

1 Introduction

In this lecture, we go through Russell's [5] proof of the dense model theorem. The dense model theorem appeared in the work of Green and Tao [2] and was then proved also in [7, 1]. In the very recent work of Reingold, Trevisan, Tulsiani and Vadhan [6], the dense model theorem was proved analogously to Holenstein's [3] proof of the hard-core set theorem. The hard-core set theorem was first obtained by Impagliazzo [4]. Impagliazzo provides two proofs of this theorem. One is non-constructive and uses Nisan's argument. The other is more involved and yields a more constructive argument in the sense that the final hard-core set can be obtained by using functions from the given test family. Holenstein improves both these proofs. Reingold et.al. obtain their result by adapting the non-constructive proof technique of Holenstein. Consequently, they do not obtain an algorithmic version of the proof.

Russell's current argument reduces the proof of the existence of dense model theorems for a set w.r.t. a class of tests to the existence of hardcore sets for a very related function w.r.t the same class of tests. This, on the one hand, unifies all treatment and at the same time makes available all proofs of the existence of hardcore sets to proving dense model theorems. In particular, using Holenstein's argument, we get a constructive version of the dense model theorem.

2 Notation and Main Theorem

Let U be a finite universe. It will be convenient for us to talk about measures over U . A *measure* μ is simply a function of the form $\mu : U \rightarrow [0, 1]$. Given a probability distribution σ over U , define the density of the measure μ w.r.t. σ , denoted by $d(\mu)$, as simply $\sum_{x \in U} \mu(x)\sigma(x)$. Every $S \subseteq U$ has a corresponding measure given by the associated characteristic function. Hence, the density of a subset S is just the probability mass endowed to S by distribution σ .

Assuming $d(\mu)$ is non-zero, the probability distribution induced by a measure μ , denoted by D_μ , is given by $D_\mu(x) = \mu(x)\sigma(x)/d(\mu)$. Hence a set $S \subseteq U$ also induces a distribution, denoted by D_S . Observe that if S has density δ , then for every function $T : U \rightarrow \{0, 1\}$,

$$\Pr_{x \sim \sigma} [T(x) = 1] \geq \Pr_{x \sim \sigma} [x \in S] \Pr [T(x) = 1 | x \in S] = \delta \Pr_{x \sim D_S} [T(x) = 1]$$

One key step in [?] is to generalize the notion of density to pseudo-density. We say that S

has ϵ -pseudo-density at least δ , w.r.t. to boolean function T if

$$\Pr_{x \sim \sigma} [T(x) = 1] \geq \delta \Pr_{x \sim D_S} [T(x) = 1] - \epsilon$$

A test is just a boolean valued function on the domain U . We consider a family or set of tests that is closed under complement and contains the constant boolean functions 1 and 0. Given such a family Γ and any positive integer t , let Γ_t be the family of tests each of whose element is obtained by taking the majority vote of up to t functions from Γ . Formally, given boolean functions T_1, \dots, T_k , let $\text{Maj}_k(T_1(x), \dots, T_k(x))$ output 1 if at least $\lceil k/2 \rceil$ T_i 's output 1. Then,

$$\Gamma_t \equiv \{\text{Maj}_k(g_1, \dots, g_k) \mid g_i \in \Gamma; 1 \leq k \leq t\}$$

Two distributions σ_1 and σ_2 are said to be ϵ -indistinguishable by tests in Γ if the following is satisfied:

$$\left| \Pr_{x \sim \sigma_1} [T(x)] - \Pr_{x \sim \sigma_2} [T(x)] \right| \leq \epsilon; \forall T \in \Gamma$$

A measure μ is an ϵ -model of a subset S if the distributions D_μ and D_S are ϵ -indistinguishable.

Theorem 1 (Impagliazzo [5]) *Given a set of tests Γ , there exists a function $t = \text{poly}(\frac{1}{\epsilon}, \frac{1}{\delta})$ such that the following holds: consider any set S that has ϵ -pseudo-density at least δ w.r.t Γ_t . There exists a $(\delta - O(\epsilon))$ -dense measure μ that is an $O(\epsilon/\delta)$ -model of S and μ is of the form $\mu(x) = \text{pl}(\sum_{i=1}^r (T_i(x)))$, $r < t$ and each $T_i \in \Gamma$ for $i \leq r$.*

This theorem easily implies a version of the dense model theorem of Reingold et.al.[6]. For their setting [6] considers the following scenario: let $S \subseteq R$ be two sets such that S is δ -dense in R w.r.t. the uniform distribution. They show that if R is ϵ -pseudorandom, i.e. D_R and the uniform distribution are ϵ -indistinguishable to tests in Γ_t , then S has an $O(\epsilon)$ -model for tests in Γ that is δ -dense. To see how such a result follows from Theorem 1, all we need to show is that the given set S has large pseudo-density. This is established in the following straight-forward fashion: $\Pr_{x \in_R U} [T(x) = 1] \geq \Pr_{x \sim D_R} [T(x) = 1] - \epsilon$, as R is ϵ -pseudorandom. But as S has density δ in R , $\Pr_{x \sim D_R} [T(x) = 1] \geq \delta \Pr_{x \sim D_S} [T(x) = 1]$ which finishes off the argument.

In the remainder of the lecture, we prove Theorem 1 using the version of the hardcore-set theorem due to Holenstein. We introduce now, the basic notions needed to state the hardcore-set theorem. A function f is called δ -hard for a family of tests Γ , on a distribution σ over the universe if $\Pr_{x \sim \sigma} [f(x) = T(x)] \leq 1 - \delta$, for each test T in Γ . Call f to be ϵ -hardcore on σ for Γ if every test can achieve advantage no better than ϵ over random guessing in predicting f , i.e. $\Pr_{x \sim \sigma} [f(x) = T(x)] \leq 1/2 + \epsilon$. Note that this is the same as saying f is $(1/2 - \epsilon)$ hard for Γ . In order to prove Theorem 1, we make use of the following:

Theorem 2 (Holenstein’s improved Hardcore-Set Theorem) *Let f be a δ -hard function under distribution σ for tests in Γ_t , where t is a function of the form $\text{poly}(\frac{1}{\epsilon}, \frac{1}{\delta})$. Then, there is a measure μ of density at least 2δ such that f is ϵ -hardcore on distribution D_μ for tests in Γ . Further, μ has the following form: $\mu(x) = p_l(\sum_{i=1}^r (T_i(x) \oplus f(x)))$, $r < t$ and each $T_i \in \Gamma$ for $i \leq r$.*

3 Proof of the Dense Model Theorem

Here, we prove Theorem 1 by constructing appropriate hardcore sets using Holenstein’s theorem. Forthwith are the details.

Let $V_S = \{(1, x) \mid x \in S\}$ and $V_U = \{(0, x) \mid x \in U\}$. Construct the larger universe (than U) $V = V_S \cup V_U$. We think of Γ as a set of tests defined on V by making each test in Γ ignore the first co-ordinate of an element in V and just act on the second co-ordinate. The hard function that we will consider is f that distinguishes which part of V does an input come from, the V_S part or the V_U part. Formally, $f(b, x) = b$. The intuitive idea is that since our tests entirely act on the second co-ordinate and since S has large pseudo-density, they fail to distinguish which part the input comes from and so they find f hard. More formally, consider the following distribution σ on V generated by the following process: with probability $\delta' = \delta/(1 + \delta)$, sample uniformly from V_S and with probability $(1 - \delta')$, sample uniformly from V_U

Claim 3 *The function f is $(\delta' - (1 - \delta')\epsilon)$ -hard under the distribution σ for tests in Γ_t .*

Proof: Assume the contrary, i.e. $\Pr_{x \sim \sigma} [T(x) = f(x)] > 1 - \delta' + \epsilon(1 - \delta')$. We show that this violates the pseudo-density of the set S . Our assumption implies,

$$\Pr_{x \sim \sigma} [T(x) = f(x)] = \delta' \Pr_{x \in RS} [T(x) = 1] + (1 - \delta')(1 - \Pr_{x \in RU} [T(x) = 1]) > 1 - \delta' + \epsilon(1 - \delta') \quad (1)$$

Dividing both sides of (1) by $(1 - \delta')$ and plugging the fact that $\delta = \delta'/(1 - \delta')$, we get

$$\delta \Pr_{x \in RS} [T(x) = 1] + 1 - \Pr_{x \in RU} [T(x) = 1] > 1 + \epsilon$$

which after rearranging yields

$$\Pr_{x \in RU} [T(x) = 1] < \delta \Pr_{x \in RS} [T(x) = 1] - \epsilon$$

which contradicts the assumption that ϵ -pseudo density of S is at least δ , under the uniform distribution over U . ■

Applying Theorem 2, we obtain a measure μ over V of density $(2\delta' - 2\epsilon(1 - \delta'))$ w.r.t. distribution σ , such that f is $\epsilon\delta'/4$ -hardcore for tests in Γ . Using the bound on the density of μ , we get

$$d(\mu) = \delta'd(\mu_S) + (1 - \delta')d(\mu_U) \geq 2\delta' - 2(1 - \delta')\epsilon \quad (2)$$

Using the fact that μ makes f hardcore for tests in Γ and that constants 0 and 1 are in Γ , we obtain that each of $|\Pr_{(b,x) \sim D_\mu} [b = f(b, x) = 0] - 1/2|$ and $|\Pr_{(b,x) \sim D_\mu} [b = f(b, x) = 1] - 1/2|$ are bounded from above by $\epsilon\delta'/4$. Consequently,

$$\left| \Pr_{(b,x) \sim D_\mu} [b = f(b, x) = 1] - \Pr_{(b,x) \sim D_\mu} [b = f(b, x) = 0] \right| \leq \epsilon\delta'/2$$

This means,

$$|\delta'd(\mu_S) - (1 - \delta')d(\mu_U)| \leq \epsilon\delta'd(\mu)/2 \leq \epsilon\delta'/2 \quad (3)$$

Solving for (2) and (3), we obtain that $d(\mu_S) \geq 1 - O(\epsilon/\delta)$ and $d(\mu_U) \geq \delta - O(\epsilon)$.

This shows that μ_U is a $(\delta - O(\epsilon))$ -dense measure on U , w.r.t. the uniform distribution. To complete the proof of Theorem 1, we just have to show that it is a good model of S . In order to do so, notice that the density of measure μ_S shows that the distribution D_{μ_S} is statistically $O(\epsilon/\delta)$ -close to the uniform distribution on S . Hence, for each test T ,

$$\left| \Pr_{x \in RS} [T(x) = 1] - \Pr_{x \sim D_{\mu_U}} [T(x) = 1] \right| \leq O\left(\frac{\epsilon}{\delta}\right) + \left| \Pr_{x \sim D_{\mu_S}} [T(x) = 1] - \Pr_{x \sim D_{\mu_U}} [T(x) = 1] \right|$$

All that we do next is show that the RHS above is small, establishing that tests cannot distinguish well S and the measure μ_U . Using the fact that f is $\epsilon\delta'/4$ -hardcore for any $T \in \Gamma$, we observe that,

$$\frac{1}{2} + \epsilon\delta'/4 \geq \Pr_{x \sim D_\mu} [f(b, x) = T(b, x)] = \Pr_{x \sim D_\mu} [x \in V_S] \Pr_{x \sim D_{\mu_S}} [T(x) = 1] + \Pr_{x \sim D_\mu} [x \in V_U] \Pr_{x \sim D_{\mu_U}} [T(x) = 0] \quad (4)$$

Note that again considering the constant tests 0 and 1, (4) implies the following

$$\frac{1}{2} - \frac{\epsilon\delta'}{4} \leq \Pr_{x \sim D_\mu} [x \in V_S], \Pr_{x \sim D_\mu} [x \in V_U] \leq \frac{1}{2} + \frac{\epsilon\delta'}{4} \quad (5)$$

Combining (4) and (5), we get

$$\Pr_{x \sim D_\mu} [x \in V_S] \Pr_{x \sim D_{\mu_S}} [T(x) = 1] - \Pr_{x \sim D_\mu} [x \in V_U] \Pr_{x \sim D_{\mu_U}} [T(x) = 1] \leq \frac{\epsilon \delta'}{2} \quad (6)$$

Now we first consider the case that, $\Pr_{x \sim D_\mu} [x \in V_S] \leq \Pr_{x \sim D_\mu} [x \in V_U]$. Hence from (5), $\Pr_{x \sim D_\mu} [x \in V_U] - \epsilon \delta' / 2 \leq \Pr_{x \sim D_\mu} [x \in V_S]$. Substituting this into (6),

$$\left| \Pr_{x \sim D_{\mu_S}} [T(x) = 1] - \Pr_{x \sim D_{\mu_U}} [T(x) = 1] \right| \leq \frac{\epsilon \delta'}{1/2 - \epsilon \delta' / 4}$$

This finishes off the argument for this case. The other case, when $\Pr_{x \sim D_\mu} [x \in V_S] \geq \Pr_{x \sim D_\mu} [x \in V_U]$, is handled completely analogously with test T replaced by its complement.

The constructive part of Theorem 1 directly follows from the constructive part of Holenstein's hardcore-set theorem.

References

- [1] T. Gowers. Decompositions, approximate structure, transference, and the Hahn-Banach theorem. Preprint, 2008.
- [2] B. Green and T. Tao. The primes contain arbitrarily long arithmetic progressions. *To appear in Annals of Mathematics*, 2004.
- [3] T. Holenstein. Key agreement from weak bit agreement. In *37th Symposium on Theory of Computing*, pages 664–673, 2005.
- [4] R. Impagliazzo. Hard-core distributions for somewhat hard problems. In *36th Symposium on Foundations of Computer Science*, pages 538–545, 1995.
- [5] R. Impagliazzo. Dense models, hardcore distributions, and computational density. Under preparation, 2008.
- [6] O. Reingold, L. Trevisan, M. Tulsiani, and S. Vadhan. Dense subsets of pseudorandom sets. In *49th Symposium on Foundations of Computer Science*, pages 76–85, 2008.
- [7] T. Tao and T. Ziegler. The primes contain arbitrarily long polynomial progressions. arXiv:math/0610050, 2006.