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

Luca Trevisan
Computer Science Division
U.C. Berkeley
luca@cs.berkeley.edu

Madhur Tulsiani
Computer Science Division
U.C. Berkeley
madhurt@cs.berkeley.edu

Salil Vadhan
School of Engineering
and Applied Sciences
Harvard University
salil@eecs.harvard.edu

***Abstract*—We show that every bounded function $g : \{0,1\}^n \to [0,1]$ admits an efficiently computable "simulator" function $h : \{0,1\}^n \to [0,1]$ such that every fixed polynomial size circuit has approximately the same correlation with g as with h. If g describes (up to scaling) a high min-entropy distribution D, then h can be used to efficiently sample a distribution D' of the same min-entropy that is indistinguishable from D by circuits of fixed polynomial size.**

We state and prove our result in a more abstract setting, in which we allow arbitrary finite domains instead of $\{0,1\}^n$, and arbitrary families of distinguishers, instead of fixed polynomial 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), 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 simulator to be exponential in the distinguishing probability, is also implicit in the Green-Tao-Ziegler proofs of the Dense Model Theorem.

I. Introduction

In this paper, we provide a new, complexity-theoretic method for approximating arbitrary functions. Specifically, we show that every bounded function $g : \{0,1\}^n \to [0,1]$ can be "simulated" by an efficiently computable function $h : \{0,1\}^n \to [0,1]$ in the sense that no circuit of fixed polynomial size can distinguish between g and h. The indistinguishability property we prove between g and h is that every circuit of fixed polynomial size, say at most s, has approximately the same correlation with h as it does with g; we allow the complexity of the "simulator" h to be polynomially larger than s. When g is a (scaled version of) a high-min-entropy probability distribution, then h can be used to construct a "simulator" in the sense usually adopted in cryptography [GMR], namely an efficient randomized algorithm whose output distribution is computationally indistinguishable from g (by circuits of fixed polynomial size).

The efficient simulation h seems to encode a lot of useful complexity-theoretic features of the function g. For example, h can be shown to be essentially the best efficiently computable approximation of g, and thus the average-case complexity of g is captured by the disagreement between g and h. In addition, we show how our efficient simulation theorem rather directly implies a number of known results in computational complexity, namely Impagliazzo's Hardcore Lemma [Imp1], Yao's XOR Lemma [Yao], [GNW], and the Dense Model Theorem of [GT], [TZ], [RTTV], [Gow]. (All of these are discussed in more detail below, in Section I-B.) In these applications, much of the computational work that would normally be done via a "reduction" is now done by the simulator h, and the proofs are mostly just calculations.

We state and prove our result in a more general form that refers to arbitrary domains X (not just $\{0,1\}^n$) and arbitrary families of distinguishers (not just small circuits). In this general form, it also implies the Weak Szemerédi Regularity Lemma of Frieze and Kannan [FK], which states that every graph can be "(weakly) approximated" by an object of "complexity" that depends only on the quality of the approximation and not on the size of the original graph. In this application, g is taken to be the adjacency matrix of the original graph, and h is its "low-complexity" approximation. Indeed, our result can be viewed as a generalized Weak Regularity Lemma, which includes both the graph-theoretic and complexity-theoretic versions as special cases.

We note that many connections between all of these previous results were already known, and indeed a version of our result is implicit in the "Constructive Dense Model Theorem" of Green, Tao, and Ziegler [GT], [TZ], but with the complexity of h depending exponentially on the distinguishing probability (making it less suitable

for computational complexity purposes). Reingold et al. [RTTV] showed how to remove this exponential dependence in the "Dense Model Theorem" (discussed below), but lost the constructivity which corresponds to the efficient simulator h. Their proof is inspired by Nisan's proof of Impagliazzo's Hardcore Lemma [Imp1], and in particular by its use of the min-max theorem of game theory. (Gowers [Gow] discovered an essentially identical proof, but casted in a purely analytic language.) Impagliazzo [Imp2] shows that a non-constructive Dense Model Theorem can be derived from a strong form of his Hardcore Lemma (such as the version proved by Holenstein [Hol]), so that any proof of the (strong form of) the Hardcore Lemma yields a non-constructive Dense Model Theorem.

Although non-constructivity seemed inherent in the proofs of [RTTV], [Imp2], [Gow], both of the proofs we give for our main theorem are based on proof techniques previously used for Impagliazzo's Hardcore Lemma, namely the min-max theorem of game theory (used in Nisan's proof) and boosting algorithms (implicitly used in Impagliazzo's original proof [Imp1] and explicitly used by Klivans and Servedio [KS]). The boosting proof becomes particularly clean when used to prove our main theorem, reinforcing the sense that we may be approaching the "right" unification of these various important results.

A. Our Main Theorem

We now state our main result precisely, 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 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 by $\mathcal{F}$:** for every $f \in \mathcal{F}$, we have

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

In almost all of our applications of this theorem, μ is the uniform measure, and this should be assumed whenever μ is not specified.

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 x on which $h(x) > g(x)$ are indistinguishable from those on which $h(x) < g(x)$ (as can be seen by noting that the indistinguishability condition can be rewritten as $|\mathbb{E}[f(x) \cdot (g(x) - h(x))]| \leq \epsilon$.

Nevertheless, it can be shown that when g is boolean, h is essentially the best efficiently computable approximation to g. Specifically, for every $f \in \mathcal{F}$, $\mathbb{E}_\mu[|f(x) - g(x)|] \leq \mathbb{E}_\mu[|h(x) - g(x)|]/2 + \epsilon$. Note that there is some slackness in this comparison, because we lose a factor of 2 in the error, and h is of higher complexity the functions in $\mathcal{F}$. (For example if $\mathcal{F}$ is the family of functions computable by circuits of size s, then h will have circuit complexity $s \cdot \mathrm{poly}(1/\epsilon)$.) In Proposition 2.3, we give evidence that this is necessary, showing that h cannot be of noticeably lower complexity than $\mathcal{F}$.

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 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.

Proving the Main Theorem: We give two proofs of our main theorem. One proof 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 $O(\epsilon^{-2})$ 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 $O(\epsilon^{-2})$. 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 uses a boosting-like argument to directly construct a function h as required of complexity $O(\epsilon^{-2})$.

B. Applications

Efficiently Simulating High-Entropy Distributions: As a first application of our main result, we discuss how to efficiently simulate any high-entropy distribution.

Corollary 1.2: Suppose that D is a distribution on $\{0,1\}^n$ of *arbitrary complexity* and min-entropy at least $n-k$ (i.e. for all x, $D(x) \leq 1/2^{n-k}$), and and choose an arbitrary size parameter $s \in \mathbb{N}$ and an approximation parameter $\epsilon > 0$. Then there is a distribution M over $\{0,1\}^n$ of min-entropy at least $n-k$, such that D and M are ϵ-indistinguishable by circuits of size s, and such that M is samplable and computable by circuits of size $s \cdot \mathrm{poly}(\epsilon^{-1}, 2^k))$.

The fact that M has the same min-entropy as D is important: without that constraint, the result can simply be obtained by the probabilistic method, letting M consist of $s \cdot \mathrm{poly}(\epsilon^{-1})$ random samples from D. (This was pointed out to us by Elad Verbin.) As with our main theorem, it would of course be preferable if the complexity of M were smaller than s, e.g. giving a simulator 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 achievable in general (similarly to our main theorem).

Deducing the above from our main theorem 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. Specifically, 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 indistinguishability parameter $\epsilon' = \epsilon \cdot 2^{-k}$.

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 at least $n-k$, and that $M(x)$, like $h(x)$, is computable by a circuit of size $O(s \cdot \mathrm{poly}(\epsilon^{-1}2^k))$. Also, M is samplable by a circuit of size $O(s \cdot \mathrm{poly}(\epsilon^{-1}2^k))$ via rejection sampling (i.e. uniformly select $x \xleftarrow{R} \{0,1\}^n$ and $r \xleftarrow{R} [0,1]$ and output x if $r \leq h(x)$, else repeat). To see that M is indistinguishable from D, observe that for every function f computable by a circuit of size $\leq s$ we have

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

The Weak Szemerédi Regularity Lemma [FK]: This is 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. Formally, the lemma states that given a graph $G = (V, E)$ and an approximation parameter ϵ, there is a partition of V into $t = 2^{\mathrm{poly}(1/\epsilon)}$ disjoint sets $V_1, \ldots, V_t$, such that the partition is "weakly ϵ-regular." To define the notion of weak regularity, define $p_{i,j}$ to be the edge density between the sets V_i and V_j, that is, $p_{i,j} := |\mathrm{edges}(V_i, V_j)|/(|V_i| \cdot |V_j|)$, and construct a complete weighted graph G' over the vertex set V such that, in G', the weight of an edge (u, v), where $u \in V_i$ and $v \in V_j$ is equal to $p_{i,j}$. Then the partition is weakly ϵ-regular if for every two disjoint sets of vertices A, B, the number of edges in G between A and B differs from the total weight of edges in G' between A and B by at most $\epsilon|V|^2$ in absolute value. The graph G' is the "low-complexity approximation" mentioned above, and indeed the adjacency matrix of G' is described by $2^{\mathrm{poly}(1/\epsilon)}$ numbers (the densities $p_{i,j}$) and $2^{\mathrm{poly}(1/\epsilon)}$ subsets of V. Such a result can be proved via a variation of the proof of the original Szemerédi Regularity Lemma (which provides a stronger notion of approximation) [Sze], which proceeds by iteratively partitioning the set of vertices. Frieze and Kannan also show how to construct an approximating weighted graph G' whose "complexity" is polynomial in ϵ^{-1}. (The weight matrix of G' is a linear combination of $\mathrm{poly}(1/\epsilon)$ "cut matrices" – adjacency matrices of bipartite complete graphs between two subsets of V).

We now show how our main theorem also provides an approximating object of complexity polynomial in ϵ^{-1}. This 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, namely the density of edges between any two disjoint sets of vertices in H is approximately the same as it is in G . Moreover, the description of h as a function of at most $k = \mathrm{poly}(1/\varepsilon)$ 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.

The Impagliazzo Hardcore Lemma [Imp1]: This is 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 distribution of high min-entropy such that the problem is hard-on-average in a must stronger sense on inputs randomly drawn from this distribution.

We deduce this lemma from our theorem as follows. 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 (which is computable by a small circuit, albeit larger than those in $\mathcal{F}$).

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 sufficiently dense, having min-entropy at least $\log \delta 2^n$. On the other hand, if f is any efficient function in $\mathcal{F}$, the indistinguishability of g and h can be used to argue that f has almost no correlation with g over the above distribution, which is thus a hardcore distribution in Impagliazzo's sense.

The Dense Model Theorem: This is a result originating in the work of Green, Tao and Ziegler [GT], [TZ] in additive combinatorics stating that if R is a (possibly very sparse) pseudorandom subset of a set X (in the usual sense, that every function of low complexity relative to some family $\mathcal{F}$ has approximately the same average on R as on X), and D is a subset of R containing a noticeable (say δ) fraction of the elements of R, then there is a large model set[1] $M \subseteq X$ that contains a noticeable fraction of all the elements of X and that is indistinguishable from D.[2]

[1]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.

[2]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.

The original proof used an iterative partition approach similar to the proof the Szemerédi Regularity Lemma. The model set M is explicitly defined in the proof, and it has complexity exponential in the approximation parameter $1/\epsilon$. (See Theorem 7.1 in [TZ].) The strength of the pseudorandomness condition required on R is also exponential in $1/\epsilon$. 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].

From our main theorem, we are able to 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. The proof proceeds roughly as follows. We take g be the characteristic function of D, and h to be the efficient approximation given by the Main Theorem, *using the uniform distribution on 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.

II. Basic Observations

In this section, we make a few basic observations regarding Theorem 1.1.

First, 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 (due to the requirement that $\mathbb{E}[h] = \mathbb{E}[g]$). We shall use the $[-1,1]$ setting in our proofs of the main theorem.

The indistinguishability condition of Theorem 1.1 is stated in terms of the correlations between f and g; the following lemma shows that, when g is boolean, it is equivalent to work with ℓ_1 distance:

Lemma 2.1: Let X be a finite set, μ a probability distribution over X, $g : X \to \{-1, 1\}$ a boolean function, and $f, h : X \to [-1, 1]$ bounded functions. Then

1) $\mathbb{E}_{x\sim\mu}[f(x)g(x)] = 1 - \mathbb{E}_{x\sim\mu}[|f(x) - g(x)|]$.
2) $|\mathbb{E}_{x\sim\mu}[f(x)g(x)] - \mathbb{E}_{x\sim\mu}[f(x)h(x)]| = |\mathbb{E}_{x\sim\mu}[|f(x) - g(x)|] - \mathbb{E}_{x\sim\mu}[|f(x) - h(x)|]|$.

Proof: Item 1 follows by case analysis on $g(x) = -1$ vs. $g(x) = 1$ for each x separately. Item 2 follows from Item 1. ∎

Next we show that the simulator h of Theorem 1.1 is the essentially "best" efficiently computable approximation to g (up to a factor of 2 in the ℓ_1 error).

Lemma 2.2: Let X be a finite domain, μ a probability distribution over X, $g : X \to \{-1, +1\}$ a boolean function, $\mathcal{F}$ a family of bounded functions $f : X \to [-1, 1]$, and $\epsilon > 0$. Suppose that $h : X \to [-1, 1]$ is such that for all $f \in \mathcal{F}$, we have

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

Then, for all $f \in \mathcal{F}$, we have

$$\mathop{\mathbb{E}}_{x\sim\mu}[|f(x) - g(x)|] \geq \mathop{\mathbb{E}}_{x\sim\mu}[|h(x) - g(x)|]/2 - \epsilon/2.$$

Proof: We omit the input x and the distribution μ from the notation for readability.

$$\begin{aligned}
&\mathbb{E}[|f - g|] \\
&= 1 - \mathbb{E}[fg] \qquad \text{(by Lemma 2.1)} \\
&= 1 - \mathbb{E}[f \cdot (g - h)/2] - \mathbb{E}[f \cdot (g + h)/2] \\
&\geq 1 - \epsilon/2 - \mathbb{E}[|g + h|/2] \\
&= 1 - \epsilon/2 - \mathbb{E}[1 - |g - h|/2] \qquad (g \in \{\pm 1\}) \\
&= 1 - \mathbb{E}[|g - h|]/2 - \epsilon/2.
\end{aligned}$$

∎

Finally, we show that, in the complexity-theoretic setting, where $\mathcal{F}$ consists of all circuits of bounded size, it is impossible for h to have lower complexity than the functions in $\mathcal{F}$. This is along the same lines (and in fact follows from) a similar argument in [TV].

Proposition 2.3: For every $n, s \in \mathbb{N}$ such that $s \geq n$, there is a function $g : \{0, 1\}^n \to \{-1, 1\}$ such that for every function $h : \{0, 1\}^n \to [-1, 1]$ computable by a circuit of size s, there is a function $f : \{0, 1\}^n \to \{-1, 1\}$ computable by a circuit of size $s \cdot \mathrm{poly}(n)$ for which $\mathbb{E}[fg] - \mathbb{E}[fh] \geq .9$.

Proof: Choose g uniformly at random from a family of t-wise independent hash functions for $t = \tilde{O}(s)$. Then with high probability over the choice of g, $\mathbb{E}[gh] \leq .1$ for every function h computable by a circuit of size s. (This follows from a tail inequality for t-wise independent hash functions and a union bound over all circuits of size s.) On the other hand, using standard hash families (e.g. polynomials of degree at most $t - 1$ over $\mathrm{GF}(2^n)$), g is computable by a circuit of size $s \cdot \mathrm{poly}(n)$ and hence we can take $f = g$ and have $\mathbb{E}[gf] = 1$. ∎

III. 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. We obtain a $O(\epsilon^{-4} \cdot \log^2 \epsilon^{-1})$ bound on the complexity of h.

In our proof we shall use twice the following result.

Lemma 3.1: 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

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

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 $x \in X$, $\mathbb{P}_{g_1,\ldots,g_k}[|\bar{g}(x) - \tilde{g}(x)| > \epsilon/2] \leq \epsilon/2$ for $k = O((1/\epsilon^2)\log(1/\epsilon))$. Thus, by linearity of expectation $\mathbb{E}_{g_1,\ldots,g_k}[\mathbb{P}_{x\sim\mu}[|\bar{g}(x) - \tilde{g}(x)| > \epsilon/2]] \leq \epsilon/2$. In particular, there exist some $g_1, \ldots, g_k$ such that $\mathbb{P}_{x\sim\mu}[|\bar{g}(x) - \tilde{g}(x)| > \epsilon/2] \leq \epsilon/2$ and hence $\mathbb{E}_{x\sim\mu}[|\bar{g}(x) - \tilde{g}(x)|] \leq \epsilon$. ∎

Theorem 3.2: 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 $f \in \mathcal{F}$,

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

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:

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

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

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 3.1, 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

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

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

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

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 3.1 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 bounded 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

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

. Since $\mathcal{F}'$ is the closure of $\mathcal{F}$ under negation, this is equivalent to:

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

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

IV. The Proof via Boosting

In this section we give a proof of Theorem 1.1 inspired by the boosting proof of the Impagliazzo Hardcore Lemma and by the Frieze-Kannan proof of the Weak Regularity Lemma. We obtain a complexity bound of $O(\epsilon^{-2})$ for h. As explained in Section II, considering functions ranging over $[-1,1]$, as we shall do below, is equivalent to considering functions ranging over $[0,1]$.

To prove our main theorem, we need to provide a bounded approximation.

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 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 $f \in \mathcal{F}$,

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

Remark 4.2: 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 larger 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. Specifically, we let $h'(x) = \gamma \cdot \mathrm{sign}(\mathbb{E}\, g - \mathbb{E}\, h) + (1-\gamma)\cdot h(x)$, where $\gamma = (\mathbb{E}\, g - \mathbb{E}\, h)/(\mathrm{sign}(\mathbb{E}\, g - \mathbb{E}\, h) - \mathbb{E}\, h) \in (0,1)$. It can be verified that $\mathbb{E}\, h' = \mathbb{E}\, g$. For the indistinguishability of h and h', we note that

$$\begin{aligned}
\mathbb{E}[|h - h'|] &= |\gamma| \cdot \mathbb{E}[|\mathrm{sign}(\mathbb{E}\, g - \mathbb{E}\, h) - h|] \\
&= |\gamma| \cdot |\mathrm{sign}(\mathbb{E}\, g - \mathbb{E}\, h) - \mathbb{E}\, h| \\
&= |\,\mathbb{E}\, g - \mathbb{E}\, h| \leq \epsilon,
\end{aligned}$$

where the second equality follows because $\mathrm{sign}(\mathbb{E}\, g - \mathbb{E}\, h) - h$ always has the same sign.

Proof: As a warm-up, we give a simpler proof that does not ensure that the function h is bounded; this is modelled after an argument of Frieze and Kannan [FK]. Let $\mathcal{F}'$ be the "closure of $\mathcal{F}$ under negation," that is, $\mathcal{F}' := \mathcal{F} \cup \{-f : f \in \mathcal{F}\}$. As in the previous proof, working with $\mathcal{F}'$ enables us to remove the absolute value from the definition of indistinguishability.

We construct h to be the function h_t at the end of the following algorithm.

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

If the algorithm terminates after T steps, then the output function h satisfies the required indistinguishability property, and it has complexity at most T relative to $\mathcal{F}$. (Indeed, it is just a weighted sum of at most T functions from $\mathcal{F}$.) We show that the algorithm must terminate within $T \leq \epsilon^{-2}$ steps via an "energy decrease" argument. We define 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 $\Delta_t = g$, 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

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

where the last inequality follows from the hypothesis that $\mathbb{E}[(g-h_t)\cdot f_{t+1}] > \epsilon$ and using $|f_{t+1}| \leq 1$.

The above construction does not necessarily produce a bounded function h; after $T = \epsilon^{-2}$ steps the values of $h = h_T$ can be as large as $T \cdot \epsilon = \epsilon^{-1}$. We now modify it to yield a bounded function h. The construction is exactly the same, except that we treat values larger than 1 (resp., smaller than -1) as 1 (resp., as -1). Specifically, we have the following algorithm:

1) $h_0 := \mathbf{0}$; $s_0 := \mathbf{0}$; $t := 0$
2) while $\exists f_{t+1} \in \mathcal{F}'$ such that $\mathbb{E}_{x\sim\mu}[f_{t+1}(x)g(x)] - \mathbb{E}_{x\sim\mu}[f_{t+1}(x)h_t(x)] > \epsilon$
   a) $s_{t+1} := s_t + \epsilon f_{t+1}$
   b) $h_{t+1} := \begin{cases} 1 & \text{if } s_{t+1} > 1 \\ s_{t+1} & \text{if } s_{t+1} \in (-1,1) \\ -1 & \text{if } s_{t+1} < -1 \end{cases}$
   c) $t := t + 1$

The algorithm is the same as the previous warm-up algorithm, except that previously the function h at every step was just a (weighted) sum of distinguishers, while now it it is a pointwise truncated version of the sum so that it is always between -1 and $+1$.

Note that, if the process stops after T steps, then the function h_{T+1} satisfies the second condition by definition. It only remains to prove that $T = O(1/\epsilon^2)$. For the analysis, we consider the error $\Delta_t(x) := (g(x) - h_t(x))$ of our bounded function at time t, and define the *overflow function*

$$O_t(x) := (h_t(x) - s_t(x)) \cdot \Delta_t(x).$$

The overflow looks at how much, if at all, s_t and h_t differ, scaled by $\Delta_t(x)$. We note that the the overflow is always non-negative, because $s_t(x) - h_t(x) > 0$ implies $h_t(x) = 1 \geq g(x)$, and similarly in the case $s_t(x) - h_t(x) < 0$.

Finally, define the *energy* at time t as

$$E_t := \mathop{\mathbb{E}}_{x\sim\mu}\left[\Delta_t^2(x) + 2O_t(x)\right]$$

Notice that, at all times t, $E_t \geq 0$, and that, at the beginning, $E_0 \leq 1$.

The proof will now follow from the following claim about these functions:

Claim 4.3: For all points $x \in X$ and for every t

$$\begin{aligned} 2\epsilon f_{t+1}(x)\Delta_t(x) \;\leq\; & \Delta_t^2(x) - \Delta_{t+1}^2(x) \\ & +\; 2O_t(x) - 2O_{t+1}(x) + \epsilon^2 \end{aligned}$$

We first show how Claim 4.3 implies the theorem. The condition that at every time step the f_{t+1} distinguishes between g and h_t implies that $\mathbb{E}_{x\sim\mu}\left[f_{t+1}(x)\Delta_t(x)\right] > \epsilon$. Using this and Claim 4.3, we have

$$E_t - E_{t+1} \geq \epsilon^2$$

and so, if the process continues for T steps,

$$1 \geq E_0 - E_T \geq \epsilon^2 T$$

which is a contradiction if $T > 1/\epsilon^2$.

We now prove the claim.

Proof (of Claim 4.3): For any function f at x, we simply write f instead of $f(x)$ in the rest of the proof. Recall that $\epsilon f_{t+1} = s_{t+1} - s_t$, that $\Delta_{t+1} = \Delta_t - (h_{t+1} - h_t)$, and that $O_t = (h_t - s_t)\cdot\Delta_t$. Proving the claim is then simply a matter of rearragnging terms appropriately:

$$\begin{aligned} & 2\epsilon f_{t+1}\Delta_t \\ &= 2(s_{t+1} - s_t)\Delta_t \\ &= 2(h_{t+1} - h_t)\Delta_t + 2(s_{t+1} - h_{t+1})\Delta_t \\ &\quad - 2(s_t - h_t)\Delta_t \\ &= 2(h_{t+1} - h_t)\Delta_t + 2(s_{t+1} - h_{t+1})\Delta_{t+1} \\ &\quad - 2(s_t - h_t)\Delta_t + 2(s_{t+1} - h_{t+1})(\Delta_t - \Delta_{t+1}) \\ &= 2(h_{t+1} - h_t)\Delta_t - 2O_{t+1} + 2O_t \\ &\quad + 2(s_{t+1} - h_{t+1})(h_{t+1} - h_t) \end{aligned}$$

We note that

$$\begin{aligned} & \Delta_t^2 - \Delta_{t+1}^2 + (h_t - h_{t+1})^2 \\ &= \Delta_t^2 - \Delta_{t+1}^2 + (\Delta_t - \Delta_{t+1})^2 \\ &= 2(h_{t+1} - h_t)\Delta_t \end{aligned}$$

to obtain

$$\begin{aligned} 2\epsilon f_{t+1}\Delta_t \;=\; & \Delta_t^2 - \Delta_{t+1}^2 + 2O_t - 2O_{t+1} \\ & +\; 2(s_{t+1} - h_{t+1})(h_{t+1} - h_t) \\ & +\; (h_{t+1} - h_t)^2 \end{aligned}$$

It remains to prove that

$$2(s_{t+1} - h_{t+1})(h_{t+1} - h_t) + (h_{t+1} - h_t)^2 \leq \epsilon^2$$

Observe that $(h_{t+1} - h_t)^2 \leq \epsilon^2$, so if the product $(s_{t+1} - h_{t+1})(h_{t+1} - h_t)$ equals zero we are done. Otherwise, the only way for $(s_{t+1} - h_{t+1})$ and $(h_{t+1} - h_t)$ to be both non-zero is to have $|s_{t+1}| > 1$, $|h_{t+1}| = 1$, and $|h_t| = |s_t| < 1$. In that case,

$$\begin{aligned} & 2(s_{t+1} - h_{t+1})(h_{t+1} - h_t) + (h_{t+1} - h_t)^2 \\ &= (s_{t+1} - h_t)^2 - (s_{t+1} - h_{t+1})^2 \\ &\leq (s_{t+1} - h_t)^2 \\ &= (s_{t+1} - s_t)^2 \;\leq\; \epsilon^2 \end{aligned}$$

■

■

V. Applications

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

A. 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

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

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

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

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. ∎

B. 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 $s = \mathrm{poly}(1/\epsilon, 1/\delta)$ such that if $g : X \to \{0,1\}$ is a function, which for all functions $f : X \to \{0,1\}$ 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

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

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 $A\Delta B$.

We now prove the Hardcore Lemma using our Main Theorem.

Proof (Theorem 5.2): 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 $\mathrm{poly}(1/\gamma)$ 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 $s = \mathrm{poly}(1/\gamma)$ 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 $\mathrm{poly}(1/\gamma)$, 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

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

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

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

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

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

■

C. Deriving the Yao XOR Lemma

It is also possible to derive the Yao XOR Lemma from our main result. This is similar to the way the XOR Lemma is derived from the Hardcore Lemma (see [Imp1], [GNW]), but somewhat more direct than combining those earlier arguments with our derivation of the Hardcore Lemma above. We present the proof in the full version of the paper.

ACKNOWLEDGMENTS

We thank Boaz Barak, Oded Goldreich, Omer Reingold, Elad Verbin, and the CCC reviewers for their comments and suggestions.

The research of L.T. and M.T was supported by the National Science Foundation under grant CCF-0729137 and by the US-Israel Binational Science Foundation under grant 2006060. This work was done while S.V. was visiting U.C. Berkeley, supported by the Miller Foundation for Basic Research in Science, a Guggenheim Fellowship, US-Israel Binational Science Foundation grant 2006060, and the Office of Naval Research grant N00014-04-1-0478.

REFERENCES

[FK] A. M. Frieze and R. Kannan. Quick Approximation to Matrices and Applications. *Combinatorica*, 19(2):175–220, 1999.

[GNW] O. Goldreich, N. Nisan, and A. Wigderson. On Yao's XOR Lemma. Technical Report TR95-50, Electronic Colloquium on Computational Complexity, 1995.

[GMR] S. Goldwasser, S. Micali, and C. Rackof. The Knowledge Complexity of Interactive Proof Systems. *SIAM Journal on Computing*, 18(1):186–208, 1989. Preliminary version in *Proc of STOC'85*.

[Gow] T. Gowers. Decompositions, Approximate Structure, Transference, and the Hahn-Banach Theorem. arXiv:0811.3103, 2008.

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

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

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

[Imp2] R. Impagliazzo. Personal Communication, 2008.

[KS] A. R. Klivans and R. A. Servedio. Boosting and Hard-Core Set Construction. *Machine Learning*, 51(3):217–238, 2003.

[RTTV] O. Reingold, L. Trevisan, M. Tulsiani, and S. Vadhan. Dense Subsets of Pseudorandom Sets. In *Proceedings of the 49th IEEE Symposium on Foundations of Computer Science*, pages 76–85, 2008.

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

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

[TZ] T. Tao and T. Ziegler. The Primes Contain Arbitrarily Long Polynomial Progressions. *Acta Mathematica*, 201:213305, 2008.

[TV] L. Trevisan and S. P. Vadhan. Extracting Randomness from Samplable Distributions. In *Proceedings of the 41st IEEE Symposium on Foundations of Computer Science*, pages 32–42, 2000.

[Yao] A. C. Yao. Theory and Applications of Trapdoor Functions. In *Proceedings of the 23th IEEE Symposium on Foundations of Computer Science*, pages 80–91, 1982.