 A Generalization of the Argument of Frieze and Kannan

In this section we show how to prove a weaker version of our result by adapting the argument of Frieze and Kannan [FK]. The statement is weaker in that h is not guaranteed to be a bounded function. The proof, however, is much simpler.

Theorem 4.1 *Let X be a finite domain, μ a probability distribution over X, $g : X \to [-1, 1]$ a bounded function, $\mathcal{F}$ a family of bounded functions $f : X \to [-1, 1]$, and $\epsilon > 0$. Then there is a function $h : X \to \mathbb{R}$ such that*

1. *h has complexity $O(\epsilon^{-2})$ with respect to $\mathcal{F}$; indeed there are functions $f_1, \ldots, f_k \in \mathcal{F}$, $k \leq \epsilon^{-2}$, and coefficients $c_1, \ldots, c_k$, such that $\sum_i c_i^2 \leq 1$ and*

$$h(x) := \sum_{i=1}^{k} c_i f_i(x)$$

2. *For all $f \in \mathcal{F}$,*

$$\left| \mathop{\mathbb{E}}_{x \sim \mu}[f(x)g(x)] - \mathop{\mathbb{E}}_{x \sim \mu}[f(x)h(x)] \right| \leq \epsilon$$

Proof: Let $\mathcal{F}'$ be the "closure of $\mathcal{F}$ under negation," that is, $\mathcal{F}' := \mathcal{F} \cup \{-f : f \in \mathcal{F}\}$.
The approximating function h is the output of the following algorithm.[7]

1. $h_0 := \mathbf{0}$; $t := 0$
2. while $\exists f_{t+1} \in \mathcal{F}'. \mathbb{E}_{x \sim \mu}[f_{t+1}(x)g(x)] - \mathbb{E}_{x \sim \mu}[f_{t+1}(x)h_t(x)] > \epsilon$
   (a) $h_{t+1} := h_t + \epsilon f_{t+1}$
   (b) $t := t + 1$

If the algorithm terminates after k steps, then the output function h satisfies the required indistinguishability probability, and it can be written as $h = \epsilon f_1 + \cdots + \epsilon f_k$, where $f_i \in \mathcal{F}'$. Thus, it can be written as $h = \sum_i c_i f_i$ with $f_i \in \mathcal{F}$, and $\sum_i c_i^2 = k\epsilon^2$.

It remains to prove that the algorithm must terminate within $k \leq \epsilon^{-2}$ steps. We do so by a "energy decrease" argument, by defining a non-negative energy function whose value is at most 1 at the beginning, and which decreases by at least ϵ^2 at each step.

For every time step, define the error function $\Delta_t := g - h_t$, and consider the "energy" $E_t := \mathbb{E}_{x \sim \mu}[\Delta_t^2(x)]$.

At time 0, $h = \mathbf{0}$, and so $d_t = g$, and the energy is $E_0 = \mathbb{E}[g^2] \leq 1$.

Going from step t to step $t + 1$, and recalling that $h_{t+1} := h_t + \epsilon f_{t+1}$, we have

$$\begin{aligned} E_t - E_{t+1} &= \mathbb{E}\left[(g - h_t)^2 - (g - h_t - \epsilon f_{t+1})^2\right] \\ &= \mathbb{E}\left[2 \cdot (g - h_t)\epsilon f_{t+1}\right] - \mathbb{E}\left[\epsilon^2 f_{t+1}^2\right] \\ &\geq 2\epsilon^2 - \epsilon^2 = \epsilon^2. \end{aligned}$$

[7] Notice the similarity with the algorithm in our boosting proof; the main difference is that in our boosting proof we bound h at every step so that it is constrained to be between -1 and 1, a condition that breaks the Frieze–Kannan-style analysis presented below.

■

5 Applications

In this section, we shall use the following definition: a distribution A has *density* δ in a distribution B (or A is δ-*dense* in B), if $\forall x.\ \mathbb{P}_A[x] \leq (1/\delta) \cdot \mathbb{P}_B[x]$.

5.1 Deriving the Dense Model Theorem

We prove the Dense Model Theorem in the following formulation:

Theorem 5.1 *Let X be a finite universe, $\mathcal{F}$ a collection of bounded functions $f : X \to [-1, 1]$, $\epsilon > 0$ an accuracy parameter and $\delta > 0$ a density parameter. Let R, D be distributions over X such that D is δ-dense in R. Then there exists $C = 1/\epsilon^{O(1)}$ such that, if, for every function f' of complexity at most C with respect to $\mathcal{F}$, we have*

$$\left| \mathop{\mathbb{E}}_{x \sim R} \left[f'(x) \right] - \mathop{\mathbb{E}}_{x \sim X} \left[f'(x) \right] \right| \leq \epsilon \ ,$$

then D has a dense model in X. That is, there exists a distribution M, which has density at least $(\delta - \epsilon)$ in X such that for all $f \in \mathcal{F}$,

$$\left| \mathop{\mathbb{E}}_{x \sim D} [f(x)] - \mathop{\mathbb{E}}_{x \sim M} [f(x)] \right| \leq O(\epsilon/\delta)$$

Proof: We start by defining the function g which we shall try to approximate.

$$g(x) = \begin{cases} 1 - 2\frac{\delta \cdot \mathbb{P}_D[x]}{\mathbb{P}_R[x]} & \mathbb{P}_R[x] > 0 \\ 1 & \text{otherwise} \end{cases}$$

Note that if we had uniform distributions over some sets R and D, with $|D| = \delta |R|$ then g would be -1 inside the set D and 1 outside. The requirement that $\mathbb{P}_D(x) \leq \frac{1}{\delta} \cdot \mathbb{P}_R(x)$ ensures that g is bounded between -1 and 1. We now apply theorem 1.1 to approximate the function g according to the distribution R. This gives a function h such that $\forall f \in \mathcal{F}.\ |\mathbb{E}_{x \in R}[(g(x) - h(x))f(x)]| \leq \epsilon$. Also h has complexity at most $1/\epsilon^{O(1)}$ with respect to $\mathcal{F}$.

It shall be more convenient to define the distribution M by defining a measure $\rho_M(x) = (1 - h(x))/2$. We will then take $\mathbb{P}_{x \sim M}[x] = \rho_M(x)/(\sum_z \rho_M(z))$. Note that $\rho_M(x) \in [0, 1]$ for every x since h is bounded between -1 and 1. To show that the distribution M is dense in X, we will need to show that $\sum_x \rho_M(x) \geq (\delta - \epsilon)|X|$. This will follow from the facts that the expectation of h over X is close to its expectation over R, which is in turn close to the expectation of g over R. We first note that

$$\mathop{\mathbb{E}}_{x \sim R} \left[\frac{1 - g(x)}{2} \right] = \mathop{\mathbb{E}}_{x \sim R} \left[\frac{\delta \cdot \mathbb{P}_D[x]}{\mathbb{P}_R[x]} \right] = \delta$$

where in the last equality, we used the fact that the support of D is contained in the support of R.

This gives

$$\begin{aligned}\left|\mathop{\mathbb{E}}_{x\sim X}[\rho_M(x)] - \delta\right| &= \left|\mathop{\mathbb{E}}_{x\sim X}\left[\frac{1-h(x)}{2}\right] - \mathop{\mathbb{E}}_{x\sim R}\left[\frac{1-g(x)}{2}\right]\right| \\ &\leq \left|\mathop{\mathbb{E}}_{x\sim X}\left[\frac{1-h(x)}{2}\right] - \mathop{\mathbb{E}}_{x\sim R}\left[\frac{1-h(x)}{2}\right]\right| + \left|\mathop{\mathbb{E}}_{x\sim R}\left[\frac{1-h(x)}{2}\right] - \mathop{\mathbb{E}}_{x\sim R}\left[\frac{1-g(x)}{2}\right]\right| \\ &\leq \frac{\epsilon}{2} + \frac{\epsilon}{2} \;=\; \epsilon\end{aligned}$$

Hence, we get that $\sum_x \rho_M(x) \geq (\delta-\epsilon)|X|$. We also get that $\sum_x \rho_M(x) \leq (\delta+\epsilon)|X|$, which we shall need below. We next need to show that M and D are indistinguishable by any any $f \in \mathcal{F}$. The indistinguishability of the functions g and h by f gives

$$\begin{aligned}\left|\mathop{\mathbb{E}}_{x\sim R}[(g(x)-h(x)f(x)]\right| \leq \epsilon \implies & \left|\mathop{\mathbb{E}}_{x\sim R}\left[\frac{(1-h(x))f(x)}{2} - \frac{(1-g(x))f(x)}{2}\right]\right| \leq \epsilon/2 \\ \implies & \left|\mathop{\mathbb{E}}_{x\sim R}\left[\frac{(1-h(x))f(x)}{2}\right] - \mathop{\mathbb{E}}_{x\sim R}\left[\left(\frac{\delta\mathbb{P}_D[x]}{\mathbb{P}_R[x]}\right)f(x)\right]\right| \leq \epsilon/2 \\ \implies & \left|\mathop{\mathbb{E}}_{x\sim R}\left[\frac{(1-h(x))f(x)}{2}\right] - \delta\cdot\mathop{\mathbb{E}}_{x\sim D}[f(x)]\right| \leq \epsilon/2 \\ \implies & \left|\mathop{\mathbb{E}}_{x\sim X}\left[\frac{(1-h(x))f(x)}{2}\right] - \delta\cdot\mathop{\mathbb{E}}_{x\sim D}[f(x)]\right| \leq 3\epsilon/2\end{aligned}$$

where the last implication used the fact that $h(x)$ has low complexity and hence so does $f(x)(1-h(x))/2$. Consequently, its expectations on the distributions R and X differ by at most ϵ.

Finally, we consider

$$\begin{aligned}\left|\mathop{\mathbb{E}}_{x\sim X}\left[\frac{(1-h(x))f(x)}{2}\right] - \delta\cdot\mathop{\mathbb{E}}_{x\sim M}[f(x)]\right| &= \left|\mathop{\mathbb{E}}_{x\sim X}[\rho_M(x)f(x)] - \delta\cdot\mathop{\mathbb{E}}_{x\sim M}[f(x)]\right| \\ &= \left|\left(\frac{\sum_z \rho_M(z)}{|X|}\right)\mathop{\mathbb{E}}_{x\sim M}[f(x)] - \delta\cdot\mathop{\mathbb{E}}_{x\sim M}[f(x)]\right| \\ &\leq \epsilon\end{aligned}$$

Combining the two bounds and using triangle inequality, we get

$$\left|\delta\cdot\mathop{\mathbb{E}}_{x\sim D}[f(x)] - \delta\cdot\mathop{\mathbb{E}}_{x\sim M}[f(x)]\right| \leq \frac{5\epsilon}{2}$$

which gives $|\mathbb{E}_{x\sim D}[f(x)] - \mathbb{E}_{x\sim M}[f(x)]| \leq \frac{5\epsilon}{2\delta}$ as claimed. ■

5.2 Deriving the Impagliazzo Hard-Core Set Lemma

Theorem 5.2 *Let $\mathcal{F}$ be a family of functions from a finite domain X to $\{0,1\}$ and $\epsilon, \delta > 0$. Then there exists an $s = \text{poly}(1/\epsilon, 1/\delta)$ such that if $g : X \to \{0,1\}$ is a function, which for all functions $f_0 : X \to \{0,1\}$ having complexity at most s w.r.t $\mathcal{F}$, satisfies*

$$\mathbb{P}_{x\sim X}[f(x) = g(x)] \leq 1 - \delta$$

Then there is a distribution μ which is δ-dense in U_X such that

$$\forall f\mathcal{F}.\;\; \mathbb{P}_{x\sim\mu}[f(x) = g(x)] \leq \frac{1}{2} + \epsilon$$

Proof Idea: We apply the Theorem 1.1 to g and obtain an efficiently computable function h that is "indistinguishable" from g. We then define the distribution μ so that $\mu(x)$ is proportional to $|g(x) - h(x)|$. It follows from the weak average-case hardness of g that $|g(x) - h(x)|$ is noticeably large on average, and from this we derive that μ has the required density. The strong average-case hardness of g on the distribution μ follows from the indistinguishability condition, a fact that requires a slightly technical proof based on the following intuition: suppose h were the characteristic function of a set B, and let A be the set $\{x : g(x) = 1\}$. Then A and B have the same size since $\mathbb{E}[g] = \mathbb{E}[h]$, and μ is uniform over the symmetric difference $A \Delta B$. The indistinguishability condition requires every efficient function f to evaluate to 1 on approximately the same number of elements in A and B, and hence approximately the same number of elements in $A - B$ and $B - A$. This means that f correctly computes g in $A - B$ on approximately as many elements as elements of $B - A$ on which f incorrectly computer g, and so f computes g correctly on approximately half the elements of $A \Delta B$.

The formal proof follows.

Proof: We apply Theorem 1.1 to g, with the approximation parameter $\gamma := \epsilon\delta$. Theorem 1.1 gives us a function $h : X \to [0, 1]$ with complexity at most $\text{poly}(1/\gamma)$ with respect to $\mathcal{F}$ such that

$$\forall f \in \mathcal{F}. \quad \mathop{\mathbb{E}}_{x \sim X}[f(x) \cdot (g(x) - h(x))] \leq \gamma$$

Let us consider now the "error function" $|g(x) - h(x)|$. The assumption that g is weakly hard on average, and the fact that h has low complexity, imply that the error must be large on average. In particular, we claim that by choosing $s = \text{poly}(1/\gamma)$ we must have

$$\mathbb{E}\left[|g(x) - h(x)|\right] \geq \delta \tag{4}$$

Indeed, consider the process of picking a random t in $[0, 1]$, and defining the function $h_t(x)$ so that $h_t(x) = 1$ if $h(x) \geq t$ and $h_t(x) = 0$ otherwise. Then, for every choice of t, h_t has complexity $\text{poly}(1/\gamma)$, and recalling that g takes values in $\{0, 1\}$, we have

$$\mathbb{P}_{x \sim X, t \sim [0,1]}[h_t(x) = g(x)] = \mathop{\mathbb{E}}_{x \sim X}\left[|g(x) - h(x)|\right]$$

In particular there is a fixed t such that

$$\mathbb{P}_{x \sim X}[h_t(x) = g(x)] \geq \mathop{\mathbb{E}}_{x \sim X}\left[|g(x) - h(x)|\right]$$

and the claim follows. Let us define the distribution μ so that the probability of a point x is proportional to $|g(x) - h(x)|$. That is

$$\mu(x) := \frac{|g(x) - h(x)|}{\sum_y |g(y) - h(y)|}$$

Note that $\mu(x) \leq 1/(\delta|X|)$ and hence μ has density at least δ. We now fix a function $f \in \mathcal{F}$, and it remains to estimate $\mathbb{P}_{x \sim \mu}[f(x) = g(x)]$, which equals, $\sum_x \mu(x) \cdot 1_{[f(x)=g(x)]}$ where $1_{[f(x)=g(x)]}$ is an indicator function. We will bound this using the identity

$$|g(x) - h(x)| \cdot 1_{[f(x)=g(x)]} = \left[\left(f(x) - \frac{1}{2}\right) \cdot (g(x) - h(x)) + \frac{1}{2}|g(x) - h(x)|\right]$$

To match this with the intuition given earlier, consider the special case that $h(x)$ is boolean and let $A = \{x \mid g(x) = 1\}$, $B = \{x \mid h(x) = 1\}$. Then $|g(x) - h(x)|$ is the characteristic function for

$A\Delta B$, with $g(x)-h(x)$ being 1 on $A\setminus B$ and -1 on $B\setminus A$. So, the above equation (summed over x) says that the number on points in $A\Delta B$ on which $f(x)=g(x)$ (counted twice) equals the number of points in $A\setminus B$ where $f(x)=1$ minus the number of points in $B\setminus A$ where $f(x)=0$, plus the number of points in $A\Delta B$. (the general case can be verified by case analysis on $f(x),g(x)\in\{0,1\}$). This gives

$$\mathop{\mathbb{E}}_{x}\left[|g(x)-h(x)|\cdot 1_{[f(x)=g(x)]}\right] \leq \gamma + \frac{1}{2}\mathop{\mathbb{E}}_{x}\left[|g(x)-h(x)|\right]$$

So, finally, recalling that $\sum_x |g(x)-h(x)| \geq \delta|X|$,

$$\mathbb{P}_{x\sim\mu}[f(x)=g(x)] \;=\; \frac{\mathbb{E}_x\left[|g(x)-h(x)|\cdot 1_{[f(x)=g(x)]}\right]}{\mathbb{E}_x\left[|g(x)-h(x)|\right]} \;\leq\; \frac{1}{2}+\frac{\gamma}{\delta} \;\leq\; \frac{1}{2}+\epsilon$$

■

5.3 Deriving the Yao XOR Lemma

It is also possible to derive the Yao XOR Lemma from our main result.

For every family $\mathcal{F}$ of functions $f_0 : X \to [-1,1]$, we define the family $\mathcal{F}^k$ of functions $f : X^k \to [-1,1]$ such that if we fix all but one input of f, then the resulting function defined on x is in $\mathcal{F}$.

Theorem 5.3 *Let $\mathcal{F}$ be a family of functions from X to $[-1,1]$, $\epsilon,\delta > 0$. Then there exists $s=\mathrm{poly}(1/\epsilon,1/\delta)$ such that if $g : X \to \{-1,1\}$ satisfies for all f_0 having complexity at most s with respect to $\mathcal{F}$*

$$\left|\mathop{\mathbb{E}}_{x\sim X}[f_0(x)g(x)]\right| \leq 1-2\delta$$

Then for $g^{\otimes k} : X^k \to \{-1,1\}$ defined as $g^{\otimes k}(x_1,\ldots,x_k)=\prod_{i=1}^k g(x_k)$, we have

$$\forall f\in\mathcal{F}^k.\ \left|\mathop{\mathbb{E}}_{z\sim X^k}\left[f(z)g^{\otimes k}(z)\right]\right| \leq (1-\delta)^k+\epsilon$$

Remark 5.4 Note that the above theorem implies the usual statement of the XOR lemma where the domain X is $\{0,1\}^n$, functions have range $\{0,1\}$, and the hypothesis is that for all f_0 with complexity at most s w.r.t $\mathcal{F}$, $\mathbb{P}_{x\in\{0,1\}}[f(x)=g(x)] \leq 1-\delta$. The required conclusion is that for $g^{\oplus k} : \{0,1\}^n \to \{0,1\}$, defined as $g^{\oplus}(x_1,\ldots,x_k)=g(x_1)\oplus\ldots\oplus g(x_k)$ and for all $f\in\mathcal{F}^k$, $\mathbb{P}_{x_1,\ldots,x_k}[f(x_1,\ldots,x_k)=g^{\oplus k}(x_1,\ldots,x_k)] \leq 1/2+(1-\delta)^k+\epsilon$. This follows from the theorem by considering the functions $(-1)^{f_0(x)}, (-1)^{g(x)}, (-1)^{f(x_1,\ldots,x_k)}$ and $(-1)^{g^{\oplus k}(x_1,\ldots,x_k)}$.

Proof: We apply Theorem 1.1 with the family X and $\gamma=\epsilon\delta$ to obtain a function h which is $\epsilon\delta$-indistinguishable from g by functions of $\mathcal{F}$. Then h has complexity $\mathrm{poly}(1/\epsilon,1/\delta)$ with respect to $\mathcal{F}$.

Even though h is indistinguishable from g by functions in $\mathcal{F}$, it is still a function of low-complexity with respect to f. Hence, if s is larger than the complexity of h, it cannot be "too correlated" with g. This intuition was also used in the proof of the hardcore lemma and we formalize it in the following claim.

Claim 5.5

$$\mathop{\mathbb{E}}_{x\sim\mu}\left[\left|\frac{g(x)+h(x)}{2}\right|\right] \leq 1-\delta$$

Proof: Since $g(x) \in \{-1, 1\}$, we have $\forall x.\ |g(x) + h(x)| + |g(x) - h(x)| = 2$. Hence, it suffices to show $\mathbb{E}_{x\sim\mu}[|g(x) - h(x)|] \leq 2\delta$. For $t \in [-1, 1]$, consider $h_t(x)$, which is 1 if $h(x) \geq t$ and -1 otherwise. Then,

$$\left|\mathop{\mathbb{E}}_{t\in[-1,1]}\left[\mathop{\mathbb{E}}_{x\sim\mu}[h_t(x)g(x)]\right]\right| = \left|1 - \mathop{\mathbb{E}}_{x\sim\mu}[|g(x) - h(x)|]\right|$$

In particular, this value is achieved for some t. Also, the complexity of h_t is just 1 more than the complexity of h. Thus, if s is large enough,

$$\left|1 - \mathop{\mathbb{E}}_{x\sim\mu}[|g(x) - h(x)|]\right| \leq 1 - 2\delta \implies \mathop{\mathbb{E}}_{x\sim\mu}[|g(x) - h(x)|] \geq 2\delta$$

■

Note that we can split $g(x)$ as $g(x) = g_1(x)+g_2(x)$ for $g_1(x) = (g-h)(x)/2$ and $g_2(x) = (g+h)(x)/2$. We can then rewrite $\prod_{j=1}^{k} g(x_j)$ as

$$\begin{aligned}
\prod_{j=1}^{k} g(x_j) &= g_1(x_1)\prod_{j=2}^{k} g(x_j) + g_2(x_1)\prod_{j=2}^{k} g(x_j) \\
&= g_1(x_1)\prod_{j=2}^{k} g(x_j) + g_2(x_1)g_1(x_2)\prod_{j=3}^{k} g(x_j) + g_2(x_1)g_2(x_2)\prod_{j=3}^{k} g(x_j) \\
&= \sum_{i=1}^{k}\left(\prod_{j=1}^{i-1} g_2(x_j)\right)\cdot g_1(x_i)\cdot\left(\prod_{j'=i+1}^{k} g(x_{j'})\right) + \prod_{j=1}^{k} g_2(x_j)
\end{aligned}$$

Then, for the second term we can bound the correlation with f as

$$\mathop{\mathbb{E}}_{x_1,\ldots,x_k}\left[f(x_1,\ldots,x_k)\prod_{j=1}^{k} g_2(x_j)\right] \leq \mathop{\mathbb{E}}_{x_1,\ldots,x_k}\left[\left|\prod_{j=1}^{k} g_2(x_j)\right|\right] = \prod_{j=1}^{k}\mathop{\mathbb{E}}_{x_j}\left[\left|\frac{g(x_j)+h(x_j)}{2}\right|\right] \leq (1-\delta)^k$$

where the last inequality used Claim 5.5. The correlation with the ith term in the summation, which has $i-1$ factors of the form $g_2(x)$ and one of the form $g_1(x)$, can be bounded as

$$\begin{aligned}
&\mathop{\mathbb{E}}_{x_1,\ldots,x_k}\left[f(x_1,\ldots,x_k)\left(\prod_{j=1}^{i-1} g_2(x_j)\right)g_1(x_i)\left(\prod_{j'=i+1}^{k} g(x_{j'})\right)\right] \\
&\leq \mathop{\mathbb{E}}_{\substack{x_1,\ldots,x_{i-1}\\ x_{i+1},\ldots,x_k}}\left[\left|\left(\prod_{j=1}^{i-1} g_2(x_j)\right)\prod_{j'=i+1}^{k} g(x_{j'})\right|\cdot\left|\mathop{\mathbb{E}}_{x_i}\left[f(x_1,\ldots,x_k)\left(\frac{g(x_i)-h(x_i)}{2}\right)\right]\right|\right] \\
&\leq \epsilon\delta\cdot\prod_{j=1}^{i-1}\mathop{\mathbb{E}}_{x_j}[|g_2(x_j)|] \leq \epsilon\delta(1-\delta)^{i-1}
\end{aligned}$$

Here the penultimate inequality used the fact that the function obtained by all variables except x_i in f, belongs to the class $\mathcal{F}$ and hence does not correlate with $(g-h)/2$. Collecting terms, we can then bound the correlation of f with $g^{\otimes k}$ as

$$\begin{aligned}
\mathop{\mathbb{E}}_{x_1,\ldots,x_k}\left[f(x_1,\ldots,x_k)\prod_{i=1}^{k} g(x_i)\right] &\leq (1-\delta)^k + \epsilon\delta\cdot[1 + (1-\delta) + \ldots + (1-\delta)^{k-1}] \\
&\leq (1-\delta)^k + \epsilon\delta\cdot\frac{1}{\delta} = (1-\delta)^k + \epsilon
\end{aligned}$$

which proves the theorem. ■

Acknowledgements

We wish to thank Ravi Kannan and Omer Reingold for helpful discussions.

References

[FK] Alan M. Frieze and Ravi Kannan. Quick approximation to matrices and applications. *Combinatorica*, 19(2):175–220, 1999. 1, 4, 11

[Gow] Timothy Gowers. Decompositions, approximate structure, transference, and the Hahn-Banach theorem. arXiv:0811.3103, 2008. 2, 3

[GT] Ben Green and Terence Tao. The primes contain arbitrarily long arithmetic progressions. *Annals of Mathematics*, 167:481–547, 2008. 1

[Hol] Thomas Holenstein. Key agreement from weak bit agreement. In *Proceedings of the 37th ACM Symposium on Theory of Computing*, pages 664–673, 2005. 2

[Imp1] Russell Impagliazzo. Hard-core distributions for somewhat hard problems. In *Proceedings of the 36th IEEE Symposium on Foundations of Computer Science*, pages 538–545, 1995. 2

[Imp2] Russell Impagliazzo. Personal Communication, 2008. 2

[KS] Adam R. Klivans and Rocco A. Servedio. Boosting and hard-core set construction. *Machine Learning*, 51(3):217–238, 2003. 2

[RTTV] Omer Reingold, Luca Trevisan, Madhur Tulsiani, and Salil Vadhan. Dense subsets of pseudorandom sets. In *Proceedings of the 49th IEEE Symposium on Foundations of Computer Science*, 2008. 2

[Tao] Terence Tao. Structure and randomness in combinatorics. In *Proceedings of the 48th IEEE Symposium on Foundations of Computer Science*, pages 3–18, 2007. 3

[TV] Luca Trevisan and Salil P. Vadhan. Extracting randomness from samplable distributions. In *Proceedings of the 41st IEEE Symposium on Foundations of Computer Science*, pages 32–42, 2000. 4

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