HARDCORE MEASURES, DENSE MODELS AND LOW COMPLEXITY APPROXIMATIONS

by

Sitanshu Gakkhar
B.S., Utah State University, 2006

A Thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science
in the
School of Computing Science
Faculty of Applied Sciences

SIMON FRASER UNIVERSITY
Summer 2012

APPROVAL

Name:	Sitanshu Gakkhar
Degree:	Master of Science
Title of Thesis:	Hardcore Measures, Dense Models and Low Complexity Approximations

Examining Committee: Dr. Ramesh Krishnamurti
Chair

_____

Dr. Valentine Kabanets, Senior Supervisor

_____

Dr. Andrei Bulatov, Supervisor

_____

Dr. David Mitchell, SFU Examiner

Date Approved: 23 July, 2012

Declaration of Partial Copyright Licence

Abstract

Continuing the study of connections amongst Dense Model Theorem, Low Complexity Approximation Theorem and Hardcore Lemma initiated by Trevisan et al. [TTV09], this thesis builds on the work of Barak et al., Impagliazzo, Reingold et al. and Zhang [BHK09, Imp09, RTTV08a, Zha11] to show the essential equivalence of these three results. The first main result obtained here is a reduction from any of the standard black-box Dense Models Theorems to the Low Complexity Approximation Theorem. The next is the extension of Impagliazzo's reduction from Strong Hardcore Lemma to Dense Model Theorem. Then using Zhang's Dense Model Theorem algorithm we reduce Weak Hardcore Lemma to Strong Hardcore Lemma. Last we distill the methods of Barak et al. and Zhang to extract a single algorithm which yields uniform constructions for all three. Putting all this together demonstrates the three results are essentially equivalent.

Acknowledgments

Without Dr. Valentine Kabanets this would neither exist nor be in this form.

Contents

Chapter 1

Introduction

1.1 Introduction

The focus of this thesis lies on the interface of additive combinatorics and computational complexity specifically exploring the connections between a fundamental result in complexity, *The Hardcore Lemma* and a recent technique in additive combinatorics, *The Dense Model Theorem*. *The Hardcore Lemma* addresses hypothesis under which *hardcore measures* exist: a *measure* can be thought of as a distribution over sets and *hardcore measure* characterizes a subset on which a given function can be guessed no better than a random coin toss (i.e. a *hardcore measure* is the "hard-core" of the function). While *The Dense Model Theorem* describes the conditions under which a set of small size is *indistinguishable* from a much larger set - this larger set (or measure) is said to be the *dense model* for the original set.

The starting point for the interplay between *dense models* and *hardcore measures* is the Weak Szemerédi Regularity Lemma of Frieze and Kannan [FK99] which addresses the existence of approximations to graphs with the complexity of the approximating graphs depending on the approximating parameter and not the size. In [TTV09] Trevisan et al. establish that Weak Szemerédi Regularity Lemma follows from their Low Complexity Approximation Theorem (LCAT) since the Low Complexity Approximation Theorem essentially says that every function can be approximated by a function with *complexity* dependent only on the approximation quality. Because the idea of *approximation* lines up with that of Weak Szemerédi Regularity Lemma this means that Low Complexity Approximation Theorem is simply Weak Szemerédi Regularity Lemma in a different language.

Trevisan et al. also show that the existence of such *low complexity approximations* can be harnessed to obtain forms of *The Hardcore Lemma* and *Yao's* XOR *Lemma* (albeit quantitatively weak). In more detail Trevisan et al. prove that if a boolean function is hard to predict for functions of certain *complexity*, then there exists a *hardcore measure* for functions of polynomi-

ally smaller *complexity* and of relatively big size. This is the standard setting for the *Hardcore Lemma*: if a function is hard to guess for a collection of functions then there exists hardcore measure for a sub-collection of those functions. *The Hardcore Lemma* has two standard forms: *The Weak Hardcore Lemma* (HCL_{WEAK}) and *The Strong Hardcore Lemma* ($\text{HCL}_{\text{STRONG}}$), the difference lies in the size of the hardcore measure whose existence is guaranteed. HCL_{WEAK} was used by Russell Impagliazzo [Imp95] to show that if a function is hard to predict then the hardness can be amplified by taking the XOR of multiple copies of the function (this is informally the Yao XOR Lemma [GNW95]). Trevisan et al. further demonstrate the multi-faceted nature of LCAT by using *low complexity approximations* to derive the *The Dense Model Theorem*

The Dense Model Theorem has roots in a technique from additive combinatorics used by Green, Tao and Ziegler [GT08, TZ08] in their papers on existence of polynomial progressions in primes. Trevisan [Tre09] boils down the essence of methods of Tao et al. as follows:

1. Every set D having positive density inside a *pseudorandom* set R of integers must have arbitrarily long arithmetic progressions.
2. The primes have positive density inside the set of almost primes, and the set of almost primes is *pseudorandom*.

The main step of (1) is showing that if D is a dense subset of a *pseudorandom* subset R of the integers, then there is a set M, of positive density in all the integers, which is is *indistinguishable* from D (where the *indistinguishability* involves *Gowers norms*).

The Dense Model Theorem ($\text{DMT}_{\text{PSEUDORANDOM}}$) formalizes this transfer from the *dense* subset of a pseudorandom set to the *dense* model. In complexity theoretic formulation though, the notion of *indistinguishability* is generalized to be with respect to any class of bounded functions now just *Gowers norms*. However, *Gowers norms* do provide an another connection between complexity and additive combinatorics: since the original use of *Gowers norms* by Timothy Gowers in an analytic proof of Szemerédi's theorem on arithmetic progression in primes [TG06], objects similar to *Gower's norms* were noticed in analysis of complexity theoretic protocols for number-on-the-forehead problem model[Tre09]. And more recently Bogdanov, Lovett and Viola [BV07, Lov09, Vio08] use them to analyze pseudorandom generator constructions, while Viola and Wigderson [VW07] apply *Gowers norms* to communication complexity and to obtain XOR Lemma type results for correlation involving low degree polynomials.

It was noted by Impagliazzo [Imp09] that if a set looks bigger than it's actual size then the set must be hard to identify. This yields a natural hard to guess function. It can be shown (after much work) that from the *hardcore measure* for this function, a *dense model* for the original set can be extracted. We make an observation of converse nature: every hardcore measure "looks"

relatively big in almost the sense needed by *The Dense Model Theorem*. So *The Dense Model Theorem* can be used on small hardcore measures to build large hardcore measures from them.

1.2 Results

With this setup an interesting question is whether there are more direct connections between LCAT, HCL, DMT (and hopefully without excessive appeal to heavy technical machinery). Proofs of each of the three results (LCAT, HCL, DMT) are known via von Neumann's Min-Max Theorem for zero sum games (alternatively the duality of Linear Programming/Hahn-Banach Theorem) with the methods of proof similar. Uniform constructions[1] as implied by Dense Model Theorem and Strong Hardcore Lemma have been given (by Zhang [Zha11] and Barak et al. [BHK09] respectively) using the framework of *Bregman Projections* and are again parallel while Trevisan et al's constructive proof of LCAT mirrors Holenstein's uniform $\text{HCL}_{\text{STRONG}}$ [Hol06, Hol05]. Furthermore, in contrapositive form all three of LCAT, HCL, DMT can also be thought of as problems in learning theory − the connection for connection for the case of HCL has been formalized by Vitaly Feldman [Fel10] as well as by Klivans and Servedio [KS99]. This lends credence to the program of trying to place LCAT, HCL, DMT on the same footing.

In [Imp09] Russell Impagliazzo gave a direct (and algorithmic) reduction from Holenstein's Strong Hardcore Lemma to a stronger form of the Dense Model Theorem ($\text{DMT}_{\text{PSEUDODENSITY}}$). Along these lines, the work that follows explores the equivalence of the Strong Hardcore Lemma, Low Complexity Approximation Theorem and Dense Model Theorem by showing the following sequence of reductions:

$$\text{HCL}_{\text{STRONG}} \xrightarrow{\text{IMPAGLIAZZO}} \text{DMT}^{\substack{\text{PSEUDODENSITY} \\ \text{PSEUDORANDOM}}}$$

Here the notation Theorem A $\rightarrow$ Theorem B means that assuming hypothesis of Theorem B along with Theorem A yields the conclusion of Theorem B. Note that both are theorems so are individually true. Therefore, for the above reduction to carry any weight it needs to be shown that there is no loss in the quality of the assumptions we need when we pass from the conclusion of one theorem to the other; furthermore, the constructive form of one should yield the constructive form of the other.

Our original contribution to this reduction here is the extension of Impagliazzo's reduction [Imp09] from sets in a probability space with base distribution fixed as uniform to distributions in arbitrary finite probability spaces[2].

Next we demonstrate:

[1] By uniform constructions we mean that the constructions carry through in the setting where all functions are replaced by uniform algorithms.

[2] The extension from sets to distributions follows from an observation by Valentine Kabanets.

$$\mathrm{DMT}^{\substack{\text{PSEUDODENSITY}\\\text{PSEUDORANDOM}}} \longrightarrow \textsc{Lcat}$$

The above reduction is obtained without loss in parameters and in most general setting. Coupled with Trevisan et al's proof of $\mathrm{HCL}_{\text{WEAK}}$ from LCAT and creative use of Zhang's $\mathrm{DMT}_{\text{ALGORITHMIC}}$ [Zha11] allows us to show:

$$\textsc{Lcat} \xrightarrow[\text{ET AL}]{\text{TREVISAN}} \mathrm{HCL}_{\text{WEAK}} \xrightarrow[\text{ALGORITHMIC}]{\text{DMT}} \mathrm{HCL}_{\text{STRONG}}$$

Again, our setting is that of arbitrary probability spaces which is more general than [Zha11]. However, because we need Trevisan et al's reduction which has weaker parameters we end up with an overall loss as we chain our reductions together to start and end at $\mathrm{HCL}_{\text{strong}}$ and this is suboptimal. Since there really is no need for such loss in our techniques, so if it were possible to bypass Trevisan et al's reduction there would be no loss. To this end we give direct and uniform constructions as needed by LCAT, HCL, DMT from a more general form of DMT ($\mathrm{DMT}_{\text{MIN-MAX}}$) based on Zhang's $\mathrm{DMT}_{\text{ALGORITHMIC}}$.

$$\begin{array}{ccccc} \mathrm{HCL}_{\text{STRONG}} & \longleftarrow & \mathrm{DMT}_{\text{MIN-MAX}} & \longrightarrow & \mathrm{DMT}^{\substack{\text{PSEUDODENSITY}\\\text{PSEUDORANDOM}}} \\ & & \downarrow & & \\ & & \textsc{Lcat} & & \end{array}$$

The $\mathrm{DMT}_{\text{MIN-MAX}}$ to $\mathrm{HCL}_{\text{STRONG}}$ reduction has much in common with Barak et al. [BHK09], and $\mathrm{DMT}_{\text{MIN-MAX}}$ to $\mathrm{DMT}^{\substack{\text{PSEUDODENSITY}\\\text{PSEUDORANDOM}}}$ is inspired by Zhang, however, our setting is more general as we work in arbitrary finite probability spaces, don't demand functions to be Boolean and give *complexity* bound on our constructions in the sense of Trevisan et al. The last reduction to LCAT is original.

With this scheme, there is no loss in terms of parameters in any of the reductions. More so, by Watson [Wat11] and Lu et al. [LTW07] in the first and the second cases the parameters obtained are known to be tight[3].

Our work also implies the converse of Trevisan et al's reduction [TTV09] from LCAT to $\mathrm{DMT}_{\text{PSEUDORANDOM}}$. In addition, we separately give the converse of their proof of $\mathrm{HCL}_{\text{WEAK}}$ from LCAT although with a polynomial loss in parameters. Exploring the methods of Trevisan et al. further we show that their reduction from LCAT to $\mathrm{HCL}_{\text{WEAK}}$ does not always extend to $\mathrm{HCL}_{\text{STRONG}}$ and we address this by generalizing LCAT to Low Complexity Rational Approximation Theorem which we show is equivalent to $\mathrm{HCL}_{\text{STRONG}}$.

Putting all this together, the work here establishes that almost equivalence ("almost" because of loss in LCAT $\rightarrow$ $\mathrm{HCL}_{\text{WEAK}}$) of LCAT, HCL, DMT and shows that tight forms of DMT, HCL and LCAT are simply manifestations of a single algorithmic technique: $\mathrm{DMT}_{\text{MIN-MAX}}$. So answers to questions regarding one are revealing about the others.

[3]We conjecture that our parameters are tight for LCAT as well

The exposition here borrows from joint work with Valentine Kabanets and Russell Impagliazzo.

Chapter 2

Main Results

2.1 Preliminaries and Observations

In this section we give the definitions and make note of some easy to see (and some not) observations.

2.1.1 Distributions and measures

2.1.1.1 Distributions

A *probability distribution* over a finite domain U is a function $\sigma : U \to [0,1]$ satisfying $\sum_{x \in U} \sigma(x) = 1$. The support of the distribution σ is $\text{SUPPORT}(\sigma) = \{x \in U \mid \sigma(x) \neq 0\}$. We define the pair $\mathcal{U} = (U, \sigma)$ to be a *finite probability space*[1]. Without loss of generality, we assume that for probability space (U, σ), $\text{SUPPORT}(\sigma) = U$ as otherwise we simply restrict U to $\text{SUPPORT}(\sigma)$. The *uniform* distribution on U is denoted by u with $u(x) = 1/|U|$ where $|U|$ the size U. Also, we use the notation $x \in \sigma$ to indicate that $x \in U$ is randomly sampled according to the probability distribution σ. Throughout U will represent a finite domain and σ a distribution over U.

Given two probability distributions ρ and σ over U, the *statistical distance* between ρ and σ, $dist(\rho, \sigma)$, is defined as the half of the ℓ_1-norm of the vector $\rho - \sigma$, i.e.,

$$dist(\rho, \sigma) = (1/2) \cdot \sum_{x \in U} |\rho(x) - \sigma(x)|.$$

For a set T and a distribution ρ, we define $T[\rho] = \mathbb{E}_\rho[T] = \sum_{x \in T} \rho(x)$. This yields an

[1] Every singleton is measurable here so without loss of generality the standard probability triple is reduced to this pair

equivalent characterization of $dist(\rho, \sigma)$:

$$dist(\rho, \sigma) = \max_{T \subseteq U} |T[\rho] - T[\sigma]|,$$

This allows us to formalize what it means for two distributions to be close: two distributions ρ and σ are considered to be ϵ-*close* (with the parameter $\epsilon \in [0, 1]$) if $dist(\rho, \sigma) \leq \epsilon$.

2.1.1.2 Measures

A *measure* over the finite domain U is any function $\mu : U \to [0, 1]$. Note that Boolean measure is simply the indicator function of the support of the measure. Informally, a measure μ can be thought of as a generalization of a set in the sense that for each $x \in U$, $\mu(x)$ measures the likelihood of x being in the set. More precisely, a measure μ specifies a probability distribution over the subsets of U, where a random subset R is picked by placing each $x \in U$ into R with probability $\mu(x)$ independently. The notation $R \in \mu$ is used to indicate that a set $R \subseteq U$ is randomly chosen according to the measure μ.

Since a set $A \subseteq U$, A can be thought of as a Boolean measure $\mathbf{1}_A : U \to \{0, 1\}$(where $\mathbf{1}_A(x) = 1$ iff $x \in A$, 0 otherwise), the complement of the set A, A^c is represented by measure $\mathbf{1}_{A^c} = \mathbf{1} - \mathbf{1}_A$ (without loss of generality, we'll use A for $\mathbf{1}_A$ when it's clear from context that the object in question is a function). Extending the idea of thinking of measures as sets, we can define the *complement* of a measure μ over U by $\bar{\mu} = 1 - \mu$.

2.1.1.3 Density

For a subset A of the domain U in the probability space (σ, U) the probability that $x \in A$, $\mathbb{P}_\sigma[x \in A]$, measures the "size" of the set A relative to U.

We formalize this idea as the density of a set A in probability space $\mathcal{U} = (U, \sigma)$, $d_\sigma(A)$. Density, the probability mass assigned to A by the distribution σ, has multiple equivalent mathematical forms:

$$d_\sigma(A) = \mathbb{P}_\sigma[x \in A] = \mathbb{E}_\sigma[A] = \sum_{x \in U} A(x)\sigma(x).$$

In case the distribution σ is the uniform distribution over the entire domain, $\mathbb{P}_u[x \in A] = |A|/|U| = d_u(A)$ which is where the idea of using density as relative size arises.

More generally, for a measure $\mu : U \to [0, 1]$, the density of the measure μ in the probability space $\mathcal{U} = (U, \sigma)$, denoted by $d_\sigma(\mu)$, is defined as

$$d_\sigma(\mu) = \mathbb{E}_\sigma[\mu] = \sum_{x \in U} \mu(x)\sigma(x).$$

Equivalently, the density of a measure is the average density of the set randomly chosen according to the measure.

Claim 2.1.1. $d_\sigma(\mu) = \mathbb{E}_{A\in\mu}[d_\sigma(A)]$.

Proof. $\mathbb{E}_{A\in\mu}[d_\sigma(A)] = \mathbb{E}_{A\in\mu}\mathbb{E}_{x\in\sigma}[A(x)] = \mathbb{E}_{x\in\sigma}\mathbb{E}_{A\in\mu}[A(x)] = \mathbb{E}_{x\in\sigma}[\mu(x)] = d_\sigma(\mu)$ □

For any non-empty subset A of U, $\sigma|_A$ denotes the *distribution σ restricted to A*, which is defined as the conditional probability of sampling an element x according to σ conditioned on $x \in A$, i.e., for each $x \in U$:

$$\sigma\big|_A(x) = \sigma(x)A(x)/d_\sigma(A)$$

The significance of density lies in that it allows for decomposition of the distribution σ in the following sense:

$$\sigma = d_\sigma(A)\frac{A\cdot\sigma}{d_\sigma(A)} + d_\sigma(\bar{A})\frac{\bar{A}\cdot\sigma}{d_\sigma(\bar{A})} = d_\sigma(A)\sigma\big|_A + d_\sigma(\bar{A})\sigma\big|_{\bar{A}} \tag{2.1}$$

Note that every measure μ over a probability space (U,σ) *induces a probability distribution*, denoted by μ_σ, where $\mu_\sigma(x) = \mu(x)\sigma(x)/d_\sigma(\mu)$ for each $x \in U$. In particular for any constant measure $c\mathbf{1}$ the induced distribution, $(c\mathbf{1})_\sigma = \sigma$. The following lemma shows that the above decomposition is simply an artifact of identity $\mathbf{1} = \mathbf{1}_A + \mathbf{1}_{A^c}$.

Lemma 2.1.2. *Suppose μ,γ,η are measures in (U,σ) with $\mu = \gamma + \eta$ and $\mathbb{E}_\sigma[\gamma]/\mathbb{E}_\sigma[\mu] = \delta, \mathbb{E}_\sigma[\eta]/\mathbb{E}_\sigma[\mu] = \bar{\delta}$ then $\mu_\sigma,\gamma_\sigma,\eta_\sigma$ satisfy $\mu_\sigma(x) = \delta\gamma_\sigma(x) + \bar{\delta}\eta_\sigma(x)$.*

Proof.

$$\mu_\sigma(x) = \frac{(\gamma(x)+\eta(x))\sigma(x)}{\mathbb{E}_\sigma[\mu]} = \frac{\gamma(x)\sigma(x)}{\mathbb{E}_\sigma[\mu]}\frac{\mathbb{E}_\sigma[\gamma]}{\mathbb{E}_\sigma[\gamma]} + \frac{\eta(x)\sigma(x)}{\mathbb{E}_\sigma[\mu]}\frac{\mathbb{E}_\sigma[\eta]}{\mathbb{E}_\sigma[\eta]} = \delta\gamma_\sigma(x) + \bar{\delta}\eta_\sigma(x)$$

□

Based on the decomposition as given by equation 2.1, the distributions $\gamma_\sigma,\eta_\sigma$ can be thought of as δ-dense, $\bar{\delta}$-dense respectively in μ_σ. The conditions $\mathbb{E}_\sigma[\gamma]/\mathbb{E}_\sigma[\mu] = \delta, \mathbb{E}_\sigma[\eta]/\mathbb{E}_\sigma[\mu] = \bar{\delta}$ also tie in with intuition of density as relative size. We can abstract away from distributions related to this additive decomposition of a measure and consider density for distributions in general by noting that for any two probability distributions ρ and σ over U, it's always possible to express σ as a convex combination of ρ and τ, for some distribution τ:

$$\sigma(x) = \delta\cdot\rho(x) + (1-\delta)\cdot\tau(x) \geq \delta\cdot\rho(x), \tag{2.2}$$

with parameter $\delta \in [0,1]$ which means that to sample from σ, we need to sample from ρ with probability at least δ. We say that the distribution ρ is δ-dense in σ. The maximum value δ, $0 \leq \delta \leq 1$, such that σ can be expressed by Eq. (2.2) is defined as the *density of ρ in σ*, and is denoted by $d_\sigma(\rho)$. That is,

$$d_\sigma(\rho) = \max\{\delta \in [0,1] \mid \forall\, x \in U \;\; \sigma(x) \geq \delta \cdot \rho(x)\}.$$

Next we focus on the relation between the density of a measure μ in probability space (U, σ) and the density of the induced distribution μ_σ in the distribution σ. To start we consider sets, i.e., Boolean measures.

Claim 2.1.3. *For every set $\emptyset \neq A \subseteq U$, $d_\sigma(A) = d_\sigma(\sigma|_A)$.*

Proof. We can write σ as the convex combination $\sigma = d_\sigma(A) \cdot \sigma|_A + d_\sigma(\bar{A}) \cdot \sigma|_{\bar{A}}$, which implies by definition that $d_\sigma(\sigma|_A) \geq d_\sigma(A)$. Next consider any $x_0 \in A$. Again by definition we get that $\sigma(x_0) \geq d_\sigma(\sigma|_A) \cdot \sigma(x_0)A(x_0)/d_\sigma(A)$, so summing over A implies that $d_\sigma(\sigma|_A) \leq d_\sigma(A)$. □

Observe that, for a set $A \subseteq U$, we have $A_\sigma = \sigma|_A$. We also have the following easy observation:

Claim 2.1.4. *For arbitrary probability space (U, σ) and any measure μ over U, we have $d_\sigma(\mu_\sigma) \geq d_\sigma(\mu)$.*

Proof. For every $x \in U$ such that $\mu(x) \neq 0$, we have $\sigma(x) = \mu_\sigma(x) \cdot (d_\sigma(\mu)/\mu(x)) \geq d_\sigma(\mu) \cdot \mu_\sigma(x)$, as $\mu(x) \leq 1$. For any $x \in U$ with $\mu(x) = 0$ (and hence $\mu_\sigma(x) = 0$), we also get that $\sigma(x) \geq d_\sigma(\mu) \cdot \mu_\sigma(x)$. □

Remark 2.1.5. *Unlike the case of sets (or Boolean measures $\mu : U \to \{0,1\}$), for $[0,1]$-valued measures μ, the density of the induced distribution μ_σ can be arbitrarily bigger than the density of the measure μ in (U, σ). For example, consider $\mu(x) = \epsilon$ for every $x \in U$, where $0 < \epsilon < 1$ is any constant. The density of the measure μ is ϵ. On the other hand, the induced distribution μ_σ is identical to the distribution σ, and so has density 1 inside σ. Even though the density is not preserved by going from a measure to the induced distribution, we know by Claim 2.1.4 that it is not reduced.*

The following claim relates a measure μ and its scaled version $\mu' = c \cdot \mu$, for a constant $c > 0$.

Claim 2.1.6. *For any measure μ over (U, σ) and any constant $c > 0$, the measure $\mu' = c \cdot \mu$ satisfies:*

1. $d_\sigma(\mu') = c \cdot d_\sigma(\mu)$

Proof. $d_\sigma(\mu') = \mathbb{E}_\sigma[c \cdot \mu] = c \cdot \mathbb{E}_\sigma[\mu] = c \cdot d_\sigma(\mu)$. □

2. $\mu_\sigma = \mu'_\sigma$

Proof. Since for every $x \in U, \mu'(x) = c\mu(x)$, therefore, $\mu'_\sigma(x) = c\mu(x)\sigma(x)/d_\sigma(\mu') = \mu(x)\sigma(x)/d_\sigma(\mu) = \mu_\sigma(x)$ □

2.1.1.4 Indistinguishability

Suppose $\mathcal{F}$ be a set of bounded functions $f : U \to [0,1]$. Usually, $\mathcal{F}$ will contain the constant functions **1** and **0** and be closed under complement; such a set $\mathcal{F}$ is called a *class* of functions. For any $[0,1]$-bounded function f, we define its complement $\bar{f} = 1 - f$. By default, when saying "class $\mathcal{F}$", we mean a class of $[0,1]$-valued functions $f : U \to [0,1]$. When we need to restrict to the class of Boolean functions $f : U \to \{0,1\}$, we will indicate this by explicitly saying "Boolean class $\mathcal{F}$".

The set $\mathcal{F}$ can be thought of as a collection of "tests" that can be used to distinguish distributions and measures from one another by noting that a bounded function can be "applied" to a distribution (or measure) as follows: for any function $f : \mathcal{U} \to [0,1]$ and a probability distribution ρ over the universe U, we denote by $f[\rho]$ the *expectation of f over the probability distribution* ρ, i.e.,

$$f[\rho] = \mathbb{E}_\rho[f] = \sum_{x \in U} f(x)\rho(x).$$

For $f : U \to [0,1]$ and any measure μ over a probability space (U, σ) we define the *expectation of f over the measure* μ, denoted by $f_\sigma[\mu]$, as the expectation of f over the distribution μ_σ, i.e.,

$$f_\sigma[\mu] = \mathbb{E}_{\mu_\sigma}[f] = \sum_{x \in U} f(x)\mu(x)\sigma(x)/d_\sigma(\mu) = \frac{1}{d_\sigma(\mu)} \cdot \mathbb{E}_\sigma[f \cdot \mu].$$

Note that $f[\rho] = d_\rho(f)$ is the density of f (when viewed as a measure) in the probability space (U, ρ); similarly, $f_\sigma[\mu]$ is the density of f in the probability space (U, μ_σ).

Recall the definition of ϵ-closeness: distributions ρ_1 and ρ_2 over U are ϵ-close if, for every subset $T \subseteq U$, we have $|T[\rho_1] - T[\rho_2]| \leq \epsilon$. By restricting our tests to the set $\mathcal{F}$, we get the notion of closeness, or indistinguishability, relative to $\mathcal{F}$. For a parameter $\epsilon \in [0,1]$, we say that two distributions ρ_1 and ρ_2 over the universe U are $(\epsilon, \mathcal{F})$-*indistinguishable* if, for every $f \in \mathcal{F}$, we have

$$|f[\rho_1] - f[\rho_2]| \leq \epsilon.$$

Definition 2.1.7. A distribution ρ over U is called $(\epsilon, \mathcal{F})$-*pseudorandom over the probability space* (U, σ) if it is $(\epsilon, \mathcal{F})$-indistinguishable from σ.

Note that for $\sigma = u$, we get the standard notion of pseudorandom distribution as indistinguishable from the uniform distribution.

For two functions $g : U \to [0,1]$ and $h : U \to [0,1]$, if, for all $f \in \mathcal{F}$,

$$|\mathbb{E}_\sigma[f(g-h)]| \leq \epsilon.$$

then, we say that h is an $(\epsilon, \mathcal{F})$-*approximation* of g over the probability space (U, σ).

2.1.2 Models and pseudodensity

Using the definition of density for distributions for a restricted collection $\mathcal{F}$ of tests, we get the following definition of *pseudodensity* for distributions.

Definition 2.1.8. For distributions σ and ρ over U, for a set $\mathcal{F}$ of $[0,1]$-valued functions over U, and parameters $\epsilon, \delta \in [0,1]$, we say[2] that the distribution ρ has $(\epsilon, \mathcal{F})$-*pseudodensity* δ *inside* σ if, for all $f \in \mathcal{F}$, $f[\sigma] \geq \delta \cdot f[\rho] - \epsilon$.

A test $f \in \mathcal{F}$ such that $f[\sigma] < \delta \cdot f[\rho] - \epsilon$ witnesses the fact that ρ has density less than δ; such a test is called (ϵ, δ)-*distinctive*. The negative ϵ term is needed to ensure that $poly(1/\epsilon)$ random samples are enough to verify that a given $f \in \mathcal{F}$ is indeed (ϵ, δ)-distinctive. Thus, ρ has pseudo-density δ inside σ if there are no (ϵ, δ)-distinctive tests in $\mathcal{F}$.

For a set $A \subseteq U$, a measure μ over the probability space (U, σ), and a parameter $\epsilon \in [0,1]$, we say that μ is an $(\epsilon, \mathcal{F})$-*model for* A if the induced distributions A_σ and μ_σ are $(\epsilon, \mathcal{F})$-indistinguishable. Similarly, for a distribution ρ over U, a measure μ over U is an $(\epsilon, \mathcal{F})$-*model* for ρ in the probability space (U, σ) if ρ and μ_σ are $(\epsilon, \mathcal{F})$-indistinguishable.

Remark 2.1.9. *Note that if μ is a model for some distribution ρ (or a set A), then, by Claim 2.1.6, so is $\mu' = c \cdot \mu$, for any constant $c > 0$. By choosing c appropriately, we can thus get from μ another model μ' of smaller density.*

We will be interested in the case where A is a set of small density, yet is indistinguishable from a measure μ of high density (so A has a "large" model μ). Such a large model for A turns out to exist if the density of A "looks" big to the tests in $\mathcal{F}$. We define the notion of pseudo-density for sets next.

By Claim 2.1.3, a set $A \subseteq U$ in a probability space (U, σ) has (true) density $d_\sigma(A)$ equal to the density of the distribution $\sigma|_A$ inside σ. Thus, the density of A is at least δ (for some $\delta \in [0,1]$) iff, for all $x \in U$, $\sigma(x) \geq \delta \cdot \sigma|_A(x)$. The latter is true iff, for all tests $T : U \to [0,1]$, we have $T[\sigma] \geq \delta \cdot T[\sigma|_A]$. By restricting the tests to the set $\mathcal{F}$, we derive the notion of *pseudodensity* for sets.

[2] Equivalently ρ is δ-$(\epsilon, \mathcal{F})$-pseudodense

Definition 2.1.10. For $A \subseteq U$ over the probability space (U, σ), a set $\mathcal{F}$ of $[0,1]$-valued functions over U, and parameters $\epsilon, \delta \in [0,1]$, we say that A has $(\epsilon, \mathcal{F})$-*pseudodensity* δ *inside* (U, σ) if, for every $f \in \mathcal{F}$, we have $f[\sigma] \geq \delta \cdot f[\sigma|_A] - \epsilon$.

Remark 2.1.11. *It follows from the definition of density that if a distribution ρ has actual density δ then for any $\epsilon > 0$ and any class of bounded functions $\mathcal{F}$, ρ will be δ-$(\epsilon, \mathcal{F})$-pseudodense*

2.1.3 Hardness

Next we introduce what it means for functions and measures to be hard and hardcore respectively. The definitions here are generalizations of those that are standardly used as we allow our class of functions to be $[0,1]$-bounded and not just $\{0,1\}$-Boolean.

Definition 2.1.12. For a Boolean function $g : U \to \{0,1\}$ over a probability space (U, σ), a collection of $[0,1]$-bounded functions $\mathcal{F}$ over U and a parameter $\delta \in [0,1]$, we say that g is $(\delta, \mathcal{F})$-hard in (U, σ) if, for all $f \in \mathcal{F}$,

$$\mathbb{E}_\sigma[|f - g|] \geq \delta.$$

Definition 2.1.13. For a Boolean function $g : U \to \{0,1\}$ over a probability space (U, σ), a collection of $[0,1]$-bounded functions $\mathcal{F}$ over U, a measure μ over U, and a parameter $\epsilon \in [0,1]$, we say that the measure μ is $(\epsilon, \mathcal{F})$-*hardcore for g in* (U, σ) if, for all $f \in \mathcal{F}$,

$$\mathbb{E}_{\mu_\sigma}[|g - f|] \leq 1/2 + \epsilon.$$

Note that for $\{0,1\}$-Boolean functions ϕ, g and any distribution ρ, $\mathbb{P}_\rho[g \neq \phi] = \mathbb{E}_\rho[|g - \phi|]$ so the above definitions reduce to the standard ones for Boolean classes.

We'll be working in the setup that $\mathcal{F}$ is a class of functions. This is reasonable to demand, since all we need is that $\mathcal{F}$ contain the constant functions and be closed under complementation. In case $\mathcal{F}$ is a class the above definitions become symmetric about half which is nice.

Lemma 2.1.14. *Given $\delta, \epsilon > 0$, a class $\mathcal{F}$ and $g : U \to \{0,1\}$, $(\delta, \mathcal{F})$-hard on (σ, U), the following hold:*

1. $\mathbb{E}_\sigma[|f - g|] \leq 1 - \delta$.

   Proof. Since $\bar{f} \in \mathcal{F}$, so $\mathbb{E}_\sigma[|\bar{f} - g|] = 1 - \mathbb{E}_\sigma[|f - g|] \geq \delta$, therefore, $1 - \delta \geq \mathbb{E}_\sigma[|f - g|]$. □

2. *A measure γ is $(\epsilon, \mathcal{F})$-hardcore for g w.r.t. σ iff $\left|\mathbb{E}_{\gamma_\sigma}[|g - f|] - \frac{1}{2}\right| + \epsilon$ for every $f \in \mathcal{F}$.*

Proof. The *only if* direction is trivial since $\left|\mathbb{E}_{\gamma_\sigma}[|g-f|]-\frac{1}{2}\right| \leq \epsilon$ implies $\mathbb{E}_{\gamma_\sigma}[|g-f|] \leq \frac{1}{2}+\epsilon$. To see that γ being $(\epsilon, \mathcal{F})$-*hardcore* implies $\left|\mathbb{E}_{\gamma_\sigma}[|g-f|]-\frac{1}{2}\right| \leq \epsilon$, note that by closure of $\mathcal{F}$ under complementation $\mathbb{E}_{\gamma_\sigma}[|g-(1-f)|] \leq \frac{1}{2}+\epsilon$ follows giving $\mathbb{E}_{\gamma_\sigma}[|g-f|] \geq \frac{1}{2}-\epsilon$, so $\left|\mathbb{E}_{\gamma_\sigma}[|g-f|]-\frac{1}{2}\right| \leq \epsilon$ as needed. □

3. *If* $2\epsilon + 2\delta \geq 1$ *then the measure* $\mathbf{1}$ *is* $(\epsilon, \mathcal{F})$-*hardcore.*

Proof. Now $2\epsilon+2\delta \geq 1$ implies $\delta \geq \frac{1}{2}-\epsilon$ and as for every $f \in \mathcal{F}$, $\mathbb{E}_\sigma[|g-f|] \leq 1-\delta$ because $\mathcal{F}$ is closed under complementation, therefore, $\mathbb{E}_{\mathbf{1}_\sigma}[|g-f|] = \mathbb{E}_\sigma[|g-f|] \leq 1-\frac{1}{2}+\epsilon = \frac{1}{2}+\epsilon$. □

2.1.4 Relative Complexity

Here we define the complexity of functions relative to the given class $\mathcal{F}$ of $[0,1]$-bounded functions, following [TTV09].

Definition 2.1.15. For a function $h : U \to [0,1]$, its *complexity relative to* $\mathcal{F}$, denoted by $\textsc{comp}_{\mathcal{F}}[h]$, is the number of operations needed to express h in terms of functions $f \in \mathcal{F}$, where the allowed operations are: *(i)* scalar multiplication, *(ii)* scalar addition, *(iii)* addition, *(iv)* subtraction, *(v)* multiplication, *(vi)* truncation (*trunc*), and *(vii)* threshold (*th*).

The latter two operations are defined as follows: Given $f : U \to \mathbb{R}$, $\theta \in \mathbb{R}$, and $a < b \in \mathbb{R}$,

$$f_\theta(x) = th_\theta[f](x) = \begin{cases} 1 & \text{if } f(x) \geq \theta \\ 0 & \text{otherwise} \end{cases} \text{ and } trunc_a^b[f](x) = \begin{cases} b & \text{if } f(x) > b \\ f(x) & \text{if } f(x) \in [a,b] \\ a & \text{if } f(x) < a. \end{cases} \tag{2.3}$$

Note that for a function $g : U \to [0,a]$ with $a > 0$, we have $trunc_0^1[g](x) = g(x)(1 - th_1[g](x)) + th_1[g](x)$, and so the operation *trunc* is not strictly necessary, but is included for convenience.

This allows us to formalize the idea of functions of small complexity w.r.t. to a class:

Definition 2.1.16. Given $t \in \mathbb{N}$, $\mathcal{F}_t = \{f : \textsc{comp}_{\mathcal{F}}[f] \leq t\}$.

In the set up of multiplicative updates that we utilize, the operation $\beta^{f(x)}$ (for any constant β and $[0,1]$-bounded f) is very useful. Although functions which involve such exponentiation don't have small complexity as defined in definition 2.1.15, they may still be efficiently describable. So we extend the permissible set of operations to include exponentiation of constants, and

consider complexity of a function h relative to a class $\mathcal{F}$ ($\textsc{comp}^*_{\mathcal{F}}[h]$) as in definition 2.1.15 but with respect to this extended set of operations. Analogous to definition 2.1.16, this yields:

Definition 2.1.17. Given $t \in \mathbb{N}$, $\mathcal{F}^*_t = \{f : \textsc{comp}^*_{\mathcal{F}}[f] \leq t\}$.

Also, of relevance are the following collections of functions:

Definition 2.1.18. Given $g : U \to \{0,1\}$ and a class $\mathcal{F}$ on U, $g \oplus \mathcal{F} = \{|g - f| : f \in \mathcal{F}\}$.

So $\mathcal{F}_t$ is the set of functions with complexity at most t w.r.t. $\mathcal{F}$, while $g \oplus \mathcal{F}$ in the case $\mathcal{F}$ is boolean consists of indicator functions of event $g(x) \neq f(x)$.

Definition 2.1.19. For any collection of $[0,1]$-valued functions $\mathcal{F}$, $\textsc{Th}_\lambda[\mathcal{F}] = \{th_\theta[\frac{1}{\lambda'}\sum_{i=1}^{\lambda'} f_i] : f_i \in \mathcal{F}, \theta \in [0,1], \lambda' \leq \lambda\}$.

That is, $\textsc{Th}_\lambda[\mathcal{F}]$ is the collection of Linear Threshold Functions of particularly simple form built from at most λ functions from $\mathcal{F}$.

Definition 2.1.20. Given $\epsilon > 0$, $\mathcal{F}^\epsilon = \{th_t[f], \overline{th_t[f]} : f \in \mathcal{F}, t \in \{0, 1, n\epsilon : n \in \mathbb{N}\} \cap [0,1]\}$.

Definition 2.1.21. Given a class $\mathcal{F}$, $\mathcal{CH}[\mathcal{F}]$ is the convex hull of $\mathcal{F}$, i.e. $\mathcal{CH}[\mathcal{F}] = \{\sum_i c_i f_i : f_i \in \mathcal{F}, c_i \in (0,1], \sum_c c_i = 1\}$.

Remark 2.1.22. *Note that $g \oplus \mathcal{F}$ is closed under complementation as for Boolean g and class $\mathcal{F}$, $\overline{|g-f|} = 1 - |g-f| = |g - \bar{f}| \in g \oplus \mathcal{F}$, however, $g \oplus \mathcal{F}$ is not a class as $\mathbf{1}, \mathbf{0} \notin g \oplus \mathcal{F}$ unless $g \in \mathcal{F}$.*

Remark 2.1.23. *If $\mathcal{F}$ is a boolean class then $\mathcal{F} = \mathcal{F}^\epsilon$ for any $\epsilon > 0$ and for any $\lambda \in \mathbb{N}$, $\textsc{Th}_\lambda[\mathcal{F}]$ is also a class.*

Remark 2.1.24. *If f_i are all $\{0,1\}-$Boolean, then $th_\theta[\frac{1}{\lambda'}\sum_{i=1}^{\lambda'} f_i] = 1$ iff $\lceil \theta\lambda' \rceil$ out of λ' f_i are* 1.

2.1.5 Bregman Projections

Both entropy and the Kullback-Leibler (KL) divergence are versatile tools, though usually only considered for finite domains with base distribution as uniform; however, they extend in the obvious way to general finite probability spaces $\mathcal{U} = (U, \sigma)$ (there is some work that needs to be done to verify that the new definition of KL-divergence is indeed a *divergence*, but we leave that for Chapter 5).

2.1.5.1 Generalized Entropy/KL-divergence

Given a finite domain U (with $N = |U|$), the elements of U can be ordered in some canonical way and assigned indices from $[N]$. The distribution σ over U can be viewed as an N-dimensional vector over $\mathbb{R}^+$ (the positive reals[3]): $\sigma = (\sigma_1, \ldots, \sigma_N)$, with σ_i being the probability mass assigned by σ to the ith element of U.

For an arbitrary vector $\mathbf{x} = (x_1, \ldots, x_N) \in (\mathbb{R}^+)^N$, we define the *generalized entropy function for* $\mathcal{U} = (U, \sigma)$ as $\mathrm{ent}(\mathbf{x}) = -\sum_{i=1}^N \sigma_i x_i \log x_i$. As for the standard entropy function $\textsc{Ent}(\mathbf{x}) = -\sum_{i=1}^N x_i \log x_i$, one can show (following [CZ97][4]) that $-\mathrm{ent}(\mathbf{x})$ is a Bregman function, with the associated *Bregman divergence* $\mathrm{D}(\mathbf{x}||\mathbf{y})$ between $\mathbf{x}$ and $\mathbf{y}$ from $(\mathbb{R}^+)^N$ defined as follows:

$$\mathrm{D}(\mathbf{x}||\mathbf{y}) = \sum_{i=1}^N \sigma_i x_i \log\left(x_i/y_i\right) - \sum_{i=1}^N \sigma_i x_i + \sum_{i=1}^N \sigma_i y_i = \mathbb{E}_\sigma[\mathbf{x} \cdot \log(\mathbf{x}/\mathbf{y})] - \mathbb{E}_\sigma[\mathbf{x}] + \mathbb{E}_\sigma[\mathbf{y}];$$

the latter generalizes the standard KL divergence $KL(\mathbf{x}||\mathbf{y}) = \sum_{i=1}^N x_i \log\left(x_i/y_i\right) - \sum_{i=1}^N x_i + \sum_{i=1}^N y_i$.

Definition 2.1.25. Let $\emptyset \neq \Gamma \subseteq [0,1]^N$ be any closed convex set of measures, and let $\phi \in [0,1]^N$ be any measure. We define the *Bregman projection* of ϕ onto Γ, denoted by $P_\Gamma \phi$, as

$$P_\Gamma \phi = \arg\min_{\gamma \in \Gamma} \mathrm{D}(\gamma||\phi).$$

By [CZ97], since $\mathrm{D}(\gamma||\phi)$ is a *Bregman Divergence*, therefore $\mathrm{D}(\gamma||\phi) \geq 0$ (with $\mathrm{D}(\gamma||\phi) = 0$ iff $\gamma = \phi$) which means $\phi \in \Gamma$ implies that $P_\Gamma \phi = \phi$.

Definition 2.1.26. For $\delta \in [0,1]$, $\Gamma_\delta = \{\phi \in [0,1]^N \mid d_\sigma(\phi) \geq \delta\}$.

Note that Γ_δ is a convex and closed set of measures. We denote by $P_\delta \phi$ the Bregman projection of a measure ϕ onto Γ_δ.

Now any Bregman divergence, hence D here, satisfies Bregman's theorem:

Theorem 2.1.27 (Bregman[Bre67])**.** *Let $\Gamma \subseteq [0,1]^N$ be any non-empty closed convex set of measures. Let γ, ϕ be measures such that $\gamma \in \Gamma$. Then $\mathrm{D}(\gamma||P_\Gamma\phi) + \mathrm{D}(P_\Gamma\phi||\phi) \leq \mathrm{D}(\gamma||\phi)$, and, in particular, $\mathrm{D}(\gamma||P_\Gamma\phi) \leq \mathrm{D}(\gamma||\phi)$.*

Bregman projections can be slightly relaxed to *Approximate Bregman Projections* which are efficiently computable when the projection is on to Γ_δ:

[3] Recall we assumed that on U σ never vanishes

[4] $\mathrm{ent}(\mathbf{x})$ is simply $\textsc{Ent}(\mathbf{x})$ with each coordinate scaled by a positive constant, and such scaling does not affect the continuity, differentiability, and limit properties, as required for Bregman functions.

Definition 2.1.28. For a non-empty closed convex set Γ of measures and a parameter $\alpha \geq 0$, a measure $\phi^* \in \Gamma$ is called an *α-approximate Bregman projection* of a measure ϕ onto Γ, denoted by α-$P_\Gamma\phi$, if for all $\gamma \in \Gamma$, $\mathrm{D}(\gamma||\phi^*) \leq \mathrm{D}(\gamma||P_\Gamma\phi) + \alpha$.

Generalizing the results in [BHK09], we get the following results about projections P_δ. The proofs (which we give in Chapter 5) are extensions of [BHK09].

Theorem 2.1.29 (generalizing [BHK09])**.** *Let ϕ be any measure over the probability space $\mathcal{U} = (U, \sigma)$ such that $d_\sigma(\textsc{support}[\phi]) \geq \delta$, for some $\delta \in [0, 1]$. Let $c \geq 1$ be the smallest constant such that the measure $\mu = trunc_0^1[c \cdot \phi]$ has $d_\sigma(\mu) = \delta$. Then $P_\delta\phi = \mu$.*

As pointed out above Approximate Bregman Projections onto Γ_δ are efficiently computable and this is formalized as follows:

Lemma 2.1.30 (generalizing [BHK09][5])**.** *For $\delta \in [0, 1]$, let ϕ be a measure over the probability space $\mathcal{U} = (U, \sigma)$ such that $P_\delta\phi = trunc_0^1[c \cdot \phi]$ for $c \in [1, 1 + \zeta]$, where $\zeta > 0$. Suppose we have oracle access to ϕ, and that we can sample an element from $\mathcal{U}$ in time t. Then, for any $0 < p < 1$, we can compute an implicitly represented approximate projection $\epsilon\delta$-$P_\delta\phi$ in time $O\left(t\delta^{-1}\epsilon^{-2}(\log\log\zeta\epsilon^{-1} + \log p^{-1})\right)$, with probability $1 - p$. Moreover, the computed approximate projection has the form $trunc_0^1[\tilde{c} \cdot \phi]$, for some $\tilde{c} \in [1, 1 + \zeta]$.*

2.1.5.2 Game playing by multiplicative updates

Next we get the following generalization of the Total Loss Lemma due to Barak et al [BHK09] for the general probability space $\mathcal{U} = (U, \sigma)$. The lemma captures the setting of a two-player game where one of the players whose strategies come from a closed and convex set of measures, Γ, starts off by playing an arbitrary measure μ^1. The other player plays penalty functions countering on the first players choice. In round t the first player updates his measure from previous round to μ^{t+1} after taking into account the "penalty" function m^t played by the second player. The update is through a simple multiplicative procedure, followed by an approximate Bregman projection onto the desired set Γ of measures (since the player's strategies must come from Γ and a simple multiplicative update cannot ensure this). Intuitively, the lemma says that the expected performance of the measures obtained in this game is not much worse than the performance achieved by an *single* measure μ where μ is completely *arbitrary* and could depend on all the penalty functions m^t played in the game.

The following lemma is fundamental to our uniform constructions and is an extension of the work of Barak et al [BHK09] to our setting of arbitrary finite probability spaces. The proof follows that of Lemma 4.1 from [BHK09] closely[6].

[5]The approximation parameter in [BHK09] is $\epsilon\delta N$ rather than $\epsilon\delta$ in our case; this difference is due to our scaling of N-dimensional vectors by the distribution σ.

[6][BHK09] had a typo: the factor $1/\epsilon$ was missing in the last term on the right-hand side of Eq. (2.4).

Lemma 2.1.31 (Generalized Total Loss Lemma, generalizing [BHK09, Lemma 4.1])**.** *For a probability space $\mathcal{U} = (U, \sigma)$ with $|U| = N$, let Γ be a non-empty closed convex set of measures over U. Let $\epsilon \in (0, \frac{1}{2})$ and let $T \in \mathbb{N}$ be arbitrary. Let $\mu^1 \in \Gamma$ be an arbitrary measure over U, and, for $1 \leq t < T$, let $m^t : U \to [0,1]$ be an arbitrary function ("penalty"). Define, for each $1 \leq t < T$, the following measures:*

$$\phi^{t+1}(x) = (1-\epsilon)^{m^t(x)} \cdot \mu^t(x), \qquad \text{and} \qquad \mu^{t+1} = \alpha\text{-}P_\Gamma \phi^{t+1}.$$

Then, for every measure $\mu \in \Gamma$, we have

$$\sum_{t=1}^{T} \mathbb{E}_\sigma(\mu^t \cdot m^t) - \tfrac{\alpha}{\epsilon} \cdot T \leq (1+\epsilon) \cdot \sum_{t=1}^{T} \mathbb{E}_\sigma(\mu \cdot m^t) + \tfrac{1}{\epsilon} \cdot \mathrm{D}(\mu||\mu^1). \tag{2.4}$$

2.2 Main Results

In this section we introduce our results with technical details and give a comparison of our results with other recent work.

2.2.1 Low Complexity Approximations

Theorem 2.2.1 (Low Complexity Approximation Theorem (LCAT) [TTV09])**.** *Given any $\epsilon > 0$, there exists a $\lambda = O(\epsilon^{-2})$ such that the following holds: suppose $\mathcal{U} = (U, \sigma)$ is any finite probability space, and $\mathcal{F}$ is any class of functions over U, then for every $g : U \to [0,1]$, there exists a function $h \in \mathcal{F}_\lambda$ such that h is an $(\epsilon, \mathcal{F})$-approximation of g w.r.t. σ.*

Remark 2.2.2. *Since $\mathbf{1} \in \mathcal{F}$, therefore $|\mathbb{E}_\sigma[h] - \mathbb{E}_\sigma[g]| \leq \epsilon$.*

Theorem 2.2.1 was first proven by Trevisan et al. [TTV09] with $\lambda = O(\epsilon^{-2}\log(\epsilon^{-1}))$ in the general case and $\lambda = O(\epsilon^{-2})$ in the case class $\mathcal{F}$ is Boolean. Our construction is uniform[7] in general with $\lambda = O(\epsilon^{-2})$. Note that this result can be thought of as special case of following class of theorems:

> **Generic Low Complexity Approximation Theorem:** For every $\epsilon > 0$, there exists a $\lambda = \mathit{poly}(1/\epsilon)$ such that the following holds for any probability space $\mathcal{U} = (U, \sigma)$ and any class $\mathcal{F}$ over U: there exists a class $\mathcal{F}'$ parametrized by $\mathcal{F}, \lambda$ such that every function $g : U \to [0,1]$ has an $(\epsilon, \mathcal{F})$-approximation $h \in \mathcal{F}'$ w.r.t. σ.

Here Theorem 2.2.1 gives λ and $\mathcal{F}'$ explicitly.

[7] As remarked in Chapter 1, by uniform proof we mean in the setting of uniform algorithms.

2.2.2 Hardcore Measures

Theorem 2.2.3 (Strong Hardcore Lemma: $\text{HCL}_{\text{STRONG}}$, [Hol05, BHK09])**.** *Given any $\epsilon, \delta \in [0,1]$, there exists a $\lambda = O(\epsilon^{-2}\log(\delta^{-1}))$ such that the following holds: suppose $\mathcal{U} = (U, \sigma)$ is any finite probability space and $\mathcal{F}$ is any class of functions over U. If a function $g : U \to \{0,1\}$ is $(\delta, \text{TH}_\lambda[\mathcal{F}])$-hard in $\mathcal{U}$, then there is a 2δ-dense measure $\mu \in [g \oplus \mathcal{F}]^*_{O(\lambda)}$ which is an $(\epsilon, \mathcal{F})$-hardcore for g in $\mathcal{U}$.*

Remark 2.2.4. *In case $\mathcal{F}$ is Boolean, or g is $(\delta, \text{TH}_\lambda[\mathcal{F}^{O(\epsilon)}])$-hard, then the 2δ-dense measure μ lies in $[g \oplus \mathcal{F}]_{O(\lambda)}$.*

Again our construction is more general than previous ones as it is uniform in an arbitrary (U, σ) and does not demand that $\mathcal{F}$ be Boolean while achieving parameters which are known to be tight [LTW07].

A weak version of Theorem 2.2.3 (Weak Hardcore Lemma: HCL_{WEAK}) with $d_\sigma(\mu) = \delta$ was first proved by Impagliazzo [Imp95], with the tight version (Strong Hardcore Lemma with $d_\sigma(\mu) = 2\delta$) implicit via a bootstrapping argument (albeit with significantly weaker parameters). Klivans and Servidio [KS99] gave a proof of the strong version, again with suboptimal parameters, via two-stage boosting and, along with Kale [Kal07], made explicit the connections to computational learning. The first uniform proof (of the strong version) was given by Holenstein [Hol05], with $\lambda = O(\epsilon^{-2}\delta^{-2})$. The best known parameter $\lambda = O(\epsilon^{-2}\log(\delta^{-1}))$ was achieved by Barak et al. [BHK09], using the computational learning framework, and this λ was shown to be tight by Lu et al. [LTW07] using Turán's Theorem and bounds on tail probabilities of binomial distributions. With the exception of Trevsian et al., all previous work is in the setting $\mathcal{U} = (U, u)$ and Boolean $\mathcal{F}$.

Trevisan et al. [TTV09] also derive a constructive form of Weak Hardcore Lemma with the additional property that the hardcore measure $\mu = |g - h|$, where $h \in \mathcal{F}_\lambda$ is the $(\epsilon\delta, \mathcal{F})$ − *approximation* to g, however, they need g to be $(\delta, \text{TH}_1[\mathcal{F}_\lambda])$ − *hard* with $\lambda = O(\epsilon^{-2}\delta^{-2})$.

Remark 2.2.5. *The form of hardcore measure derived by Trevisan et al. is particularly nice as it can be shown that if the δ-dense $(\epsilon, \mathcal{F})$-hardcore measure is of form $|g - h|$ with $h \in \mathcal{F}_\lambda$, then h must be $(\epsilon, \mathcal{F})$-approximation (as against the ideal $(\epsilon\delta, \mathcal{F})$-approximation) to g, giving us a weak converse to Trevisan et al.'s reduction from* LCAT *to* HCL_{WEAK}.

As before these various formulations are simply explicit forms of the *Generic Constructive Strong Hardcore Lemma*:

> **Generic Constructive Strong Hardcore Lemma:** For every $\epsilon, \delta \in (0,1)$, there is a $\lambda = \textit{poly}(1/\epsilon, 1/\delta)$ so that the following holds for any probability space $\mathcal{U} = (U, \sigma)$, any class $\mathcal{F}$, and any Boolean function g over U: there exists classes $\mathcal{F}', \mathcal{F}''$

parametrized by $\mathcal{F}, \lambda$ such that if g is $(\delta, \mathcal{F}')$-hard w.r.t. σ, then there is a 2δ-dense measure $\mu \in \mathcal{F}''$ that is $(\epsilon, \mathcal{F})$-hardcore for g w.r.t. σ.

2.2.3 Dense Models

Theorem 2.2.6 (Dense Model Theorem, Pseudorandom formulation: $\text{DMT}_{\text{PSEUDORANDOM}}$ [TTV09, RTTV08c, Zha11, GT08, TZ08]). *Given $\epsilon, \delta \in (0,1)$, there is a $\lambda = O(\epsilon^{-2} \log \delta^{-1})$ such that for any finite probability space $\mathcal{U} = (U, \sigma)$ and any class $\mathcal{F}$ over U, the following two implications hold for a distribution ρ inside $\mathcal{U}$:*

1. *If ρ is δ-dense inside a distribution τ, where τ is $(\epsilon\delta, \text{TH}_\lambda[\mathcal{F}])$-pseudorandom in $\mathcal{U}$, then there is a δ-dense measure $\mu \in \mathcal{F}^*_{O(\lambda)}$ which is a $(O(\epsilon), \mathcal{F})$-model for ρ in $\mathcal{U}$.*
2. *If ρ is δ-dense inside a distribution τ, where τ is $(\epsilon\delta, \text{TH}_\lambda[\mathcal{F}^{O(\epsilon)}])$-pseudorandom in $\mathcal{U}$, then there is a δ-dense measure $\mu \in \mathcal{F}_{O(\lambda)}$ which is a $(O(\epsilon), \mathcal{F})$-model for ρ in $\mathcal{U}$.*

Note that the σ is free, whereas the versions of $\text{DMT}_{\text{PSEUDORANDOM}}$ due to Zhang and Trevisan et al. are in the fixed probability space (U, u). $\text{DMT}_{\text{PSEUDORANDOM}}$ was originally proved by Green, Tao, and Ziegler [GT08, TZ08] with regards to arithmetic and polynomial progressions in primes, however, the result was formulated in functional analytic language rather than complexity theoretic; Trevisan et al. [TTV09] were the first to derive the result in the complexity theoretic setting, where they show that $(\epsilon\delta, \mathcal{F})$-approximation to ρ w.r.t. τ is the required model with the parameter $\lambda = O(\epsilon^{-2}\delta^{-2})$. This λ is suboptimal, and the need for τ to be $(\epsilon\delta, \mathcal{F}_\lambda)$-pseudorandom is more than required (we demand $(\epsilon\delta, \text{TH}_\lambda[\mathcal{F}])$-pseudorandomness). Zhang [Zha11], using the online learning setup, achieves the the parameter $\lambda = O(\epsilon^{-2} \log(\delta^{-1}))$ and shows that it is tight in a certain sense with methods reminiscent of Lu et al. and their analysis of Hardcore Lemma constructions. A similar tightness analysis of query as well as advice complexity was given by Watson [Wat11].

We give a generalization of the Pseudorandom formulation of the Dense Model Theorem showing that a weaker Pseudodensity condition (originally due to Russell Impagliazzo [Imp09]) suffices to give existence and constructions of dense models:

Theorem 2.2.7 (Dense Model Theorem, Pseudodensity formulation: $\text{DMT}_{\text{PSEUDODENSITY}}$ [Imp09, RTTV08a]). *Given $\epsilon, \delta \in (0,1)$, there is a $\lambda = O(\epsilon^{-2} \log \delta^{-1})$ such that for any finite probability space $\mathcal{U} = (U, \sigma)$ and any class $\mathcal{F}$ over U, the following two implications hold for a distribution ρ inside $\mathcal{U}$:*

1. *If ρ has $(\epsilon\delta, \text{TH}_\lambda[\mathcal{F}])$-pseudodensity δ w.r.t. σ, then there is a δ-dense $(O(\epsilon), \mathcal{F})$-model $\mu \in \mathcal{F}^*_{O(\lambda)}$ for ρ in $\mathcal{U}$.*

2. *If ρ has $(\epsilon\delta, \mathrm{TH}_\lambda[\mathcal{F}^{O(\epsilon)}])$-pseudodensity δ w.r.t. σ, then there is a δ-dense $(O(\epsilon), \mathcal{F})$-model $\mu \in \mathcal{F}_{O(\lambda)}$ for ρ in $\mathcal{U}$.*

Remark 2.2.8. *Note that in case $\mathcal{F}$ is Boolean the two implications in Theorems 2.2.6, 2.2.7 are identical. So not only do we show that dense models exists but give the conditions under which they also have small complexity as considered by Trevisan et al.*

It is easy to see that $\mathrm{DMT}_{\mathrm{PSEUDORANDOM}}$ follows from $\mathrm{DMT}_{\mathrm{PSEUDODENSITY}}$ since, as shown in the following lemma, if a distribution ρ is δ-dense inside a $(\epsilon, \mathcal{F})$-pseudorandom distribution τ, then ρ also has $(\epsilon, \mathcal{F})$-pseudodensity at least δ.

Lemma 2.2.9. *For $\mathcal{U} = (U, \sigma)$, suppose that distribution ρ has density δ inside a distribution τ where τ is $(\epsilon, \mathcal{F})$-pseudorandom in $\mathcal{U}$. Then ρ has $(\epsilon, \mathcal{F})$-pseudodensity at least δ inside σ.*

Proof. By assumption, $\tau(x) \geq \delta \cdot \rho(x)$ for all $x \in U$. Hence, for every $f \in \mathcal{F}$, we have $f[\tau] \geq \delta \cdot f[\rho]$. By the definition of pseudorandomness, we have for every $f \in \mathcal{F}$ that $|f[\tau] - f[\sigma]| \leq \epsilon$. It follows that, for every $f \in \mathcal{F}$, $f[\sigma] \geq f[\tau] - \epsilon \geq \delta \cdot f[\rho] - \epsilon$, as required. □

A non-constructive form of 2.2.7 (with $\sigma = u$, $\lambda = O(\epsilon^{-2}\log(\epsilon^{-1}\delta^{-1}))$) was also given by Reingold et al in [RTTV08a, RTTV08b] and this is the approach we willl also follow.

Impagliazzo [Imp09] gave the constructive proof of Theorem 2.2.7 in the setting $\mathcal{U} = (U, u)$, with ρ restricted to the uniform distribution on sets, and the Boolean class $\mathcal{F}$. Impagliazzo's reduction is from Holenstein's Strong Hardcore Lemma and generalizes to arbitrary probability spaces (U, σ) and extension to $[0, 1]$-bounded $\mathcal{F}$ is natural. However, the parameter λ, is the suboptimal $O(\epsilon^{-2}\delta^{-2})$ in [Imp09] *even if we assume the optimal $\lambda = O(\epsilon^{-2}\log(2/\delta))$ in the Strong Hardcore Lemma*; furthermore, the measure μ in [Imp09] is an $(O(\epsilon\delta^{-1}), \mathcal{F})$-model with density $\delta - O(\epsilon)$. Still as Impagliazzo's reduction is from the Strong Hardcore Lemma it does give us one direction of the equivalence between Strong Hardcore Lemma and Dense Model Theorems that we want, with the parameters still polynomial in δ and ϵ.

Not only is the Pseudodensity formulation of the Dense Model Theorem more general, but we show that in this form Dense Model Theorem reduces to the Strong Hardcore Lemma ($\mathrm{HCL}_{\mathrm{STRONG}}$) with parameters still $poly(\epsilon^{-1}, \delta^{-1})$. This allows us to demonstrate the equivalence of Algorithmic Dense Model Theorems ($\mathrm{DMT}_{\mathrm{ALGORITHMIC}}$) and the $\mathrm{HCL}_{\mathrm{STRONG}}$ since we can use known $\mathrm{DMT}_{\mathrm{ALGORITHMIC}}$ as a blackbox to yield $\mathrm{HCL}_{\mathrm{STRONG}}$ (in fact we can use any construction of dense model satisfying a certain condition to this end). We prove and use the following formulation of the $\mathrm{HCL}_{\mathrm{STRONG}}$ which we finesse to obtain $\mathrm{DMT}_{\mathrm{ALGORITHMIC}}$.

Along the lines of Generic Low Complexity Approximation Theorem and Generic Constructive Strong Hardcore Lemma, Theorems 2.2.7 and 2.2.7 are regarded as special cases of corresponding classes of theorems:

Generic Constructive Dense Model Theorem (Pseudodensity Version): For every $\epsilon, \delta \in (0,1)$, there is a $\lambda = poly(1/\epsilon, 1/\delta)$ such that the following holds for any probability space $\mathcal{U} = (U, \sigma)$, any class $\mathcal{F}$, and any distribution ρ over U: there exists classes $\mathcal{F}', \mathcal{F}''$ parametrized by $\mathcal{F}, \lambda$ such that if ρ has $(\epsilon, \mathcal{F}')$-pseudodensity δ inside σ, then there is a δ-dense $(O(\epsilon/\delta), \mathcal{F})$-model $\mu \in \mathcal{F}''$ for ρ w.r.t σ.

Generic Constructive Dense Model Theorem (Pseudorandom Version): For every $\epsilon, \delta \in (0,1)$, there is a $\lambda = poly(1/\epsilon, 1/\delta)$ such that the following holds for any probability space $\mathcal{U} = (U, \sigma)$, any class $\mathcal{F}$, and any distributions ρ and τ over U: there exists classes $\mathcal{F}', \mathcal{F}''$ parametrized by $\mathcal{F}, \lambda$ such that if ρ is δ-dense inside τ, where τ is $(\epsilon, \mathcal{F}')$-pseudorandom in $\mathcal{U}$, then there is a δ-dense $(O(\epsilon/\delta), \mathcal{F})$-model $\mu \in \mathcal{F}''$ for ρ w.r.t. σ.

2.2.4 Learning Theory Connections

Consider Theorems 2.2.1, 2.2.3, 2.2.6, 2.2.7 in contrapositive form:

1. Theorem 2.2.1 (LCAT) says that for any bounded function there is exists an efficiently constructible function which is indistinguishable from that function i.e. every function can be "learnt" efficiently.

2. Theorem 2.2.3 ($\text{HCL}_{\text{STRONG}}$) says that if on every 2δ-dense measure we can guess a function g correctly with probability better than $\frac{1}{2} + \epsilon$ then we can efficiently come up with a function that guesses g correctly with probability better than $1 - \delta$, i.e. we can boost the probability of making a correct guess.

3. Theorem 2.2.6 ($\text{DMT}_{\text{PSEUDORANDOM}}$) says that if for every δ-dense distribution μ_σ we can come up with a function which distinguishes ρ and μ_σ (i.e. ρ has no $\delta - dense$ model), then we can learn the base distribution σ well enough that we can tell it apart from any distribution τ in which ρ is δ-dense (i.e. τ cannot be pseudorandom).

4. Theorem 2.2.7 ($\text{DMT}_{\text{PSEUDODENSITY}}$) says that assuming we can tell ρ apart from every candidate model then we can learn ρ well enough to get a (ϵ, δ)-distinctive test, that is, we can come up with a function which has a disproportionately higher correlation with ρ than with σ.

All four results are about using witnesses to failure of a certain condition to show that we can produce a witness to failure of a stronger related condition; so in a sense we are learning from one set of witnesses to produce a more powerful witness. In particular, boosting as given by 2 has been explicitly addressed in learning theory: Feldman [Fel10] showed that any hardcore

measure construction satisfying the Strong Hardcore Lemma is a distribution specific agnostic boosting algorithm. Furthermore, the framework of online learning has been utilized to address both dense model and hardcore constructions [BHK09, Zha11].

With this in mind we frame a more general version of Dense Model Theorem, $\text{DMT}_{\text{Min-Max}}$, which better captures this idea of a collection of witnesses being used to build another witness, and so can be used to give a black box proof of all of the above theorems.

Theorem 2.2.10 (Min-Max Formulation of Dense Model Theorem $\text{DMT}_{\text{Min-Max}}$)**.** *There is a universal constant $c > 0$ such that for every $\epsilon, \delta \in (0,1)$, there is a $\lambda = poly(1/\epsilon, 1/\delta)$ so that exactly one of the following two conditions holds for any probability space $\mathcal{U} = (U, \sigma)$, any class $\mathcal{F}$, and any distribution ρ over U:*

- *Either ρ has a δ-dense $(\epsilon, \mathcal{F})$-model $\mu \in \mathcal{F}'' \subseteq \mathcal{F}^*_{O(\lambda)}$ w.r.t σ, i.e. there are no witnesses that can distinguish ρ from δ-dense μ better than ϵ.*
- *Or there is a function F, the average of fewer than λ functions f from $\mathcal{F}$ such that F is an ϵ/c-distinguishes ρ and any given δ-dense measure, i.e. for every δ-dense measure γ, $F[\gamma_\sigma] - F[\rho] > \epsilon/c$.*

Remark 2.2.11. *As has been the case so far if $\mathcal{F}$ is Boolean then $\mathcal{F}'' = \mathcal{F}_{O(\lambda)}$.*

The Min-Max apropos $\text{DMT}_{\text{Min-Max}}$ alludes to fact that a result of this flavor can be obtained by setting up a two-player, zero sum game between a player who plays candidate models for ρ and another who plays over possible witnesses showing the failure of the other player's choice of model and appealing to von Neuman's Min-Max Theorem.

2.3 The Scope

We first show direct reductions amongst Theorems 2.2.1, 2.2.3, 2.2.6, 2.2.7 pointing out various artifacts of interest along the way. Next we generalize the online-learning framework of [BHK09, Zha11] to first obtain $\text{DMT}_{\text{Min-Max}}$ and then use the essence of our reductions to obtain all four of the results from $\text{DMT}_{\text{Min-Max}}$.

Chapter 3

Low Complexity Approximations

3.1 Hardcore Measures and Low Complexity Approximations

To ease into technical details we start in the framework of Trevisan et al and demonstrate a converse (with some relaxation) to their reduction from Low Complexity Approximation Theorem to Weak Hardcore Lemma.

Being able to interchangeably work with $[0,1]$-bounded functions and $[-1,1]$ bounded functions lends transparency to computations. So for any $\phi : U \to [0,1]$ consider the corresponding $\phi^\dagger$ defined as follows:

Definition 3.1.1. For $\phi : U \to [0,1]$, $\phi^\dagger = 2\phi - 1$.

Remark 3.1.2. *Note that $\phi^\dagger : U \to [-1,1]$ with $\phi^\dagger = 1$ iff $\phi = 1$ and $\phi^\dagger = -1$ iff $\phi = 0$. Also $\phi = (\phi^\dagger + 1)/2$ and $(\bar{f})^\dagger = -(f^\dagger)$.*

Definition 3.1.3. For a $[0,1]$-bounded class $\mathcal{F}$ on U, $\mathcal{F}^\dagger$ denotes the collection $\{f^\dagger : f \in \mathcal{F}\}$.

Remark 3.1.4. *Since $\mathcal{F}$ is a class $\mathbf{1} \in \mathcal{F}$ meaning $\mathbf{1} \in \mathcal{F}^\dagger$ and for every $f \in \mathcal{F}$, $\pm f^\dagger \in \mathcal{F}^\dagger$.*

As f and $f^\dagger$ are related by an affine transformation, the estimates relevant here turn out to be the same up to a constant multiple. The following lemma (proof deferred till A.2.1) gives some algebraic relations between f and $f^\dagger$ and relates estimates for functions in the two ranges:

Lemma 3.1.5. *Given $\epsilon \in (0, \frac{1}{2})$ for a $[0,1]$-bounded class $\mathcal{F}$ on (U,σ) and functions $g, h : U \to [0,1]$:*

1. *If g is $\{0,1\}$-Boolean then $2|g - f| = (1 - f^\dagger g^\dagger)$ and $2|g - \bar{f}| = (1 + f^\dagger g^\dagger)$.*
2. *If g is $\{0,1\}$-Boolean then $g^\dagger |g - h| = g - h$.*
3. *$2|\mathbb{E}_\sigma[f^\dagger(g - h)]| = |\mathbb{E}_\sigma[f^\dagger(g^\dagger - h^\dagger)]|$.*

4. *For every $f^\dagger$ in class $\mathcal{F}^\dagger$, $|\mathbb{E}_\sigma[f^\dagger(g^\dagger - h^\dagger)]| = O(\epsilon)$ iff $|\mathbb{E}_\sigma[f(g-h)]| = O(\epsilon)$ for every f in class $\mathcal{F}$.*

With this in hand, consider Trevisan et al's following result:

Theorem 3.1.6 (Trevisan et al's LCAT-HCL$_{\text{WEAK}}$-reduction)**.** *Suppose in a finite probability space (U,σ), given parameters $\epsilon,\delta > 0$ and a class of functions $\mathcal{F}$, $h : U \to [0,1]$ is an $(\epsilon\delta,\mathcal{F})$-approximation to $g : U \to \{0,1\}$ w.r.t σ. If $h \in \mathcal{F}_\lambda$ (for some $\lambda = poly(1/\epsilon, 1/\delta)$) with g $(\delta, \text{TH}_1[\mathcal{F}_\lambda])$-hard w.r.t. σ, then the measure $\mu = |g-h|$ is δ-dense, $(O(\epsilon),\mathcal{F})$-hardcore.*

Sketch of Trevisan et al's reduction. Using that $\mathbb{E}_\sigma[\mathbb{E}_{t\sim[0,1]}[|g - th_t[h]|]] = \mathbb{E}_\sigma[|g-h|]$ along with $(\delta, \text{TH}_1[\mathcal{F}_\lambda])$-hardness of g, it's easy to show $\mathbb{E}_\sigma[|g-h|] \geq \delta$. By the fact that h is an $(\epsilon\delta,\mathcal{F})$ approximation to g along with Lemma 3.1.5,

$$\mathbb{E}_\sigma[\mu|g-\bar{f}|] = \tfrac{1}{2}\mathbb{E}_\sigma[|g-h|(1+f^\dagger g^\dagger)] = \tfrac{1}{2}d_\sigma(\mu) + \tfrac{1}{2}\mathbb{E}_\sigma[(g-h)f^\dagger] \leq \tfrac{1}{2}d_\sigma(\mu) + O(\epsilon\delta)$$

Therefore, it follows that $\mathbb{E}_{\mu_\sigma}[|g-\bar{f}|] = \mathbb{E}_\sigma[|g-h||g-\bar{f}|]/d_\sigma(\mu)$ is as claimed. □

So Trevisan et al. can use the existence of $(\epsilon\delta,\mathcal{F})$-approximations of low complexity to derive their formulation of Weak Hardcore Lemma:

Theorem 3.1.7 (Trevisan et al Weak Hardcore Lemma)**.** *Given parameters $\epsilon,\delta > 0$, there exists $\lambda = poly(1/\epsilon, 1/\delta)$ such that if in a finite probability space (U,σ) with a class of functions $\mathcal{F}$ on U, a function $g : U \to \{0,1\}$ is $(\delta, \text{TH}_1[\mathcal{F}_\lambda])$-hard w.r.t. σ, then there exists $h \in \mathcal{F}_\lambda$ such that the measure $|g-h|$ is δ-dense, $(O(\epsilon,\mathcal{F}))$-hardcore.*

We will demonstrate the following which shows that assuming existence of hardcore measures of the form given by Trevisan et al is enough to show existence of low complexity approximations:

Theorem 3.1.8. *In any finite probability space (U,σ), given $\delta > 0$, $g : U \to \{0,1\}$ and a class $\mathcal{F}$, then Trevisan et al's Weak Hardcore Lemma (Theorem 3.1.7) implies that there exists $h \in \mathcal{F}_\lambda$ such that h is an $(O(\delta),\mathcal{F}) - approximation$ to g w.r.t. σ with $\lambda = poly(1/\delta)$.*

Proof. Fix $\epsilon = \delta/4$ and choose $\lambda' = poly(1/\epsilon, 1/\delta)$ as needed in Theorem 3.1.7 with parameters ϵ,δ implying $\lambda' = poly(1/\delta)$ and set $\lambda = \lambda' + 1 = poly(1/\delta)$.

There are two mutually exclusive possibilities:

1. $\exists f^* \in \mathcal{F}_\lambda$ with $\mathbb{E}_\sigma[|g-f^*|] \leq \delta$
2. $\forall f \in \mathcal{F}_\lambda,\ \mathbb{E}_\sigma[|f-g|] > \delta$

If 1 holds then $\forall f \in \mathcal{F}_\lambda$, $|\mathbb{E}_\sigma[f(g-f^*)]| \leq \mathbb{E}_\sigma[|f(g-f^*)|] \leq \mathbb{E}_\sigma[|g-f^*|]| \leq \delta$, so $f^* \in \mathcal{F}_\lambda$ is the required $(\delta, \mathcal{F})$-approximation.

So assume 2. Since $\textsc{Th}_1[\mathcal{F}_{\lambda'}] \subset \mathcal{F}_\lambda$, therefore, g is $(\delta, \textsc{Th}_1[\mathcal{F}_{\lambda'}]) - hard$. As λ' is exactly as needed by Theorem 3.1.7, there exists $h \in \mathcal{F}_{\lambda'}$ such that $\mu = |g-h|$ is δ-dense-$(\epsilon, \mathcal{F})$-hardcore.

As μ is $(\epsilon, \mathcal{F})$-hardcore,

$$\epsilon \geq |\mathbb{E}_{\mu_\sigma}[|g-f|] - \tfrac{1}{2}| = \tfrac{1}{2}|\mathbb{E}_{\mu_\sigma}[f^\dagger g^\dagger]| = \mathbb{E}_\sigma[f^\dagger g^\dagger |g-h|]/d_\sigma(\mu)$$

By Lemma 3.1.5,

$$\forall f \in \mathcal{F},\ 2\epsilon d(\mu) \geq |\mathbb{E}_\sigma[(2f-1)g^\dagger |g-h|]| \geq |\mathbb{E}_\sigma[2fg^\dagger |g-h|]| - |\mathbb{E}_\sigma[g^\dagger |g-h|]|$$

Therefore,

$$\forall f \in \mathcal{F},\ 2\epsilon d(\mu) + \mathbb{E}_\sigma[g^\dagger |g-h|]| \geq 2|\mathbb{E}_\sigma[f(g-h)]|$$

Consider $|\mathbb{E}_\sigma[g^\dagger |g-h|]| = |\mathbb{E}_\sigma[(2g-1)\mu]|$. Using Lemma 3.1.5 and that $g = |g - \mathbf{0}|$, $|\mathbb{E}_\sigma[(2g-1)\mu]| = \mathbb{E}_{\mu_\sigma}[2|g-\mathbf{0}|-1]d_\sigma(\mu)$. And since $|\mathbb{E}_{\mu_\sigma}[|g-\mathbf{0}| - \frac{1}{2}]| \leq \epsilon$ because $\mathbf{0} \in \mathcal{F}$ and μ is $(\epsilon, \mathcal{F})$-hardcore for g:

$$|\mathbb{E}_\sigma[g^\dagger |g-h|]| = 2d_\sigma(\mu)|\mathbb{E}_{\mu_\sigma}[|g-\mathbf{0}| - \tfrac{1}{2}]| \leq 2d_\sigma(\mu)\epsilon$$

This yields

$$\forall f \in \mathcal{F},\ \delta \geq 2d_\sigma(\mu)\epsilon \geq |\mathbb{E}_\sigma[f(g-h)]|$$

.

Thus $h \in \mathcal{F}_{\lambda'}$ is the required $(\delta, \mathcal{F})$-approximation.

Remark 3.1.9. *If we could guarantee $d_\sigma(\mu) = O(\delta)$, then we could have set ϵ to a small absolute constant instead; however, the best upper bound we can guarantee for density of a hardcore measure is $1-\delta$. The choice of $\epsilon = O(\delta)$ makes the complexity parameter achieved quadratically worse as compared to 2.2.1. The obvious solution to scale μ to have density exactly δ does not resolve this as then the measure has form $\mu = k|g-h|$ for some constant k and this k carries throughout.*

□

3.2 Dense Models to Low Complexity Approximations

Now we give reductions from existence of dense models to existence of low complexity approximations:

$$\mathrm{DMT}^{\text{PSEUDODENSITY}}_{\text{PSEUDORANDOM}} \to \textsc{Lcat}.$$

In particular as Trevisan et al derive Pseudorandom form of Dense Model Theorem from existence of low complexity approximations, the reductions here imply the equivalence of the two results. Also, there is a direct correspondence between the parameters we obtain i.e. there is no loss in parameters when going from assumptions of Dense Model Theorem to conclusion of Low Complexity Approximation Theorem.

Theorem 3.2.1. *Suppose $g : U \to [0,1]$ is arbitrary. Then in any probability space (U, σ), for any class $\mathcal{F}$ on U, assuming $\mathrm{DMT}_{\text{PSEUDODENSITY}}$ (Theorem 2.2.7), yields $h \in \mathcal{F}_\lambda$ which is $(O(\epsilon), \mathcal{F})$ − approximation to g w.r.t. σ, with $\lambda = O(\epsilon^{-2})$ and any $0 < \epsilon < \frac{1}{2}$.*

Proof. Without loss of generality we can assume that $\alpha = d_\sigma(g)$ is known since we can always estimate it within ϵ by taking $poly(\epsilon^{-1})$ samples, and our parameters would be worse only by a constant multiple of ϵ because of this slack. Also, we can assume $\alpha \geq \frac{1}{2}$ as otherwise we can choose to work with $\bar{g}$ since $\mathbb{E}_\sigma[f(g-h)] = -\mathbb{E}_\sigma[f(\bar{g}-\bar{h})]$.

Notice that $\mathbb{E}_\sigma[g] = \alpha$ implies, by Claim 2.1.4, that the induced distribution g_σ has density at least α inside σ, and hence also $(\epsilon, \textsc{Th}_\lambda[\mathcal{F}^{O(\epsilon)}])$-pseudodensity at least α for every $\epsilon > 0, \lambda \in \mathbb{N}$, hence also for $\lambda = \lambda(\epsilon^{-1}, \alpha^{-1})$ as needed by Theorem 2.2.7. Note that since $\alpha \geq \frac{1}{2}$, Theorem 2.2.7 fixes $\lambda = O(\epsilon^{-2})$.

By Theorem 2.2.7, g has a α-dense $(\epsilon', \mathcal{F})$-model $h \in \mathcal{F}_\lambda$, for $\epsilon' = O(\epsilon)$. Suppose $\beta = d_\sigma(h)$ is the true density of h. Note that $\beta \geq \alpha$. Using standard concentration bounds, by random sampling β can be estimated within an additive error ϵ with high probability. Let β' denote the estimate.

Now $\beta' \geq \beta - \epsilon \geq \alpha - 2\epsilon$. In case $\beta' > \alpha$, replace h by $h' = c \cdot h$ with $c = \alpha/\beta'$. By Remark 2.1.9, h' is also a model for g_σ, and $d_\sigma(h') = c \cdot d_\sigma(h) = \alpha\beta/\beta'$. By using the fact that $\beta' \in [\beta - \epsilon, \beta + \epsilon]$ with high probability, $d_\sigma(h') \geq \alpha - \epsilon$ and $d_\sigma(h') \leq \alpha + 2\epsilon$, also with high probability. If $\beta' \leq \alpha$ then keep h as the model with h satisfying $\alpha - \epsilon \leq d_\sigma(h) \leq \alpha + \epsilon$.

In either case, we get a measure in $\mathcal{F}_{\lambda+1}$ which is a model for distribution g_σ with the density of the measure within an additive 2ϵ from the density α of g. For simplicity of notation, we will denote this model by h.

By definition of being a model, we have for every $f \in \mathcal{F}$, $|f[h_\sigma] - f[g_\sigma]| \leq \epsilon'$, which is equivalent to

$$\left| \frac{\mathbb{E}_\sigma[f \cdot h]}{d_\sigma(h)} - \frac{\mathbb{E}_\sigma[f \cdot g]}{d_\sigma(g)} \right| \leq \epsilon'.$$

The following simple estimate (proof deferred) which formalizes the idea that if two fractions are close and their denominators are also close, then their numerators are also close allows us to complete the proof.

Lemma 3.2.2. *For any a, b, x, y, if $|x-y| \leq \epsilon_1$ and $|a/x - b/y| \leq \epsilon_2$, then $|a-b| \leq (b/y)\epsilon_1 + x\epsilon_2$.*

Applying Lemma 3.2.2 with $a = \mathbb{E}_\sigma[f \cdot h]$, $b = \mathbb{E}_\sigma[f \cdot g]$, $x = d_\sigma(h)$, $y = d_\sigma(g)$, $\epsilon_1 = 2\epsilon$, and $\epsilon_2 = \epsilon'$, observe that $b \leq y$ and $x \leq 1$ (since f and h are at most 1 on any input in U). Thus, we get that $|a - b| \leq \epsilon_1 + \epsilon_2$, implying that $|\mathbb{E}_\sigma[f \cdot (h - g)]| \leq 2\epsilon + \epsilon'$. So h is an $(O(\epsilon), \mathcal{F})$-approximation to g, and the complexity of h relative to $\mathcal{F}$ is $O(\epsilon^{-2})$ as required. □

Along similar lines, we derive LCAT from $\text{DMT}_{\text{PSEUDORANDOM}}$.

Theorem 3.2.3. *Suppose $g : U \to [0, 1]$ is arbitrary. Then for the probability space $\mathcal{U} = (U, \sigma)$ and any class $\mathcal{F}$ on U, assuming $\text{DMT}_{\text{PSEUDORANDOM}}$ (Theorem 2.2.6), yields an $h \in \mathcal{F}_\lambda$ such that h is a $(\epsilon, \mathcal{F})$-approximation to g in $\mathcal{U}$, for $\lambda = O(\epsilon^{-2})$ and any $0 < \epsilon < \frac{1}{2}$.*

Proof. We simply use Theorem 2.2.6, with $\rho = g_\sigma$ and $\tau = \sigma$. For $\alpha = d_\sigma(g)$, we get by Claim 2.1.4, that the density $d_u(g_u) \geq \alpha$. Trivially, for every class $\mathcal{G}$, σ is $(O(\epsilon), \mathcal{G})$-pseudorandom in the space (U, σ). Hence, by Theorem 2.2.6, there exists δ-dense $(O(\epsilon), \mathcal{F})$-model $h \in \mathcal{F}_\lambda$ for g. The rest is exactly as in proof for Theorem 3.2.1. □

3.3 Low Complexity Rational Approximations

In section 3.1 we gave a reduction (albeit with a relaxation in parameters) from Trevisan et al. Weak Hardcore Lemma (3.1.7) to LCAT (Theorem 2.2.1). Then we showed that Trevisan et al. original reduction is strictly weaker than the Strong Hardcore Lemma. We note that LCAT as formulated by Trevisan et al. intuitively says that if $\mathbb{E}_\sigma[f(g - h)]$ is small for every $f \in \mathcal{F}$ then h "looks" like g to $\mathcal{F}$. Here h is pretty much a polynomial made from functions from $\mathcal{F}$ (with the exception of truncation and threshold operations, though these can be approximated well by polynomials of bounded degree (see Reingold et al.[RTTV08b])); however, an interesting generalization is to consider a rational function approximation to g. Since division is not a permissible operation, we think of g as being approximated by a rational function u_1/u_2 with both $u_1, u_2 \in \mathcal{F}_\lambda$ if for every $f \in \mathcal{F}$, $\mathbb{E}_\sigma[f(u_2 g - u_1)]$ is small. This "rational function" approximation to g is sufficient to get at Holenstein's Strong Hardcore Lemma (2.2.3) which was beyond the reach of LCAT with Trevisan et al's reduction. To motivate this Low Complexity Rational Approximation Theorem, we first consider an equivalent formulation of LCAT inspired by the techniques of section 3.1, before formulating and proving the Low Complexity Rational Approximation Theorem.

3.3.1 Low Complexity Approximation Theorem: Weak Alternative Form

Assuming the Trevisan et al. Weak Hardcore Lemma (3.1.7), Low Complexity Approximation Theorem 2.2.1 is equivalent to Proposition 3.3.1 (though the parameter λ is quadratically worse again as remarked in 3.1.9).

Proposition 3.3.1. *In any probability space (U, σ), given $\epsilon, \epsilon' > 0$, $\{0,1\}-$boolean class $\mathcal{F}$ on U, a $\{0,1\}-$boolean g, there exists $\lambda = poly(1/\epsilon, 1/\epsilon')$ such that one of the two conditions must hold:*

(a) $\exists u^* \in \mathcal{F}_\lambda$ satisfying $\mathbb{E}_\sigma[|g - u^*|] \leq \epsilon$.

(b) $\exists u^* \in \mathcal{F}_\lambda$ satisfying $\forall f,\ |\mathbb{E}_\sigma[f(g-u^*)]| \leq \mathbb{E}_\sigma[|g-u^*|]\epsilon'$ with $\mathbb{E}_\sigma[|g-u^*|] \in [\epsilon, 1]$.

Theorem 3.3.2. *Assuming Trevisan et al's Weak Hardcore Lemma, Theorem 2.2.1 is equivalent to Proposition 3.3.1*

Proof. To see 3.3.1 implies 2.2.1 fix $\epsilon' = \epsilon$ and choose λ' as needed by Trevisan et al Weak Hardcore Lemma with hardness parameter ϵ and hardcore parameter ϵ' and set $\lambda = \lambda' + 1$. If (a) holds then $\forall f \in \mathcal{F},\ \epsilon \geq \mathbb{E}_\sigma[|g-u^*|] \geq \mathbb{E}_\sigma[|f(g-u^*)|] \geq |\mathbb{E}_\sigma[f(g-u^*)]|$, thus, u^* is the approximation as needed by Theorem 2.2.1.

And if (b) holds then g is $(\epsilon, \mathcal{F}_\lambda)$-hard and, therefore, $(\epsilon, \mathrm{Th}_1[\mathcal{F}_{\lambda'}])$-hard and by choice of λ' and Theorem 3.1.8 there exists a ϵ-dense $(\epsilon', \mathcal{F})$ hardcore measure $\mu = |g - u^*|$ with $u^* \in \mathcal{F}_\lambda$. And as in proof of Theorem 3.1.8 this implies u^* is an $(O(\epsilon'), \mathcal{F})$-approximation to g.

The other direction (Theorem 2.2.1 implies Proposition 3.3.1) is very clean: simply consider the two mutually exclusive possibilities that either g is well approximated by a low complexity function which gives us (a), or g is $(\epsilon, \mathcal{F}_\lambda)$-hard which by 3.1.8 (note we already have Theorem 2.2.1 so we have 3.1.8) gives (b) where $\lambda - 1$ is chosen to ensure the existence of ϵ'-hardcore, ϵ-dense measure of form $|g - u^*|$ (all needed properties follow exactly as in section 3.1). □

3.3.2 Low Complexity Rational Approximation Theorem

Since Theorem 3.3.2 relies almost entirely on section 3.1, it serves mostly a didactic purpose. It illustrates the *either a function has small complexity w.r.t. a class, or it can be "approximated" by a function which has small complexity w.r.t. the class* structure that we exploit next:

Theorem 3.3.3 (Low Complexity Rational Approximation Theorem)**.** *In any probability space (U, σ), given $\epsilon > 0, \epsilon' \in (0, 1/5]$, $[0,1]$-bounded class $\mathcal{F}$, a $\{0,1\}-$boolean g, there exists $\lambda = poly(1/\epsilon, 1/\epsilon')$ such that one of the two conditions must hold:*

1. *$\exists u \in \mathrm{Th}_\lambda[\mathcal{F}]$ satisfying $\mathbb{E}_\sigma[|g-u|] \leq \epsilon$ (equivalently $\mathbb{E}_\sigma[|g^\dagger - u^\dagger|] \leq 2\epsilon$).*

2. $\exists u_1^*, u_2^* \in \mathcal{F}_{O(\lambda)}$ *satisfying* $\forall f \in \mathcal{F},\ |\mathbb{E}_\sigma[f^\dagger(u_1^* g^\dagger + u_2^*)]| \leq 2\epsilon\epsilon'$ *with* $\mathbb{E}_\sigma[u_1^* + u_2^* g^\dagger] = 2\epsilon$ *and* $(u_1^* + u_2^* g^\dagger)$ *is* $[0,1]$*-bounded.*

Remark 3.3.4. *3.3.3.1 implies* $\forall f \in \mathcal{F},\ |\mathbb{E}_\sigma[f^\dagger(g^\dagger + u^\dagger)]| \leq 2\epsilon\epsilon'$ *with* $u \in \textsc{Th}_\lambda[\mathcal{F}]$ *(so* $u^\dagger \in \mathcal{F}_{O(\lambda)}$*). As such 3.3.3.1 is a special case of 3.3.3.2 where* u_1^* *can be fixed to be* 1 *and* ϵ' *is a small constant. Note that if* u_1^* *can be fixed to* 1 *then 3.3.3.2 is simply 2.2.1 (even though 3.3.3.2 is in range* $[-1,1]$*, Lemma 3.1.5 renders this inconsequential). However, note that the result only holds for Boolean* g*.*

Theorem 3.3.5. *Strong Hardcore Lemma (2.2.3) is equivalent to Theorem 3.3.3*

Proof. To see 2.2.3 implies 3.3.3, given $\mathcal{F}$, ϵ, g, choose λ as needed by the Strong Hardcore Lemma for $(\epsilon', \mathcal{F})$-hardcore measure to exists if g is $(\epsilon, \textsc{Th}_\lambda[\mathcal{F}])$-hard. Once λ is fixed there are two possibilities: either g is $(\epsilon, \textsc{Th}_\lambda[\mathcal{F}])$-hard or it isn't. If g is not $(\epsilon, \textsc{Th}_\lambda[\mathcal{F}])$-hard then there exists $u \in \textsc{Th}_\lambda[\mathcal{F}]$ such that $\mathbb{E}[|g - u|] \leq \epsilon$ and this is 3.3.3.1.

Otherwise g is $(\epsilon, \textsc{Th}_\lambda[\mathcal{F}])$-hard, so by Strong Hardcore Lemma there exists a 2ϵ-dense $(\epsilon', \mathcal{F})$-hardcore measure μ (guaranteed by choice of λ and Strong Hardcore Lemma). This gives

$$\forall f^\dagger \in \mathcal{F}^\dagger,\ |\mathbb{E}_\sigma[\mu f^\dagger g^\dagger]| \leq \epsilon' d(\mu).$$

Now $\mu \in [g \oplus \mathcal{F}]_\lambda^*$, so by Shannon expansion $\mu = g v_2 + (1-g) v_1$ with $v_1, v_2 \in \mathcal{F}_\lambda^*$ yielding

$$\mu = v_2 \left(\frac{g^\dagger + 1}{2}\right) + v_1 \left(1 - \frac{g^\dagger + 1}{2}\right) = \frac{(v_2 - v_1)}{2} g^\dagger + \frac{(v_1 + v_2)}{2} \equiv u_2^* g^\dagger + u_1^*.$$

Note $u_1^*, u_2^* \in \mathcal{F}_{3\lambda}^*$ and that μ being a measure is $[0,1]$-bounded. Without loss of generality μ can be scaled (by multiplying u_1^*, u_2^* by a constant) so that $d_\sigma[\mu] = \mathbb{E}_\sigma[u_1^* + u_2^* g^\dagger] = 2\epsilon$.

Remark 3.3.6. *Since we have ability to make the measure the exact density we want, therefore, unlike the results Theorem 3.3.2 and results from section 3.1 there's no loss in parameter* ϵ *since* ϵ' *can be fixed to be a small constant (say* $1/10$*) as remarked in 3.1.9.*

Because μ is $(\epsilon', \mathcal{F})$-hardcore with $d_\sigma(\mu) = 2\epsilon$, so it follows that for every $f^\dagger \in \mathcal{F}$, $\mathbb{E}_\sigma[f^\dagger g^\dagger(u_1^* + u_2^* \hat{g})] = \mathbb{E}_\sigma[f^\dagger(u_1^* g^\dagger + u_2^*)] \leq 2\epsilon'\epsilon$ and this is 3.3.3.2.

For the other direction, 3.3.3 implies 2.2.3, we want to show that there is a good λ such that g being $(\epsilon, \textsc{Th}_\lambda[\mathcal{F}])$-hard implies there exists a 2ϵ-dense $(\epsilon', \mathcal{F})$-hardcore measure. Choose this λ to be as needed by 3.3.3.2. So if we have the requisite hardness then 3.3.3.2 comes into play and it follows from properties of $\mu = u_1^* + u_2^* \hat{g}$ that μ is 2ϵ-dense $(\epsilon', \mathcal{F})$-hardcore measure. □

Chapter 4

Hardcore Measures and Dense Models

4.1 Dense Models and Weak Hardcore Lemma

Recall from Trevisan et al. [TTV09] that if $g : U \to \{0,1\}$ is $(\delta, \text{TH}_1[\mathcal{F}_\lambda])$-hard with $\lambda = O(\epsilon^{-2}\delta^{-2})$ then h (which is an $(\epsilon\delta, \mathcal{F})$-approximation to g) given by LCAT has the property that the measure $\mu = |g - h|$ is δ-dense $(\epsilon, \mathcal{F})$-hardcore. Putting this together with Theorems 3.2.1 and 3.2.3 allows us to get HCL_{WEAK} from various formulations of DMT.

We note that the reduction of [TTV09] has two weak points (besides the issue with density being δ and not 2δ):

1. To get $(O(\epsilon\delta), \mathcal{F})$-approximation to g, one needs $\lambda = O(\epsilon^{-2}\delta^{-2})$ instead of the tight $O(\epsilon^{-2}\log(\delta^{-1}))$.

2. g must be $(\delta, \text{TH}_1[\mathcal{F}_\lambda])$-hard versus $(\delta, \text{TH}_\lambda[\mathcal{F}])$-hard as in Theorem 2.2.3 .

On the other hand the reduction of Trevisan et al. [TTV09] raises the question: does every $(\epsilon, \mathcal{F})$-hardcore measure μ looks like $|g - h|$ for some $h \in \mathcal{F}_\lambda$? We show that the answer is no in case $\text{TH}_1[\mathcal{F}_\lambda]$ is closed under complementation. Since we need the closure under complementation to enforce the very reasonable condition that g and $\bar{g}$ are *both* $(\delta, \text{TH}_1[\mathcal{F}_\lambda])$-hard, the answer is likely no in general. This is captured by the following lemma:

Lemma 4.1.1. *Suppose $g : U \to \{0,1\}$ and $\bar{g}$ are both $(\delta, \mathcal{F}_{\lambda+1})$-hard for some Boolean class $\mathcal{F}$ over the probability space (U, σ), and some parameter $\lambda \geq 1$. Then for any $h \in \mathcal{F}_\lambda$, we have $\delta \leq d_\sigma(|g - h|) \leq 1 - \delta$.*

Proof. It's easy to show that $\mathbb{E}_{\theta\in[0,1]}[\mathbb{E}_\sigma[|g - th_\theta[h]|]] = \mathbb{E}_\sigma[|g - h|] = d_\sigma(|g - h|)$, where θ is chosen uniformly at random from the interval $[0, 1]$ (see Lemma A.2.2.4).

Let $h' = th_\theta[h]|$. Since both g and h' are Boolean functions, we have $\mathbb{E}_\sigma[|g - h'|] = \mathbb{P}_{x\in\sigma}[g(x) \neq h'(x)] = 1 - \mathbb{E}_\sigma[|\bar{g} - h'|] = 1 - \mathbb{P}_{x\in\sigma}[\bar{g}(x) \neq h'(x)]$ and by hardness assumption for $g, \bar{g}$, $1 - \delta \geq \mathbb{E}_\sigma[|g - h'|] \geq \delta$. As this holds for all $\theta \in [0, 1]$, therefore, it holds on averaging over θ uniformly chosen from $[0, 1]$. □

By Lemma 4.1.1, for a function g that is δ-hard for $\delta > 1/3$, a hardcore measure $\mu = |g - h|$ for some low complexity h can only be of density at most $2/3$. On the other hand, by the Strong Hardcore Lemma we know that g must have a hardcore measure of density $2\delta > 2/3$. So, in general, the form $|g-h|$, with a low-complexity h, is insufficient for describing a arbitrary hardcore measure for a function g.

An explicit example of this insufficiency is as given below:

Example 4.1.2. *Suppose $\mathcal{U} = (U, u)$ with $U = \{x_i : i \in \mathbb{N}, i \leq 10\}$ (i.e. $|U| = 10$) and $\mathcal{F} = \{\mathbf{0}, \mathbf{1}\}$. Note that $\mathcal{F}_\lambda = \mathcal{F}$ for every $\lambda \in \mathbb{N}$. Choose $\mathcal{H} = \{x_i : i \in \mathbb{N}, i \leq 8\}$. In $\mathcal{H}$ set $g(x_i) = 1$ if i is odd, 0 otherwise. Outside of $\mathcal{H}$ set $g = 0$ identically. For any $\epsilon > 0$, $\mathcal{H}$ is $(\epsilon, \mathcal{F})$-hardcore for g, and as $\textsc{Th}_1[\mathcal{F}_\lambda] \subset \mathcal{F}_{\lambda+1} = \mathcal{F}$ for every λ, therefore, g is $(2/5, \textsc{Th}_1[\mathcal{F}_\lambda])$-hard, and by Strong Hardcore Lemma has a $4/5$-dense, $(\epsilon, \mathcal{F})$-hardcore measure which is obviously $\mathcal{H}$. But $\textsc{max}_{h\in\mathcal{F}_\lambda}[d_u(|g-h|)] = \textsc{max}[d_u(|g-\mathbf{1}|), d_u(|g-\mathbf{0}|)] = 3/5$. So there's no hardcore measure of tight density of form $|g - h|$ with h of low complexity for small ϵ.*

4.1.1 Hardcore measures and Pseudodensity

A natural approach to get $\text{HCL}_{\text{STRONG}}$ from HCL_{WEAK} is to argue that any hardcore measure μ is indistinguishable from a hardcore measure of optimal density, by showing first that μ has large pseudodensity against an appropriate class and then appealing to the Dense Model Theorem. And this with some work can be made to go the distance as even a small hardcore measure for a Boolean function g (relative to the class $\mathcal{F}$ of tests) induces a distribution that has optimal pseudodensity for the collection $g \oplus \mathcal{F}$ which characterizes the fraction of the domain where g does not agree with a given function.

Lemma 4.1.3. *For $\mathcal{U} = (U, \sigma)$ and a Boolean class $\mathcal{F}$ on U, if a measure μ is $(\epsilon, \mathcal{F})$-hardcore for a Boolean function g which is $(\delta, \mathcal{F}')$-hard with $\mathcal{F} \subset \mathcal{F}'$, then, within $\mathcal{U}$, the distribution μ_σ has $(O(\epsilon\delta), \mathcal{CH}[g \oplus \mathcal{F}])$-pseudodensity 2δ.*

Proof. Consider any $f \in \mathcal{F}$, and let $\phi = f \oplus g$. By $(\delta, \mathcal{F}')$-hardness of g, we get $\phi[\sigma] = \mathbb{E}_\sigma[g \oplus f] \geq \delta$. Since μ is $(\epsilon, \mathcal{F})$-hardcore, we get $\phi[\mu_\sigma] = \mathbb{E}_{\mu_\sigma}[g \oplus f] \in [\frac{1}{2} - \epsilon, \frac{1}{2} + \epsilon]$. It follows that $2\delta\phi[\mu_\sigma] - 2\delta\epsilon \leq \delta \leq \phi[\sigma]$. Since f was arbitrary and using linearity of expectation, it follows that μ_σ is 2δ-$(O(\epsilon\delta), \mathcal{CH}[g \oplus \mathcal{F}])$-pseudodense. □

However, the result does not directly follow since to make direct use of $\text{DMT}_{\text{PSEUODENSITY}}$ to get at $\text{HCL}_{\text{STRONG}}$, the induced distribution μ_σ must have pseudodensity 2δ against the class, $\text{TH}_\lambda[g \oplus \mathcal{F}] \cup \{\mathbf{0}, \mathbf{1}\}$ (and not just $g \oplus \mathcal{F}$) and this class is not quite easy to handle because to be able to use the hardness condition we need to factor g out of the threshold: suppose $th_\theta[\sum_{[\lambda']} |g - f_i|/\lambda'] \in \text{TH}_\lambda[g \oplus \mathcal{F}]$ then factoring out g yields

$$th_\theta\left[\sum_{[\lambda']} \frac{|g - f_i|}{\lambda'}\right] = gth_\theta\left[\sum_{[\lambda']} \frac{\bar{f_i}}{\lambda'}\right] + \bar{g}th_\theta\left[\sum_{[\lambda']} \frac{f_i}{\lambda'}\right] \equiv gth_\theta[\bar{\phi}] + \bar{g}th_\theta[\phi] \neq |g - th_\theta[\phi]|$$

with $th_\theta[\phi] \in \text{TH}_\lambda[\mathcal{F}]$, so it is not clear how to bound the expectation of $th_\theta[\sum_{[\lambda']} |g - f_i|]$ over σ to establish the needed lower bound on pseudodensity against a rich enough class to get a 2δ-dense model.

Still by same argument as 4.1.3 it's possible to argue that any $(\epsilon, \text{TH}_\lambda[\mathcal{F}])$-hardcore measure for g (with g $(\delta, \text{TH}_\lambda[\mathcal{F}])$-hard) is 2δ-$(O(\epsilon\delta), g \oplus \text{TH}_\lambda[\mathcal{F}] \cup \{\mathbf{0}, \mathbf{1}\})$-pseudodense. Using Holenstein's Derandomization (see section 6.3.3.2) it turns out that this modified pseudodensity condition is sufficient to get the 2δ-dense $(O(\epsilon), g \oplus \mathcal{F})$-model for any hardcore measure and this model will be the $(O(\epsilon), \mathcal{F})$-hardcore measure as needed by the Strong Hardcore Lemma. But since for $\text{HCL}_{\text{STRONG}}$ the assumption is that g is $(\delta, \text{TH}_\lambda[\mathcal{F}])$-hard, so even small $(\epsilon, \text{TH}_\lambda[\mathcal{F}])$-hardcore measures may not exist. To side step this such a source hardcore measure will be artificially seeded in the domain.

With this we will be able to produce 2δ-dense hardcore measures by using an algorithmic $\text{DMT}_{\text{MIN-MAX}}$ (inspired by Zhang [Zha11]) as a black box.

4.2 Strong Hardcore Lemma implies Dense Model Theorem

Since the machinery to get at $\text{HCL}_{\text{STRONG}}$ via $\text{DMT}_{\text{MIN-MAX}}$ is a little heavy, we first give the generalization of Impagliazzo's reduction [Imp09] from $\text{HCL}_{\text{STRONG}}$ to $\text{DMT}_{\text{PSEUODENSITY}}$.

4.2.1 The key idea

To motivate the reduction we consider Impagliazzo's original set up of sets under uniform distributions. Suppose a set S has $(\epsilon, \text{TH}_\lambda[\mathcal{F}])$-pseudodensity δ (in (U, u)) meaning $f[U] \geq \delta f[S] - \epsilon$ for every $f \in \text{TH}_\lambda[\mathcal{F}]$ but the actual density of S is $\delta/2$, i.e. S looks twice as big as it is. In the extreme case that the indicator function of S was in $\text{TH}_\lambda[\mathcal{F}]$ it's easy to see the psedudensity condition fails badly as then for $f = \mathbf{1}_S$, $f[U] = \delta/2$ while $\delta f[S] = \delta$. This says is that no function in $\text{TH}_\lambda[\mathcal{F}]$ can have a disproportionately higher expectation on S than U, so the indicator function of S must be somewhat hard to guess. We'd want to utilize the hardness

of $\mathbf{1}_S$ to build a hardcore measure and then argue that this measure is both a model for S and sufficiently dense.

However, if the discrepancy between the actual density and the pseudodensity of the set S is too large then the it may be that the hardness is insufficient to get a δ-dense model for S. To this end Impagliazzo warps the original uniform distribution into a new distribution in which S is actually δ-dense. With respect to this distribution $\mathbf{1}_S$ is sufficiently hard so as to yield a hardcore measure which when considered in the original distribution is the needed model. This transference from hardcore measure to model is not obvious. Impagliazzo [Imp09] achieves this via a sequence of inequalities; here we show that the similar inequalities actually hold more generally than Impagliazzo's original setting.

4.2.2 Details

Assuming $\text{HCL}_{\text{STRONG}}$ (Theorem 2.2.3) we'll prove $\text{DMT}_{\text{PSEUDODENSITY}}$ (Theorem 2.2.7.1) in the form given below. On taking Remark 2.2.4 into account Theorem 2.2.7.2 follows from Theorem 2.2.3 as well.

Theorem 4.2.1. *Given $\epsilon, \delta \in (0,1)$ such that $\epsilon \leq \delta/3$, there exists a $\lambda = O(\epsilon^{-2} \log \delta^{-1})$ such that the following holds. Let $\mathcal{U} = (U, \sigma)$ be any finite probability space, and let $\mathcal{F}$ be any class of functions over U. If a distribution ρ over U has $(\epsilon, \text{TH}_\lambda[\mathcal{F}])$-pseudodensity δ in σ, then there exists a $(12 \cdot \epsilon/\delta, \mathcal{F})$-model $\mu \in \mathcal{F}_\lambda^*$ for ρ of density $\delta - 3\epsilon$.*[1]

Proof. Set $\bar{\delta} := \delta/(1+\delta)$ (implying $\delta = \bar{\delta}/(1-\bar{\delta})$) and $\bar{\epsilon} := \epsilon/6$. Define $\hat{U} := \{(0,x) \mid x \in U\} \cup \{(1,x) \mid x \in \text{SUPPORT}[\rho]\}$, along with the following distribution $\hat{\sigma}$ over $\hat{U}$: for $(b,x) \in \hat{U}$, where $b \in \{0,1\}$ and $x \in U$, define

$$\hat{\sigma}(b,x) = \begin{cases} (1-\bar{\delta}) \cdot \sigma(x) & \text{if } b = 0 \\ \bar{\delta} \cdot \rho(x) & \text{if } b = 1. \end{cases}$$

Associate with each $f \in \mathcal{F}$ a function $\hat{f} : \hat{U} \to \{0,1\}$ such that, for any $(b,x) \in \hat{U}$, $\hat{f}(b,x) := f(x)$. Define $\hat{\mathcal{F}} = \{\hat{f} \mid f \in \mathcal{F}\}$. Note that $\hat{\mathcal{F}}$ is a class of Boolean functions over $\hat{U}$.

Consider $g : \hat{U} \to \{0,1\}$ where $g(b,x) = b$ for every $(b,x) \in \hat{U}$. Our tests $\hat{f} \in \hat{\mathcal{F}}$, on input (b,x), ignore b and use x only. Such tests have difficulty in computing g, as we show next.

Claim 4.2.2. *For $\hat{\delta} := \bar{\delta} - \epsilon + \epsilon\bar{\delta}$, the function g is $(\hat{\delta}, \text{TH}_\lambda[\hat{\mathcal{F}}])$-hard in $(\hat{U}, \hat{\sigma})$.*

Proof. Suppose there is $\hat{\phi} \in \text{TH}_\lambda[\hat{\mathcal{F}}]$, corresponding to $\phi \in \text{TH}_\lambda[\mathcal{F}]$, such that $\mathbb{E}_{\hat{\sigma}}[|g - \hat{\phi}|] < \hat{\delta}$. By considering separately inputs $\{1\} \times U$ and $\{0\} \times U$, we have $\mathbb{E}_{\hat{\sigma}}[|g - \hat{\phi}|] = \bar{\delta} \cdot \mathbb{E}_\rho[1 - \phi] + (1 - \bar{\delta}) \cdot \mathbb{E}_\sigma[\phi]$,

[1] The 3ϵ slack in the density can be moved into the error term by averaging the measure μ with the constant 1 measure as in claim 6.3.14.

and so $\bar{\delta} \cdot (1 - \phi[\rho]) + (1 - \bar{\delta}) \cdot \phi[\sigma] < \bar{\delta} - \epsilon(1 - \bar{\delta})$. Dividing both sides of this inequality by $(1 - \bar{\delta})$ and using $\delta = \bar{\delta}/(1 - \bar{\delta})$, we get $\phi[\sigma] < \delta \cdot \phi[\rho] - \epsilon$, contradicting the pseudodensity δ of ρ. □

By the Strong Hardcore Lemma, for $\lambda = O(\bar{\epsilon}^{-2} \log(\hat{\delta}^{-1})) = O(\epsilon^{-2} \log \delta^{-1})$, there exists an $(\bar{\epsilon}, \hat{\mathcal{F}})$-hardcore measure $\eta \in [|g - \hat{\mathcal{F}}|]^*_\lambda$ of density at least $2\hat{\delta}$ over $(\hat{U}, \hat{\sigma})$. Define $\eta_1(x) := \eta(1, x)$ and $\eta_0(x) := \eta(0, x) \in \mathcal{F}^*_\lambda$. We get

$$d_{\hat{\sigma}}(\eta) = \bar{\delta} \cdot d_\rho(\eta_1) + (1 - \bar{\delta}) \cdot d_\sigma(\eta_0) \geq 2(\bar{\delta} - \epsilon + \epsilon\bar{\delta}). \tag{4.1}$$

We will argue that η_0 is a dense model for ρ. First we lower-bound $d_\sigma(\eta_0)$ and $d_\rho(\eta_1)$.

Claim 4.2.3. *$d_\sigma(\eta_0) \geq \delta - (7/3)\epsilon$, and $d_\rho(\eta_1) \geq 1 - (7/3)\epsilon/\delta$.*

Proof. Since $\mathbf{0}, \mathbf{1} \in \hat{\mathcal{F}}$, by the definition of hardcore we get that both $\mathbb{P}_{\eta_{\hat{\sigma}}}[g = 1]$ and $\mathbb{P}_{\eta_{\hat{\sigma}}}[g = 0]$ are in the interval $[\frac{1}{2} - \bar{\epsilon}, \frac{1}{2} + \bar{\epsilon}]$, and so are within $2\bar{\epsilon}$ of each other. We have $\mathbb{P}_{\eta_{\hat{\sigma}}}[g = 1] = \sum_{x \in U} \eta_{\hat{\sigma}}(1, x) = (1/d_{\hat{\sigma}}(\eta)) \cdot \sum_{x \in U} \eta_1(x)\hat{\sigma}(1, x) = (1/d_{\hat{\sigma}}(\eta)) \cdot \bar{\delta} \cdot d_\rho(\eta_1)$, and similarly, $\mathbb{P}_{\eta_{\hat{\sigma}}}[g = 0] = \sum_{x \in U} \eta_{\hat{\sigma}}(0, x) = (1/d_{\hat{\sigma}}(\eta)) \cdot \sum_{x \in U} \eta_0(x)\hat{\sigma}(0, x) = (1/d_{\hat{\sigma}}(\eta)) \cdot (1 - \bar{\delta}) \cdot d_\sigma(\eta_0)$. It follows that $|\bar{\delta} \cdot d_\rho(\eta_1) - (1 - \bar{\delta}) \cdot d_\sigma(\eta_0)| \leq 2\bar{\epsilon} \cdot d_{\hat{\sigma}}(\eta) \leq 2\bar{\epsilon}$. Together with Eq. (4.1), this implies $d_\sigma(\eta_0) \geq (\hat{\delta} - \bar{\epsilon})/(1 - \bar{\delta}) = (\bar{\delta} - \epsilon + \epsilon\bar{\delta} - \bar{\epsilon}) \cdot (1 + \delta) \geq \delta - 2(\epsilon + \bar{\epsilon}) = \delta - 7\epsilon/3$, and $d_\rho(\eta_1) \geq (\hat{\delta} - \bar{\epsilon})/\bar{\delta} = (\hat{\delta} - \bar{\epsilon})(1 + \delta)/\delta \geq (\delta - (7/3)\epsilon)/\delta = 1 - (7/3)\epsilon/\delta$. □

Next we show that η_0 is a model for ρ in (U, σ) by first arguing that $(\eta_0)_\sigma$ is indistinguishable from $(\eta_1)_\rho$ and $(\eta_1)_\rho$ is indistinguishable from ρ by tests in $\mathcal{F}$, then applying the triangle inequality will then conclude the proof of the theorem.

Claim 4.2.4. *The distributions $(\eta_0)_\sigma$ and $(\eta_1)_\rho$ are $(\epsilon, \mathcal{F})$-indistinguishable.*

Proof. Let $f \in \mathcal{F}$ be arbitrary. For the corresponding test $\hat{f} \in \hat{\mathcal{F}}$, we get by the definition of hardcore that $\mathbb{E}_{\eta_{\hat{\sigma}}}[|\hat{f} - g|] \in [\frac{1}{2} - \bar{\epsilon}, \frac{1}{2} + \bar{\epsilon}]$. Conditioning on $g = 0$ and $g = 1$, we get

$$\mathbb{E}_{\eta_{\hat{\sigma}}}[|\hat{f} - g|] = \mathbb{E}_{\eta_{\hat{\sigma}}}[f \mid g = 0] \cdot \mathbb{P}_{\eta_{\hat{\sigma}}}[g = 0] + \mathbb{E}_{\eta_{\hat{\sigma}}}[1 - f \mid g = 1] \cdot \mathbb{P}_{\eta_{\hat{\sigma}}}[g = 1]. \tag{4.2}$$

We have

$$\mathbb{E}_{\eta_{\hat{\sigma}}}[f \mid g = 0] = \frac{\sum_{x \in U} f(x)\eta_0(x)\sigma(x)(1 - \bar{\delta})}{\sum_{x \in U} \eta_0(x)\sigma(x)(1 - \bar{\delta})} = \mathbb{E}_{(\eta_0)_\sigma}[f] = f[(\eta_0)_\sigma], \tag{4.3}$$

and, similarly,

$$\mathbb{E}_{\eta_{\hat{\sigma}}}[1 - f \mid g = 1] = \frac{\sum_{x \in U}(1 - f(x))\eta_1(x)\rho(x)\bar{\delta}}{\sum_{x \in U} \eta_1(x)\rho(x)\bar{\delta}} = \mathbb{E}_{(\eta_1)_\rho}[1 - f] = 1 - f[(\eta_1)_\rho]. \tag{4.4}$$

Also, since $\mathbf{0}, \mathbf{1} \in \hat{\mathcal{F}}$, we get by the definition of hardcore that both $\mathbb{P}_{\eta_{\hat{\sigma}}}[g = 0]$ and $\mathbb{P}_{\eta_{\hat{\sigma}}}[g = 1]$ are in the interval $[\frac{1}{2} - \bar{\epsilon}, \frac{1}{2} + \bar{\epsilon}]$. Combining this with Eqs. (4.2)–(4.4) yields

$$1 - \epsilon \leq \frac{1 - 2\bar{\epsilon}}{1 + 2\bar{\epsilon}} \leq 1 - f[(\eta_1)_\sigma] + f[(\eta_0)_\rho] \leq \frac{1 + 2\bar{\epsilon}}{1 - 2\bar{\epsilon}} \leq 1 + \epsilon,$$

where we used that $\bar{\epsilon} = \epsilon/6 \leq 1/6$. We conclude that $|f[(\eta_0)_\rho] - f[(\eta_1)_\sigma]| \leq \epsilon$, as required. □

Claim 4.2.5. *The distributions $(\eta_1)_\rho$ and ρ are $(11 \cdot \epsilon/\delta, \mathcal{F})$-indistinguishable.*

Proof. Let $f \in \mathcal{F}$ be arbitrary. We have that $f[(\eta_1)_\rho] - f[\rho]$ equals

$$\frac{\mathbb{E}_\rho[f \cdot \eta_1]}{d_\rho(\eta_1)} - \mathbb{E}_\rho[f] = \frac{1}{d_\rho(\eta_1)} \cdot \mathbb{E}_\rho[f \cdot (\eta_1 - d_\rho(\eta_1))] \leq \frac{1}{d_\rho(\eta_1)} \cdot \mathbb{E}_\rho[1 - d_\rho(\eta_1)] = \frac{1}{d_\rho(\eta_1)} - 1,$$

where for the inequality we first used $f(x) \geq 0$ for all $x \in U$ to get $f(x) \cdot (\eta_1(x) - d_\rho(\eta_1)) \leq f(x) \cdot (1 - d_\rho(\eta_1))$, and then used $1 - d_\rho(\eta_1) \geq 0$ to get $f(x) \cdot (1 - d_\rho(\eta_1)) \leq 1 \cdot (1 - d_\rho(\eta_1))$. By Claim 4.2.3, $d_\rho(\eta_1) \geq 1 - (7/3)\epsilon/\delta$, and so, $1/d_\rho(\eta_1) - 1 \leq (\epsilon/\delta)/(3/7 - \epsilon/\delta) \leq (21/2)\epsilon/\delta$, where we used our assumption that $\epsilon \leq \delta/3$ to get the lower bound $3/7 - \epsilon/\delta \geq 3/7 - 1/3 = 2/21$.

Thus, $f[(\eta_1)_\rho] - f[\rho] \leq (10.5)\epsilon/\delta$, for every $f \in \mathcal{F}$. Since $\mathcal{F}$ is closed under complement, we also get for every $f \in \mathcal{F}$ that $1 - f[(\eta_1)_\rho] - (1 - f[\rho]) = f[\rho] - f[(\eta_1)_\rho] \leq (10.5)\epsilon/\delta$. □

Finally, we argue that η_0 is a model for ρ in σ. Let $f \in \mathcal{F}$ be arbitrary. By the triangle inequality and Claims 4.2.4 and 4.2.5, we get $|f[\rho] - f[(\eta_0)_\sigma]| \leq |f[\rho] - f[(\eta_1)_\rho]| + |f[(\eta_1)_\rho] - f[(\eta_0)_\sigma]| \leq (10.5)\epsilon/\delta + \epsilon \leq 12\epsilon/\delta$. Hence, η_0 is a $(12 \cdot \epsilon/\delta, \mathcal{F})$-model for ρ in (U, σ) of density at least $\delta - 3\epsilon$. □

4.3 Tightness of Model quality

From the above reduction we see that for a δ-$(O(\epsilon), \textsc{Th}_\lambda[\mathcal{F}])$-pseudodense distribution, it is possible to construct a measure of small complexity which is a $(O(\epsilon/\delta), \mathcal{F})$-model. In [Imp09], Impagliazzo gave a construction which demonstrates that this is tight up to a constant factor in the sense that there exists a class $\mathcal{F}$ and a δ-$(O(\epsilon), \textsc{Th}_\lambda[\mathcal{F}])$-pseudodense distribution which (for appropriate universal non-zero constant) cannot be $(k\epsilon/\delta, \mathcal{F})$-indistinguishable from any δ-dense distribution.

This question can be asked of $\text{DMT}_{\textsc{Pseudorandom}}$ (Theorem 2.2.6) as the same relation between the density of target distribution in the pseudorandom distribution and the model parameter holds there. We show this is indeed true.

First we revisit Impagliazzo's construction and then give the analogous construction for pseudorandom $\text{DMT}_{\textsc{Pseudorandom}}$:

Theorem 4.3.1 (Impagliazzo)**.** *In (U, u), there exists $S \subset U$, a class $\mathcal{F}$, such that S has $(\epsilon, \mathcal{F}_\lambda)$-pseudodensity δ (where λ is as needed by $\mathrm{DMT}_{\textsc{Pseudodensity}}$) and there exists a function $f \in \mathcal{F}$ which $O(\epsilon/\delta)$ distinguishes S from every δ-dense distribution.*

Proof. Let $T \subset S \subset U$ with $d_U[S] = \epsilon, d_S[T] = \epsilon/\delta$. Choose $\mathcal{F} = \{t \doteq \mathbf{1}_T, \bar{t}, \mathbf{1}, \mathbf{0}\}$, so $\mathcal{F}_\lambda = \mathcal{F}$ for every λ (permissible).

Note for $f = \mathbf{1}, \mathbf{0}$, $f[U] \geq \delta f[S] - \epsilon$ is trivially true. Now $t[U] = d_U[\bar{S}]t[\bar{S}] + d_U[S]t[S] = 0 + \epsilon \cdot \epsilon/\delta = O(\epsilon)$, while $t[S] = \epsilon/\delta$ by construction. So $t[U] = O(\epsilon) \geq \delta\epsilon/\delta - \epsilon = \delta t[S] - \epsilon$.

Similarly for $\bar{t}$, $\bar{t}[U] = 1 - t[U] = 1 - O(\epsilon)$ and $\bar{t}[S] = 1 - t[S] = 1 - \epsilon/\delta$, therefore, for appropriate δ, $t[U] = 1 - \epsilon \geq \delta(1 - \epsilon/\delta) - \epsilon = \delta t[S] - \epsilon$ holds.

Thus, S has $(\epsilon, \mathcal{F}_\lambda)$-pseudodensity δ, and by $\mathrm{DMT}_{\textsc{Pseudodensity}}$ has a δ-dense model (ignoring the ϵ slack in model density). But for every μ with density δ , $t[\mu] = d_u[\mu] \cdot \mathbb{E}_{\mu_u}[t] = \delta \cdot (d_{\mu_u}[\bar{S}]t[\bar{S}] + d_{\mu_u}[S]t[S]) \in [0, \epsilon]$.

While $t[S] = \epsilon/\delta$, giving $|t[S] - t[\mu]| = O(\epsilon/\delta)$ implying that t is the required distinguisher. □

To give the analogous example demonstrating tightness of relation between density and model parameters for $\mathrm{DMT}_{\textsc{Pseudorandom}}$, we'll finesse with the above construction to give a $(O(\epsilon), \mathcal{F}_\lambda)$-pseudorandom R with S δ-dense in R, and the rest of the argument is identical.

Define $R = S \cup V$ where $V \subset \bar{S}$ is such that $d_R[S] = \delta, d_R[V] = \bar{\delta}$ (any choice of V works). We'll show that R is $(O(\epsilon), \mathcal{F}_\lambda)$-pseudorandom.

Since $\mathcal{F} = \mathcal{F}_\lambda$, and for $f = \mathbf{1}, \mathbf{0}$, $f[U] - f[R] = 0$ obviously and while $\bar{t}[U] - \bar{t}[R] = t[R] - t[U]$, to show that R is $(O(\epsilon), \mathcal{F}_\lambda)$-pseudorandom, we just need to show $|t[R] - t[U]| \leq O(\epsilon)$.

Note that $t[U] = O(\epsilon)$ as before. While $t[R] = d_R[V]t[V] + d_R[S]t[S] = 0 + \delta\epsilon/\delta = O(\epsilon)$. Therefore, as needed $|t[R] - t[U]| \leq O(\epsilon)$.

So we have the following corollary to Theorem 4.3.1:

Corollary 4.3.2. *In (U, u), there exists $S, R \subset U$, a class $\mathcal{F}$, such that S has density δ inside R and R is $(\epsilon, \mathcal{F}_\lambda)$-pseudorandom(where λ is as needed by $\mathrm{DMT}_{\textsc{Pseudorandom}}$) and there exists a function $f \in \mathcal{F}$ which $O(\epsilon/\delta)$ distinguishes S from every δ-dense distribution.*

4.4 Pseudodensity to Pseudorandandomness

Lemma 2.2.9 shows that if a distribution ρ has density δ inside a distribution τ where τ is $(\epsilon, \mathcal{F})$-pseudorandom in $\mathcal{U}$. Then ρ has $(\epsilon, \mathcal{F})$-pseudodensity at least δ inside σ. We'll prove a relaxed version of the converse which says that if a distribution is ρ is δ-$(\epsilon, \mathcal{F}')$-pseudodense then there must be some $(O(\epsilon), \mathcal{F})$-pseudorandom distribution inside which it's δ-dense is also true (where $\mathcal{F}' = \mathrm{TH}_1[\mathcal{CH}[\mathcal{F}]]$). Note it's easy to see that for every fixed choice of f there exists

a distribution τ_f which looks pseudorandom to f and inside which ρ is $\delta - dense$; the problem is that such τ_f may look pseudorandom only to that function f.

Since $\mathcal{F}$ is a strictly smaller than $\mathcal{F}'$, the result is weak; however, this sort of loss is expected as we need to come up with a universal distribution which looks pseudorandom to every permissible function (i.e. the quantifiers need to be changed from *there exists* to *for all*), where as in Lemma 2.2.9 no such universal distribution is needed, which makes this pseudodensity condition less demanding. This switching of quantifiers is exactly the reason that in Dense Model Theorems and Hardcore Lemmas the hypothesis are against $\textsc{Th}_\lambda[\mathcal{F}]$ while the conclusion holds for $\mathcal{F}$. Given this it remains unclear is the result obtained here can be improved in general.

Theorem 4.4.1. *Given $\epsilon > 0$ and finite probability space (U, u), if $S \subset U$ has $(\epsilon, \textsc{Th}_1[\mathcal{CH}[\mathcal{F}]])$-pseudodensity δ then there's a measure μ with S_u δ-dense in μ_u and μ_u $(O(\epsilon), \mathcal{F})$-pseudorandom.*

Proof. Assume there exists no measure γ with S_u δ-dense in γ_u and γ_u $(2\epsilon, \mathcal{F})$-pseudorandom, i.e. for every γ with S_u δ-dense in γ_u, there exists f such that $f[\gamma_u] - f[u] > \epsilon$. Define $\mathcal{R}_\delta = \{\gamma : U \to [0,1] | \gamma_u \geq \delta S_u\}$, i.e. $\gamma \in \mathcal{R}$ implies S_u is δ-dense in γ_u.

Consider the two player zero-sum game: *player A* plays measures $\gamma \in \mathcal{R}_\delta$ and *player B* $f \in \mathcal{F}$ with *Player B's* payoff $f[\gamma_u] - f[u]$ which *player B* is trying to maximize. By the *Min-Max Theorem*, there is a value α for this game such that there exists optimal strategies $\gamma^* \in \mathcal{CH}[\mathcal{R}_\delta], f^* \in \mathcal{CH}[\mathcal{F}]$ for *players A, B* satisfying:

$$\forall \gamma \in \mathcal{R}_\delta,\ f^*[\gamma_u] - f^*[u] \geq \alpha \text{ and } \forall f \in \mathcal{F},\ f[\gamma_u^*] - f[u] \leq \alpha$$

Note that $\mathcal{CH}[\mathcal{R}_\delta] = \mathcal{R}_\delta$, so by assumption for γ^* there exists $f_{\gamma^*} \in \mathcal{F}$ such that $\alpha \geq f_{\gamma^*}[\gamma_u^*] - f_{\gamma^*}[u] > \epsilon$, therefore, $f^*[\gamma_u] - f^*[u] > \epsilon$.

Remark 4.4.2. *Ideally f^* should be replaceable by an average of some poly(ϵ^{-1}) functions from $\mathcal{F}$, f°, with only $\epsilon/2$ loss, however, sampling from the distribution governing f^* to get f° fails since the choice of samples is not independent of choice of γ.*

Claim 4.4.3. *There exists $V \subset U$ such that $f^*[u] \geq \delta f^*[S_u] + \bar{\delta} f^*[V_u] - \epsilon$.*

Proof. Suppose not, so for every set V, $f^*[u] < \delta f^*[S_u] + \bar{\delta} f^*[V_u] - \epsilon$. Fix $V = \{\textsc{argmin}_U[f^*]\}$. By averaging there is threshold θ such that $f_\theta^*[u] < \delta f_\theta^*[S_u] + \bar{\delta} f_\theta^*[V_u] - \epsilon$. Now $f_\theta^*[u] < 1 - \epsilon$ so for some $z \in U$ $f_\theta^*(z) = 0$ implying $f_\theta^*[V] = 0$ (since $x \in V$ yields $f^*(x) \leq f^*(z)$ giving $f_\theta^*(x) \leq f_\theta^*(z) = 0$), thus $f_\theta^*[u] < \delta f_\theta^*[S_u] - \epsilon$ with $f_\theta^* \in \textsc{Th}_\lambda[\mathcal{F}]$ contradicting the $(\epsilon, \textsc{Th}_\lambda[\mathcal{F}])$-pseudodensity condition. □

So $f^*[u] \geq \delta f^*[S_u] + \bar{\delta} f^*[V_u] - \epsilon$ for some V. Define a measure $\mu = cS + c_1 V$ (with $c, c_1 \in \mathbb{R}^{\geq 0}$ such that $\mathbb{E}_u[cS]/\mathbb{E}_u[\mu] = \delta$). The existence of such μ is easy to see: $\mu' = S + kV$

with $k = (1/\delta - 1)\mathbb{E}_u[S]/\mathbb{E}_u[V]$ satisfies $\mathbb{E}_u[S]/\mathbb{E}_u[\mu'] = \delta$; if μ' is a valid measure then μ' is the required μ, otherwise $\mu = \mu'/\max_U\{\mu'(x)\}$ is as needed.

By lemma 2.1.2, $\mu_u = \delta S_u + \bar{\delta} V_u$ and $(cS)_u = S_u$ is δ-dense in μ_u implying $\mu \in \mathcal{R}_\delta$ and $f^*[\mu_u] = \mathbb{E}_{\mu_u}[f^*] = \delta f^*[S_u] + \bar{\delta} f^*[V_u]$. Therefore, by construction, $f^*[u] \geq \delta f^*[S] + \bar{\delta} f^*[V] - \epsilon = f^*[\mu_u] - \epsilon$ giving $\epsilon \geq f^*[\mu_u] - f^*[u]$ with $\mu \in \mathcal{R}_\delta$: contradiction. □

Chapter 5

The Bregman Projection Framework

5.1 Bregman Functions and Generalized Entropy

The reduction from $\text{HCL}_{\text{STRONG}}$ to $\text{DMT}_{\text{PSEUDODENSITY}}$ given in 4.2.2 relies on $\text{HCL}_{\text{STRONG}}$ for the space $\hat{U}$ with the distribution $\hat{\sigma}$. Even when the base distribution is uniform, $\hat{\sigma}$ can be wild. The reduction that we'll demonstrate from $\text{DMT}_{\text{MIN-MAX}}$ to $\text{HCL}_{\text{STRONG}}$ also uses a trick of this sort; so we need DMT and HCL w.r.t. arbitrary probability distributions. Since the online learning approach of Zhang and Barak et al. [Zha11, BHK09] achieves tight parameters this is the direction we want to follow. However, Zhang's work is based on results of Barak et al. who in turn use that the entropy function is a *Bregman function* and generates the Kullback-Leibler Divergence which satisfies Bregman's Theorem (Theorem 5.2.1). To push this approach through the setting of general finite probability spaces, we need analogues of entropy and Kullback-Leibler Divergence in such spaces, and we need to show that Bregman projection framework of Barak et al. holds. We start by introducing Bregman functions. The exposition in 5.1.1 follows Censor and Zenios [CZ97].

5.1.1 Bregman Functions

For a sufficiently differentiable function, $f : \Lambda \subset \mathbb{R}^n \to \mathbb{R}$, and a non-empty open convex set $\mathcal{S}$ with its closure $\bar{\mathcal{S}} \subset \Lambda$ define $D_f : \bar{\mathcal{S}} \times \mathcal{S} \to \mathbb{R},\ D_f(x, y) = f(x) - f(y) - \langle \nabla f(y), x - y \rangle$ where ∇f is the gradient of f. Also, define the partial level sets of D_f in the first and second variable:

1. $L_1^f(y, \alpha) = \{x \in \bar{\mathcal{S}} : D_f(x, y) \leq \alpha\}$ with $\alpha \in \mathbb{R}$
2. $L_2^f(x, \alpha) = \{y \in \mathcal{S} : D_f(x, y) \leq \alpha\}$ with $\alpha \in \mathbb{R}$

Definition 5.1.1 (Bregman function)**.** A function $f : \Lambda \subset \mathbb{R}^n \to \mathbb{R}$ is a Bregman function, if there exists a nonempty, open convex set $\mathcal{S}$ (the *zone*) with $\bar{S} \subset \Lambda$ satisfying

1. f is strictly convex and continuous on $\bar{\mathcal{S}}$.
2. $\forall i \in [n],\ \partial f/\partial x_i$ is continuous over $\mathcal{S}$.
3. $\forall \alpha \in \mathbb{R}, y \in \mathcal{S}, x \in \bar{\mathcal{S}}$, the partial level sets $L_1^f(y,\alpha), L_2^f(x,\alpha)$ are bounded.
4. If $y^v \in \mathcal{S}$ for all $v \geq 0$ and $\lim_{v\to\infty} y^v = y^*$ then $\lim_{v\to\infty} D_f(y^*, y^v) = 0$.
5. If $y^v \in \mathcal{S}, x^v \in \bar{\mathcal{S}}$ for all $v \geq 0$, $\lim_{v\to\infty} D_f(x^v, y^v) = 0$, $\lim_{v\to\infty} y^v = y^*$ with $\{x^v\}$ bounded then $\lim_{v\to\infty} x^v = y^*$.

Definition 5.1.2 (Generalized Distance)**.** For $f \in \mathcal{B}(S)$ (the space of Bregman function with zone S), define the *generalized distance function (D-function)* to be $D_f(x,y) : \bar{S} \times S \subset \mathbb{R}^{2n} \to \mathbb{R}$ by

$$D_f(x,y) \equiv f(x) - f(y) - \langle \nabla f(y), x - y \rangle \tag{5.1}$$

Definition 5.1.3 (Generalized projections)**.** Given $\Omega \subset \mathbb{R}^n, f \in \mathcal{B}(S), y \in S$, the generalized projection of y onto Ω is defined to be $x^* \in \Omega \cap \bar{S}$ such that

$$\min_{z \in \Omega \cap \bar{S}} D_f(z,y) = D_f(x^*, y) \tag{5.2}$$

Definition 5.1.4. For a distribution σ on U with $\sigma_i = \sigma(x_i)$, define the generalized entropy function $\text{ent}(\mathbf{x}) = -\sum_U \sigma_i x_i \log(x_i)$ *(where* $\log = \ln$*)*.

Remark 5.1.5. *Note that ent is simply* $\text{Ent} = -\sum_U x_i \log(x_i)$ *except with* i^{th} *coordinate scaled by* σ_i*. Also, from [CZ97],* $-\text{Ent}$ *is known to be a Bregman Function with associated Bregman Divergence* $\sum_U x_i \log\left(x_i y_i^{-1}\right) - \sum_U x_i + \sum_U y_i$ *which is the Kullback-Leibler Divergence.*

Remark 5.1.6. *Without loss of generality assume that* $\text{supp}[\sigma] = U$ *since if* $\sigma(x) = 0$*, then we may as well work with the space* $U - \{x\}$*.*

5.1.2 Generalized Entropy/KL Divergence

With the definitions introduced we now give the analogues of Entropy and Kullback-Leibler Divergence for a general (U, σ).

Theorem 5.1.7. *Suppose* σ *is a distribution on* U *(*$|U| = n$*) with* $\sigma_i = \sigma(x_i)$*, then* $-ent(\mathbf{x}) = \sum_U \sigma_i x_i \log(x_i)$ *is a Bregman Function with zone* $\mathcal{S} = \{\mathbf{x} \in \mathbb{R}^n : \forall j, x_j > 0\}$ *and associated Bregman Divergence* $D_{-ent} = \sum_U \sigma_i x_i \log\left(x_i y_i^{-1}\right) - \sum_U \sigma_i x_i + \sum_U \sigma_i y_i$*.*

Proof. $[-\nabla \text{ent}(y)]_i = \sigma_i \log(y_i) + \sigma_i$, so

$$-\langle \nabla[-\text{ent}(y)], x - y\rangle = -\sum_U \sigma_i(x_i \log(y_i) + x_i - y_i \log(y_i) - y_i) \text{ giving}$$

$$D_{-\text{ent}}(x, y) = \sum_U \sigma_i x_i \log(x_i) - \sum_U \sigma_i y_i \log(y_i) - \sum_U \sigma_i(x_i \log(y_i) + x_i - y_i \log(y_i) - y_i)$$

$$\text{Thus, } D_{-\text{ent}}(x, y) = \sum_U \sigma_i x_i \log\left(\frac{x_i}{y_i}\right) - \sum_U \sigma_i x_i + \sum_U \sigma_i y_i$$

Now we'll show that $f = -\text{ent}$ satisfies, all properties of definition 5.1.1.

Claim 5.1.8. *$-ent$ is strictly convex and continuous on $\bar{\mathcal{S}}$*

Proof. Obviously $-\text{ent}$ is continuous on $\mathcal{S}$; furthermore, continuity holds on $\bar{\mathcal{S}}$ as well, since the function $g(y) = y \log(y)$ is continuous at $y = 0$ as $\lim_{y\to 0} g(y) = 0 = g(0)$ (since $0 \log(0) = 0$ by convention) and so it follows that $-\text{ent}(\mathbf{x}) = \sum_U \sigma_i x_i \log(x_i)$ is continuous on $\bar{\mathcal{S}}$.

Now to show that $-\text{ent}$ is strictly convex we need that for all $t \in (0, 1)$, $\mathbf{x}, \mathbf{y} \in \bar{\mathcal{S}}$ with $\mathbf{x} \neq \mathbf{y}$, $-\text{ent}(t\mathbf{x} + (1-t)\mathbf{y}) < -t\text{ent}(\mathbf{x}) - (1-t)\text{ent}(\mathbf{y})$, that is,

$$\sum_U \sigma_i(tx_i + (1-t)y_i) \log(tx_i + (1-t)y_i) < \sum_U \sigma_i \left(tx_i \log(x_i) + (1-t)y_i \log(y_i)\right)$$

So if $g(y) = y \log(y)$ is strictly convex on $\mathbb{R}^{\geq 0}$ then it follows that $-\text{ent}$ is strictly convex as well. Note that on $\mathbb{R}^{>0}$, $d^2 g/dy^2 = \frac{1}{y}$ which never vanishes on $\mathbb{R}^{>0}$, so g is stricly convex [NP05]. To see that strict convexity extends to $\mathbb{R}^{\geq 0}$, we need that $\forall t \in (0, 1),\ (tx + (1-t)y) \log(tx + (1-t)y) < tx \log(x) + (1-t)y \log(y)$ with $y = 0, x > 0$. As $0 \log(0) = 0$, therefore, this becomes $\forall t \in (0, 1),\ tx \log(tx) < tx \log(x)$ which is obviously true since $tx \log(tx) = tx \log(x) + tx \log(t)$ and $\log(t) < 0$ as $t \in (0, 1)$. □

Claim 5.1.9. $\dfrac{\partial(-ent)}{\partial x_i}$ *is continuous on $\mathcal{S}$*

Proof. $\dfrac{\partial(-\text{ent})}{\partial x_i} = \sigma_i(\log(x_i) + 1)$ which is continuous on $\mathcal{S}$ as for all $x \in \mathcal{S}, x_i > 0$. □

Claim 5.1.10. *$\forall \alpha \in \mathbb{R}, y \in \mathcal{S}, x \in \bar{\mathcal{S}}$, the partial level sets $L_1^{-ent}(y, \alpha), L_2^{-ent}(x, \alpha)$ are bounded.*

Proof. The argument for boundedness of $L_1^{-\text{ent}}, L_2^{-\text{ent}}$ both is by contrapositive: we'll show that if $L_1^{-\text{ent}}(y, \alpha)$ is unbounded, that is, for some y, α, there exists $x \in L_1^{-\text{ent}}(y, \alpha)$ such that for some i the x_i goes to infinity, then $D_{-\text{ent}}(x, y) \leq \alpha$ cannot hold. And similarly for $L_2^{-\text{ent}}$. Also, throughout the argument, we'll implicitly be using that $\sigma_i \in (0, 1]$ for all i, and that U is finite.

So assume $x \in L_1^{-\text{ent}}(y, \alpha)$ for some finite α, a fixed $y \in \mathcal{S}$ and x_i goes to infinity for some coordinates. By definition $D_f(x, y) = \sum_U \sigma_i x_i \log(x_i y_i^{-1}) - \sum_U \sigma_i x_i + \sum_U \sigma_i y_i$. Obviously $\sum_U \sigma_i y_i$ is bounded as y is fixed. So consider $\sum_U \sigma_i x_i(\log(x_i y_i^{-1}) - 1)$. For every coordinate i such that x_i goes to infinity, $\sigma_i x_i(\log(x_i y_i^{-1} - 1) - 1)$ goes to positive infinity as well (using $\sigma_i > 0$ for all i) otherwise it is finite, therefore $D_{-\text{ent}}(x, y) \leq \alpha$ cannot hold.

Now suppose $y \in L_2^{-\text{ent}}(x, \alpha)$ for some finite α, a fixed $x \in \bar{\mathcal{S}}$ and y_i goes to infinity for some coordinates. Considering $D_{-\text{ent}}(x, y) = \sum_U \sigma_i x_i \log(x_i y_i^{-1}) - \sum_U \sigma_i x_i + \sum_U \sigma_i y_i$ $= \sum_U \sigma_i[x_i \log(x_i) - x_i \log(y_i) - x_i + y_i]$ again, note that the term $\sum_U \sigma_i[x_i \log(x_i) - x_i]$ is bounded for fixed x, while the term $\sigma_i[-x_i \log(y_i) + y_i]$ goes to positive infinity if y_i goes to positive infinity and is finite otherwise, so $D_{-\text{ent}}(x, y) \leq \alpha$ cannot hold.

Thus, it follows that $L_1^{-\text{ent}}, L_2^{-\text{ent}}$ must both be bounded. □

Claim 5.1.11. *If $y^v \in \mathcal{S}$ for all $v \geq 0$ and $\lim_{v\to\infty} y^v = y^*$ then $\lim_{v\to\infty} D_{-ent}(y^*, y^v) = 0$.*

Proof. $D_{-\text{ent}}(y^*, y^v) = \sum_U \sigma_i[y_i^* \log(y_i^*) - y_i^* \log(y_i^v) - y_i^* + y_i^v]$, as $\lim_{v\to\infty} y^v = y^*$ we have $\lim_{v\to\infty} y_i^v = y_i^*$ for each i, hence, obviously, $\lim_{v\to\infty} D_{-\text{ent}}(y^*, y^v) = 0$. □

Claim 5.1.12. *If $y^v \in \mathcal{S}, x^v \in \bar{\mathcal{S}}$ for all $v \geq 0$, $\lim_{v\to\infty} D_{-ent}(x^v, y^v) = 0$,$\lim_{v\to\infty} y^v = y^*$ with $\{x^v\}$ bounded then $\lim_{v\to\infty} x^v = y^*$.*

Proof. Suppose $y^v \in \mathcal{S}, x^v \in \bar{\mathcal{S}}$ for all $v \geq 0$, $\lim_{v\to\infty} D_{-\text{ent}}(x^v, y^v) = 0$,$\lim_{v\to\infty} y^v = y^*$ with $\{x^v\}$ bounded. We just need to show that any convergent subsequence $\{x^{v_s}\}$ (with $\lim_{s\to\infty}\{x^{v_s}\} = \bar{x}$) of $\{x^v\}$ converges to y^*; this is because $\{x^v\}$ being bounded implies that it must have a convergent subsequence $\{x^{v_s}\}$; now either $\{x^v\} - \{x^{v_s}\}$ is also an infinite sequence, in which case we can extract another convergent subsequence and iterate on this argument, or it is finite in which case we can group it with any one of the convergent subsequences and still have the convergence, so we partition $\{x^v\}$ into convergent subsequences, and if each one of them has the same limit then so does $\{x^v\}$.

Consider $t_i : \mathbb{R}^{\geq 0} \times \mathbb{R}^{>0} \to \mathbb{R}, t_i(x, y) = \sigma_i(x(\log(x/y) - 1) + y)$ with $x \geq 0, y > 0$. As $\sigma_i > 0$ we have that $t_i(x, y) = 0$ iff $x = y$. Now $\lim_{s\to\infty} D_{-\text{ent}}(x^{v_s}, y^{v_s}) = 0 \lim_{s\to\infty} \sum_U t_i(x_i^{v_s}, y_i^{v_s}) = 0$ It follows that for each i, $\lim_{s\to\infty} t_i(x_i^{v_s}, y_i^{v_s}) = 0$. Notice that $\{x_i^{v_s}\} \to \bar{x}_i$ and $\{y_i^{v_s}\} \to y_i^*$.

If $y_i^* > 0$ then as $t_i(x, y) = 0$ iff $x = y$ implies that $\bar{x}_i = y_i^*$, otherwise if $y_i^* = 0$ then $0 = t(\bar{x}_i, y_i^*) = t(\bar{x}_i, 0)\sigma_i(\bar{x}_i \log(\bar{x}_i) - \bar{x}_i \log(0) - \bar{x}_i)$ can hold only if $\bar{x}_i = 0$ as well, giving that $\bar{x}_i = y_i^*$ so $\lim_{v\to\infty} x^{v_s} = y^*$, which yields the claim. □

So, it follows that $f = -\text{ent}$ is a Bregman function. □

5.2 Bregman's Theorem

For completeness we give the proof of Bregnan's Theorem [CZ97, Bre67].

Theorem 5.2.1 (Bregman [Bre67]). *Suppose $f \in \mathcal{B}(S)$, $\Omega \subset \mathbb{R}^n$ be closed and convex with $\Omega \cap S \neq \emptyset$. Assume that $y \in S$ so $P_\Omega(y) \in S$. Let $z \in \Omega \cap \bar{S}$ then $\forall y \in S$*

$$D_f(z, P_\Omega(y)) + D_f(P_\Omega(y), y) \leq D_f(z, y) \tag{5.3}$$

Proof. Define $G(u) \equiv D_f(u, y) - D_f(u, P_\Omega(y))$ which implies

$$G(u) = f(u) - f(y) - \langle \nabla f(y), u - y\rangle - f(u) + f(P_\Omega(y)) + \langle \nabla f(P_\Omega(y)), u - P_\Omega(y)\rangle$$

$$\text{So } G(u) = f(P_\Omega(y)) - f(y) + \langle \nabla f(P_\Omega(y)), u - P_\Omega(y)\rangle - \langle \nabla f(y), u - y\rangle$$

$$\implies G(u) = f(P_\Omega(y)) - f(y) + \langle \nabla f(P_\Omega(y)), u\rangle - \langle \nabla f(P_\Omega(y)), P_\Omega(y)\rangle - \langle \nabla f(y), u\rangle + \langle \nabla f(y), y\rangle$$

This gives $\implies G(u) = \langle u, a\rangle + b$ with $a, b \in \mathbb{R}^n$, therefore, $G(u)$ is convex.

For $u_\lambda = \lambda z + (1-\lambda)P_\Omega(y)$ by convexity of $G(u)$, $G(u_\lambda) = D_f(u_\lambda, y) - D_f(u_\lambda, P_\Omega(y)) \leq \lambda G(z) + (1-\lambda)G(P_\Omega(y))$ implying

$$G(u_\lambda) \leq \lambda(D_f(z, y) - D_f(z, P_\Omega(y))) + (1-\lambda)(D_f(P_\Omega(y), y) - D_f(P_\Omega(y), P_\Omega(y)))$$

$$\implies D_f(u_\lambda, y) - D_f(u_\lambda, P_\Omega(y)) \leq \lambda(D_f(z, y) - D_f(z, P_\Omega(y))) + (1-\lambda)D_f(P_\Omega(y), y)$$

$$\implies D_f(z, y) - D_f(z, P_\Omega(y)) - D_f(P_\Omega(y), y) \geq \tfrac{1}{\lambda}\left(D_f(u_\lambda, y) - D_f(P_\Omega(y), y)\right) - \tfrac{1}{\lambda}D_f(u_\lambda, P_\Omega(y))$$

Note that $D_f(u_\lambda, y) - D_f(P_\Omega(y), y) \geq 0$ by definition of $P_\Omega(y)$.

Consider

$$\begin{aligned}
\lim_{\lambda\to 0} \tfrac{1}{\lambda} D_f(u_\lambda, P_\Omega(y)) &= \lim_{\lambda\to 0} \tfrac{1}{\lambda} D_f(\lambda z + (1-\lambda)P_\Omega(y), P_\Omega(y)) \\
&= \lim_{\lambda\to 0} \frac{D_f(\lambda(z - P_\Omega(y)) + P_\Omega(y), P_\Omega(y)) - D_f(P_\Omega(y), P_\Omega(y))}{\lambda} \\
&\implies \lim_{\lambda\to 0} \tfrac{1}{\lambda} D_f(u_\lambda, P_\Omega(y)) = \nabla_{\overrightarrow{z - P_\Omega(y)}} D_f(x, P_\Omega(y))\big|_{P_\Omega(y)} \\
\implies \lim_{\lambda\to 0} \tfrac{1}{\lambda} D_f(u_\lambda, P_\Omega(y)) &= \langle \nabla_x D_f(x, P_\Omega(y))\big|_{x=P_\Omega(y)}, z - P_\Omega(y)\rangle
\end{aligned}$$

Now $[\nabla_x D_f(x, P_\Omega(y))]_i$ is given by

$$\lim_{\lambda\to 0} \frac{f(x + \lambda e_i) - f(x) - \langle \nabla f(P_\Omega(y)), x + \lambda e_i - P_\Omega(y)\rangle + \langle \nabla f(P_\Omega(y)), x - P_\Omega(y)\rangle}{\lambda}$$

which implies

$$[\nabla_x D_f(x, P_\Omega(y))]_i = \lim_{\lambda\to 0} \frac{f(x+\lambda e_i) - f(x)}{\lambda} - \lim_{\lambda\to 0} \frac{\langle \nabla f(P_\Omega(y)), \lambda e_i\rangle}{\lambda}$$

$$\text{So } [\nabla_x D_f(x, P_\Omega(y))]_i = [\nabla f(x)]_i - \lim_{\lambda\to 0} \frac{\lambda\langle \nabla f(P_\Omega(y)), e_i\rangle}{\lambda} = [\nabla f(x)]_i - [\nabla f(P_\Omega(y))]_i$$

Thus, $\nabla_x D_f(x, P_\Omega(y))\big|_{x=P_\Omega(y)} = \mathbf{0}$ which yields $D_f(z,y) - D_f(z, P_\Omega(y)) - D_f(P_\Omega(y), y) \geq 0$ and so

$$D_f(z,y) \geq D_f(z, P_\Omega(y)) + D_f(P_\Omega(y), y)$$

□

5.3 Extending Barak et al

Next we give the extensions of results of Barak et al. [BHK09] by modifying their proofs to target (U, σ).

Theorem 5.3.1 (generalizing [BHK09])**.** *Let $\gamma \notin \Gamma_\delta$ be a measure over the probability space $\mathcal{U} = (U, \sigma)$ such that $d_\sigma(\text{SUPPORT}[\gamma]) \geq \delta$, for some $\delta \in [0,1]$. Let $c \geq 1$ be the smallest constant such that the measure $\mu = trunc_0^1[c \cdot \gamma]$ has $d_\sigma(\mu) = \delta$. Then $P_\delta \gamma = \mu$.*

Remark 5.3.2. *Since $d_\sigma(\text{SUPPORT}[\gamma]) \geq \delta$ such c exists and $c \leq min_{x_i \in \text{SUPPORT}[\gamma]}\{\gamma_i\}^{-1}$. In case $\gamma \in \Gamma_\delta$ then trivially $P_\delta\gamma = \gamma$.*

Proof. The proof follows the that of [BHK09] closely: we consider

$$\tilde{f}(\mathcal{M}) := \mathrm{D}(\mathcal{M}||\gamma) = \sum_U \sigma_i \mathcal{M}_i \log(\mathcal{M}_i/\gamma_i) + \sum_U \sigma_i\gamma_i - \sum_U \sigma_i \mathcal{M}_i$$

over the polytope Γ_δ. To show that μ is the projection we need to establish that μ minimizes $\tilde{f}$. Now if $\gamma_i = 0$ then $[\tilde{f}(\mathcal{M})]_i = 0$ if $\mathcal{M}_i = 0$ and $[\tilde{f}(\mathcal{M})]_i = \infty$ otherwise, therefore if γ_i vanishes then for any candidate minimizer $\mathcal{M}^*, \mathcal{M}_i^*$ must vanish as well.

Using that $\gamma_i = 0$ implies $\mu_i = 0$, to argue that μ minimizes $\tilde{f}(\mathcal{M})$ over Γ_δ it is sufficient to argue that μ minimizes

$$f(\mathcal{M}) := \mathrm{D}(\mathcal{M}||\gamma) = \sum_{\text{SUPPORT}[\gamma]} \sigma_i \mathcal{M}_i \log(\mathcal{M}_i/\gamma_i) + \sum_{\text{SUPPORT}[\gamma]} \sigma_i\gamma_i - \sum_{\text{SUPPORT}[\gamma]} \sigma_i \mathcal{M}_i$$

over $\Gamma_\delta^\gamma = \{\mathcal{M} \in \Gamma_\delta : \text{SUPPORT}[\mathcal{M}] \subset \text{SUPPORT}[\gamma]\}$.

Note that f is differentiable and convex in every variable. The convexity follows from convexity of the constant functions and the function $y\log(y)$, while $\partial f/\partial x_i = [\nabla f(\mathcal{M})]_i = \sigma_i \log(\mathcal{M}_i/\gamma_i)$ if $x_i \in \text{SUPPORT}[\gamma]$ and 0 otherwise which shows differentiability of each coordinate.

Now the tangent plane to $f(\mathcal{M})$ at $\mathcal{M} = \mu$ is given by $f(\mu)+\nabla f(\mu)^T(\mathcal{M}-\mu)$. By convexity we must have $f(\mathcal{M}) \geq f(\mu) + \nabla f(\mu)^T(\mathcal{M} - \mu)$ for every $\mathcal{M} \in \Gamma_\delta^\gamma$. To show that $P_\delta\gamma = \mu$, we need that μ minimizes f which follows if $\nabla f(\mu)^T(\mathcal{M} - \mu) \geq 0$ for every $\mathcal{M} \in \Gamma_\delta^\gamma$.

First consider x_i such that $\mu_i = 1$, since $\mu = trunc_0^1[c \cdot \gamma]$, therefore, $\gamma_i \geq \frac{1}{c}$. So $[\nabla f(\mu)]_i = \sigma_i \log(1/\gamma_i) \leq \sigma_i \log(c)$. Note that $\mathcal{M}_i - \mu_i = \mathcal{M}_i - 1 \leq 0$ and $\sigma_i \log(c) \geq 0$, giving

$$[\nabla f(\mu)]_i(\mathcal{M}_i - \mu_i) \geq \sigma_i \log(c)(\mathcal{M}_i - \mu_i) \tag{5.4}$$

And if $\mu_i < 1$ then $\mu_i = c\gamma_i$ so $[\nabla f(\mu)]_i = \sigma_i \log(\mu_i/\gamma_i) = \sigma_i \log(c)$, yielding

$$[\nabla f(\mu)]_i(\mathcal{M}_i - \mu_i) = \sigma_i \log(c)(\mathcal{M}_i - \mu_i) \tag{5.5}$$

Using 5.4, 5.5, $\nabla f(\mu)^T(\mathcal{M} - \mu) \geq \sum_{\text{SUPPORT}[\gamma]} \sigma_i \log(c)(\mathcal{M}_i - \mu_i) = \log(c)(d_\sigma[\mathcal{M}] - d_\sigma[\mu])$ since $\text{SUPPORT}[\mu] = \text{SUPPORT}[\gamma]$ and we are considering $\mathcal{M}$ with $\text{SUPPORT}[\mathcal{M}] \subset \text{SUPPORT}[\gamma]$, furthermore $\log(c)(d_\sigma[\mathcal{M}] - d_\sigma[\mu])$ because $log(c) \geq 0$ and $d_\sigma[\mathcal{M}] \geq d_\sigma[\mu] = \delta$ as otherwise $d_\sigma[\mathcal{M}] < \delta$ but $\mathcal{M} \in \Gamma_\delta$, so $\nabla f(\mu)^T(\mathcal{M} - \mu) \geq 0$ as needed. □

Lemma 5.3.3 (generalizing [BHK09][1])**.** *For $\delta \in [0, 1]$, let γ be a measure over the probability space $\mathcal{U} = (U, \sigma)$ such that $P_\delta\gamma = trunc_0^1[c \cdot \nu]$ for $c \in [1, 1 + \zeta]$, where $\zeta > 0$. Suppose we have oracle access to ν, and that we can sample an element from $\mathcal{U}$ in time t. Then, for any $0 < p < 1$, we can compute an implicitly represented approximate projection $\epsilon\delta$-$P_\delta\gamma$ in time $O\left(t\delta^{-1}\epsilon^{-2}(\log\log\zeta\epsilon^{-1} + \log p^{-1})\right)$, with probability $1-p$. Moreover, the computed approximate projection has the form $trunc_0^1[\tilde{c} \cdot \nu]$, for some $\tilde{c} \in [1, 1 + \zeta]$.*

Proof. As before, we follow [BHK09]: Suppose $\mu^* = trunc_0^1[\tilde{c} \cdot \nu]$ for some $\tilde{c} \in [1, 1 + \zeta]$ satisfies $d_\sigma[\mu^*] \in [\delta, (1 + \epsilon)\delta]$, then we claim that $\mu^* = \epsilon\delta$-$P_\delta\gamma$.

To see this, note that μ^* satisfies that for every i $\mu_i^* \geq P_\delta\gamma_i$ because we know $P_\delta\gamma = trunc_0^1[c \cdot \nu]$ and as $\mu^* = trunc_0^1[\tilde{c} \cdot \nu]$ with $d_\sigma[\mu^*] \geq \delta$ hence $\tilde{c} \geq c$, and $d_\sigma[\mu^*] - d_\sigma[P_\delta\gamma] \leq \epsilon\delta$ which comes from 5.3.1: as $d_\sigma[P_\delta\gamma] = \delta$, and by definition of μ^*, $d_\sigma[\mu^*] \in [\delta, (1 + \epsilon)\delta]$.

[1]The approximation parameter in [BHK09] is $\epsilon\delta N$ rather than $\epsilon\delta$ in our case; this difference is due to our scaling of N-dimensional vectors by the distribution σ.

Now to show that μ^* is $\epsilon\delta$-$P_\delta\gamma$, we need that $\mu^* \in \Gamma_\delta$ and $\mathrm{D}(\mathcal{M}||\mu^*) \leq \mathrm{D}(\mathcal{M}||P_\delta\gamma) + \epsilon\delta$. Obviously $\mu^* \in \Gamma_\delta$ because $d_\sigma[\mu^*] \geq \delta$, and to see the second note,

$$\mathrm{D}(\mathcal{M}||\mu^*) - \mathrm{D}(\mathcal{M}||P_\delta\gamma) = \sum_{\mathcal{U}} \sigma_i \mathcal{M}_i \log\left(\frac{P_\delta\gamma_i}{\mu_i^*}\right) - \sum_{\mathcal{U}} \sigma_i P_\delta\gamma_i + \sum_{\mathcal{U}} \sigma_i \mu_i^*$$

Since $\forall x_i \ P_\delta\gamma_i/\mu_i^* \leq 1$ so $\log(P_\delta\gamma_i/\mu_i^*) \leq 0$ giving $\mathrm{D}(\mathcal{M}||\mu^*) - \mathrm{D}(\mathcal{M}||P_\delta\gamma) \leq \sum_{\mathcal{U}} \sigma_i\mu_i^* - \sum_{\mathcal{U}} \sigma_i P_\delta\gamma_i = d_\sigma[\mu^*] - d_\sigma[P_\delta\gamma] \leq \epsilon\delta$.

The constructive part of 5.3.3 is identical to that of [BHK09]: $\tilde{c}$ as needed can be found by binary search over $[1, 1+\zeta]$ and the *with high probability clause* comes from using Hoeffding's bounds to estimate the density by sampling at each iteration. □

5.3.1 Generalized Barak et al's Total Loss Lemma

The extensions of Barak et al. results tie in to yield the following Lemma (as in Barak et al's original framework) which captures the game playing setup as introduced in section 2.1.5.2. This lemma is fundamental to the analysis of our constructions.

Theorem 5.3.4 (*Generalized* Total Loss Lemma (Barak et al))**.** *Suppose (U, σ) is a finite probability space and let Γ be a closed convex set of measures on U. Let $\epsilon \in (0, \frac{1}{2})$, $\mu^1 \in \Gamma$ be arbitrary and $m^t(x)$ be arbitrary penalty. Define $\gamma^{t+1} = (1 - \frac{\epsilon}{4})^{m^t(x)}\mu^t(x)$ and $\mu^{t+1} = \alpha - approx\ P_\Gamma\gamma^{t+1}$. Then $\forall \mu \in \Gamma$*

$$\sum_{i=1}^{\lambda} \mathbb{E}_\sigma[\mu^t, m^t] - \tfrac{\alpha}{\epsilon}\lambda \leq (1+\epsilon)\sum_{i=1}^{\lambda} \mathbb{E}_\sigma[\mu, m^t] + \tfrac{1}{\epsilon}\mathrm{D}(\mu||\mu^1)$$

where $\lambda \in \mathbb{N}$ and $\mathrm{D}(\cdot||\cdot)$, the generalized Bregman Divergence, D_{-ent} w.r.t. distribution σ.

Proof. The proof follows by the same computations as Barak et al: by linearity of expectation σ factors out at every step. To start we have:

$$\begin{aligned}\mathrm{D}(\mu||\gamma^{t+1}) - \mathrm{D}(\mu||\mu^t) = &\sum_U \sigma_i\mu(x_i)\log\left(\frac{\mu(x_i)}{\gamma^{t+1}(x_i)}\right) - \sum_U \sigma_i\mu(x_i) + \sum_U \sigma_i\gamma^{t+1}(x_i) \\ &- \left(\sum_U \sigma_i\mu(x_i)\log\left(\frac{\mu(x_i)}{\mu^t(x_i)}\right) - \sum_U \sigma_i\mu(x_i) + \sum_U \sigma_i\mu^t(x_i)\right)\end{aligned}$$

This yields $\mathrm{D}(\mu||\gamma^{t+1}) - \mathrm{D}(\mu||\mu^t) = \sum_U \sigma_i\mu(x_i)\log\left(\frac{\mu^t(x_i)}{\gamma^{t+1}(x_i)}\right) - \sum_U \sigma_i\mu^t(x_i) + \sum_U \sigma_i\gamma^{t+1}(x_i)$. Using $\frac{\mu^t(x_i)}{\gamma^{t+1}(x_i)} = (1-\epsilon)^{-m_i^t}$ gives:

$$\sum_U \sigma_i \mu(x_i) \log\left(\frac{\mu^t(x_i)}{\gamma^{t+1}(x_i)}\right) = \sum_U \sigma_i \mu(x_i) \log\left((1-\epsilon)^{-m_i^t}\right) = -\sum_U \sigma_i \mu(x_i) m_i^t \log(1-\epsilon).$$

By $-\log(1-\epsilon) \leq \epsilon(1+\epsilon)$ for $\epsilon \leq \frac{1}{2}$,

$$\sum_U \sigma_i \mu(x_i) \log\left(\frac{\mu^t(x_i)}{\gamma^{t+1}(x_i)}\right) \leq \epsilon(1+\epsilon) \sum_U \sigma_i \mu(x_i) m_i^t = \epsilon(1+\epsilon)\mathbb{E}_\sigma[\mu, m^t]$$

Now $\sum_U \sigma_i \gamma^{t+1}(x) = \sum_U (1-\epsilon)^{m_i^t} \sigma_i \mu^t(x) \leq \sum_U (1 - m_i^t \epsilon)\sigma_i \mu^t(x)$ meaning:

$$\sum_U \sigma_i \gamma^{t+1}(x) \leq \sum_U \sigma_i \mu^t(x) - \sum_U \epsilon \sigma_i m_i^t \mu^t(x) = \sum_U \sigma_i \mu^t(x) - \epsilon \mathbb{E}_\sigma[M^t, m^t].$$

Hence $\mathrm{D}(\mu||\gamma^{t+1}) - \mathrm{D}(\mu||\mu^t) \leq \epsilon(1+\epsilon)\mathbb{E}_\sigma[\mu, m^t] - \epsilon\mathbb{E}_\sigma[M^t, m^t]$, and by definition of $(\alpha, \sigma)-$ *approx* $P_\delta \gamma^{t+1}$ $\mathrm{D}(\mu||\gamma^{t+1}) \geq \mathrm{D}(\mu||P_\delta\gamma^{t+1})$ so $\mathrm{D}(\mu||\gamma^{t+1}) \geq \mathrm{D}(\mu||\mu^{t+1}) - \alpha$ and, therefore, $\mathrm{D}(\mu||\mu^{t+1}) - \mathrm{D}(\mu||\mu^t) \leq \epsilon(1+\epsilon)\mathbb{E}_\sigma[\mu, m^t] - \epsilon\mathbb{E}_\sigma[M^t, m^t] + \alpha$. This implies $\sum_{t=1}^{\lambda} \mathrm{D}(\mu||\mu^{t+1}) - \mathrm{D}(\mu||\mu^t) \leq \sum_{t=1}^{\lambda} \left(\epsilon(1+\epsilon)\mathbb{E}_\sigma[\mu, m^t] - \mathbb{E}_\sigma[M^t, m^t] + \alpha\right)$ yielding:

$$\frac{\mathrm{D}(\mu||\mu^{\lambda+1})}{\epsilon} - \frac{\mathrm{D}(\mu||\mu^1)}{\epsilon} \leq \sum_{t=1}^{\lambda} \left((1+\epsilon)\mathbb{E}_\sigma[\mu, m^t] - \mathbb{E}_\sigma[M^t, m^t]\right) + \frac{\lambda\alpha}{\epsilon}$$

So it follows that $\sum_{t=1}^{\lambda} \mathbb{E}_\sigma[M^t, m^t] - \frac{\lambda\alpha}{\epsilon} \leq \sum_{t=1}^{\lambda}(1+\epsilon)\mathbb{E}_\sigma[\mu, m^t] - \frac{\mathrm{D}(\mu||\mu^{\lambda+1})}{\epsilon} + \frac{\mathrm{D}(\mu||\mu^1)}{\epsilon}$. Thus,

$$\sum_{t=1}^{\lambda} \mathbb{E}_\sigma[M^t, m^t] - \frac{\lambda\alpha}{\epsilon} \leq \sum_{t=1}^{\lambda}(1+\epsilon)\mathbb{E}_\sigma[\mu, m^t] + \frac{\mathrm{D}(\mu||\mu^1)}{\epsilon}$$

□

Chapter 6

Uniform Constructions

6.1 Algorithmic $\mathrm{DMT}_{\text{MIN-MAX}}$

6.1.1 On-Line Learning Algorithm

Next we introduce the *Online Learning Algorithm*[1] **OLL** (due to Zhang[Zha11] and Barak et al.[BHK09]), generalized to arbitrary finite probability space $\mathcal{U} = (U, \sigma)$. A variant of **OLL** was utilized by Zhang to give an analysis of query complexity of $\mathrm{DMT}_{\text{PSEUDORANDOM}}$ while Barak et al[BHK09] used **OLL** to get at hardcore measures.

The algorithm takes as input a distribution ρ over U, a class $\mathcal{F}$ of $[0,1]$-bounded functions over U along with parameters $0 < \epsilon, \delta < 1$. Starting with the constant δ-dense measure $\mu^1 = \delta \cdot \mathbf{1}$, the algorithm iterates doing multiplicative updates and approximate Bregman projections on to the space of δ-dense measures to get a new δ-dense measure μ^{t+1} from the current measure μ^t. The algorithm uses as "penalty" m^t the function f which witness that μ^t is not a model for ρ. In usual setup **OLL** has access to an oracle with produces such witnesses.

After at most $T = O(\epsilon^{-2} \log \delta^{-1})$ iterations, either we get, for some $1 \leq t \leq T$, a measure μ^t that is a δ-dense model for the distribution ρ, or we get that the average F of all penalty functions m^t is a "universal" distinguisher in the sense that the same function F is a witness to failure of *every* δ-dense measure to be a model for ρ. More precisely, we have the following:

Theorem 6.1.1 (Analysis of **OLL**). *For a finite probability space $(\mathcal{U}, \sigma)$, a class $\mathcal{F}$, a distribution ρ over U, and parameters $\epsilon, \delta \in (0,1)$, the algorithm* **OLL** *finds*

1. *Either a δ-dense $(\epsilon, \mathcal{F})$-model μ^t for ρ, for some $1 \leq t \leq T$ with $\mu^t \in \mathcal{F}^*_{O(T)}$,*
2. *Or a test $F : U \to [0,1]$ such that $F[\mu_\sigma] - F[\rho] > \epsilon/4$ for every measure μ with $d_\sigma(\mu) = \delta$ (thereby witnessing that ρ has no δ-dense $(\epsilon/4, \mathcal{F})$-model) where $F = \frac{1}{T}\sum_{t=1}^{T} m^t$ with $m^t \in \mathcal{F}$.*

[1] For more details on-line learning algorithms setup see Blum [Blu96]

Algorithm 1 OLL$[\rho, \mathcal{F}, \delta, \epsilon]$ (Generalized DMT$_{\text{ALGORITHMIC}}$[Zha11])

$t = 1$, define the measure μ^1 by $\mu^1(x) = \delta$, for each $x \in U$
$T = \frac{16}{\epsilon^2} \log(\frac{1}{\delta})$, $\alpha = \frac{\epsilon^2 \delta}{16}$
while $t < T$ **do**
  if $\exists f \in \mathcal{F}$ such that $f[(\mu^t)_\sigma] - f[\rho] > \epsilon$ **then**
    $m^t = f$ for some $f \in \mathcal{F}$ satisfying $f[(\mu^t)_\sigma] - f[\rho] > \epsilon$
    $\nu^{t+1}(x) = (1 - \frac{\epsilon}{4})^{m^t(x)} \mu^t(x)$
    $\mu^{t+1} = \alpha\text{-}\ P_\delta \nu^{t+1}$
    $t = t + 1$
  else
    return MODEL, μ^t
  end if
end while
return DISTINGUISHER, $F = \frac{1}{T} \sum_{t=1}^{T} m^t$

Proof. Observe that the algorithm **OLL** outputs "MODEL" only if, for some μ^t, it holds that $f[(\mu^t)_\sigma] - f[\rho] \leq \epsilon$ for every $f \in \mathcal{F}$. Since the class $\mathcal{F}$ is closed under negation, the same inequality holds also for $\bar{f}$, which implies that $f[\rho] - f[(\mu^t)_\sigma] \leq \epsilon$ for every $f \in \mathcal{F}$ as well. So we get that $|f[(\mu^t)_\sigma - f[\rho]| \leq \epsilon$ for all $f \in \mathcal{F}$, and hence, μ^t is a model for ρ. By construction, all measures μ^j, $1 \leq j \leq T$, are δ-dense, and so μ^t is a δ-dense $(\epsilon, \mathcal{F})$-model for ρ.

Now suppose that the algorithm **OLL** outputs "DISTINGUISHER". We will show that in this case the function F constructed by the algorithm is such that, for every measure μ over U with $d_\sigma(\mu) = \delta$,

$$F[\mu_\sigma] - F[\rho] > \tfrac{\epsilon}{4}. \tag{6.1}$$

Observe that F has complexity T relative to $\mathcal{F}$ as $F = \frac{1}{T} \sum_{t=1}^{T} m^t$ for $m^t \in \mathcal{F}$. Using Eq. (6.1) and the fact that F is an average of tests m^t from $\mathcal{F}$, we get that some m^t (with t depending upon μ) is a witness to μ not being an $(\epsilon/4, \mathcal{F})$-model for ρ. Hence, ρ has no δ-dense $(\epsilon/4, \mathcal{F})$-model.

Now we prove Eq. (6.1). First, by the construction of F, we have

$$\sum_T \mathbb{E}_{(\mu^t)_\sigma}[m^t] > \sum_T \mathbb{E}_\rho[m^t] + \epsilon T, \tag{6.2}$$

where $\sum_T$ denotes the summation $\sum_{t=1}^{T}$. By the Total Loss Lemma, Lemma 5.3.4, we get

$$\sum_T \mathbb{E}_\sigma[\mu^t \cdot m^t] \leq (1 + \tfrac{\epsilon}{4}) \sum_T \mathbb{E}_\sigma[\mu \cdot m^t] + \tfrac{4}{\epsilon} \mathrm{D}(\mu || \mu^1) + \tfrac{4\alpha}{\epsilon} T. \tag{6.3}$$

We have $\mathbb{E}_\sigma[\mu^t \cdot m^t] = d_\sigma(\mu^t) \cdot \mathbb{E}_{(\mu^t)_\sigma}[m^t]$, and hence, using the fact that $d_\sigma(\mu^t) \geq \delta$ for all $1 \leq t \leq T$ and Eq. (6.2), we conclude that

$$\sum_T \mathbb{E}_\sigma[\mu^t \cdot m^t] \geq \delta \cdot \sum_T \mathbb{E}_{(\mu^t)_\sigma}[m^t] > \delta \left(\sum_T \mathbb{E}_\rho[m^t] + \epsilon \cdot T \right) = \delta \cdot T \cdot (\mathbb{E}_\rho[F] + \epsilon). \tag{6.4}$$

Similarly, we have that $\mathbb{E}_\sigma[\mu \cdot m^t] = d_\sigma(\mu) \cdot \mathbb{E}_{\mu_\sigma}[m^t] = \delta \cdot \mathbb{E}_{\mu_\sigma}[m^t]$, and so,

$$\sum_T \mathbb{E}_\sigma[\mu \cdot m^t] = \delta \cdot T \cdot \mathbb{E}_{\mu_\sigma}[F]. \tag{6.5}$$

We also have

$$\mathrm{D}(\mu||\mu^1) = \mathbb{E}_\sigma \left[\mu \cdot \log \left(\tfrac{\mu}{\mu^1} \right) \right] + d_\sigma(\mu_1) - d_\sigma(\mu) \leq \delta \log \delta^{-1}, \tag{6.6}$$

using the fact that $\mu^1(x) = \delta$ for all $x \in U$, and hence $d_\sigma(\mu^1) = \delta = d_\sigma(\mu)$, as well as the inequality $\mu(x) \cdot \log(\mu(x)/\delta) \leq \mu(x) \cdot \log(1/\delta)$, which follows from $0 \leq \mu(x) \leq 1$ and $0 \log 0 = 0$.

Using Eqs. (6.4)–(6.6) inside Eq. (6.3), we get

$$\delta T \cdot (\mathbb{E}_\rho[F] + \epsilon) < \left(1 + \frac{\epsilon}{4}\right) \delta T \cdot \mathbb{E}_{\mu_\sigma}[F] + \frac{4}{\epsilon} \delta \log \delta^{-1} + \frac{4\alpha}{\epsilon} \cdot T. \tag{6.7}$$

Dividing both sides of Eq. (6.7) by δT, and using the definition of $T = 16\epsilon^{-2} \log \delta^{-1}$ and $\alpha = \epsilon^2 \delta / 16$, we get $\mathbb{E}_{\mu_\sigma}[F] - \mathbb{E}_\rho[F] > \epsilon - (\epsilon/4) \cdot \mathbb{E}_{\mu_\sigma}[F] - (\epsilon/4) - (\epsilon/4)$. The latter is at least $\epsilon/4$, since $\mathbb{E}_{\mu_\sigma}[F] \leq 1$. □

Remark 6.1.2. *Consider a two-player zero-sum game where the first player chooses $f \in \mathcal{F}$ and the second player chooses a δ-dense measure μ, with the payoff for the first player given by $f[\mu_\sigma] - f[\rho]$. Then* **OLL**$[\rho, \mathcal{F}, \delta, \epsilon]$ *returns either a mixed strategy for the second player which does well against every strategy of first player, or vice versa for the first player. Thus* **OLL** *finds approximately optimal mixed strategies for this class of zero-sum games between δ-dense measures and algorithms $\mathcal{F}$. (This topic has been explored by by Vadhan and Zheng [VZ11] as well.) In fact, the algorithm* **OLL** *is the Multiplicative Weights Algorithm of Freund and Schapire [FS99] for approximately solving two-player zero-sum games, combined with taking Bregman projections (as in the Smooth Boosting Algorithm of Kale[Kal07] and Barak et al. [BHK09]); the projections onto the set Γ_δ of δ-dense measures are taken to ensure that the strategy of the second player at each stage is a δ-dense measure.*

It's easily evident that the function F computed in the "DISTINGUISHER" branch of the algorithm **OLL** satisfies the requirements of the Constructive $\mathrm{DMT}_{\text{Min-Max}}$ (Theorem 2.2.10) with $\lambda = 16\epsilon^{-2} \log \delta^{-1}$. We just need to establish the complexity bounds on the model to show **OLL** gives an explicit form of $\mathrm{DMT}_{\text{Min-Max}}$.

Observe that the complexity of the measures μ^j, for $1 < j \leq T$, produced by the algorithm is $O(j)$. Indeed, μ^1 has complexity 1. For $j > 1$, ν^j is obtained from μ^{j-1} using one exponentiation and multiplication. Finally, the approximate Bregman projection of ν^j requires two extra operations: scalar multiplication and truncation, by Lemma 5.3.3. Overall, μ^j is obtained from μ^{j-1} using a constant number of operations. So the complexity of model μ^t returned by the algorithm **OLL** in the "MODEL" branch is at most $O(t) \leq O(\lambda)$; more precisely:

Lemma 6.1.3. *For all measures μ^t, $1 \leq t \leq T$ computed by* **OLL***, if the witnesses m^s (with $s \leq t$) produced by the oracle are $\{0,1\}$-Boolean then $\mu^t \in \mathcal{F}_{5t}$, otherwise $\mu^t \in \mathcal{F}^*_{5t}$.*

Proof. The proof as outlined above is by induction on t. For $t = 1$, $\textsc{comp}_{\mathcal{F}}[\mu^1] = 1 \leq 5$, since $\mu^1 = \delta \cdot \mathbf{1}$. Assuming the claim holds for t, consider the complexity of updating m^t: $\nu^{t+1}(x) = (1 - \frac{\epsilon}{4})^{m^t(x)}\mu^t(x)$. In case m^t is Boolean $(1 - \frac{\epsilon}{4})^{m^t(x)}\mu^t(x) = (m^t(x) \cdot (1 - \frac{\epsilon}{4}) + (1 - m^t(x))) \cdot \mu^t(x) = (1 - \frac{\epsilon}{4}m^t(x)) \cdot \mu^t$ giving $\textsc{comp}_{\mathcal{F}}[\nu^{t+1}] \leq \textsc{comp}_{\mathcal{F}}[\mu^t] + 3$ which is at most $5 \cdot t + 3$ by the inductive assumption.

If m^t is not Boolean then only a single exponentiation operation is needed so the above complexity bound still holds but w.r.t to the to set of operations inclusive of exponentiation. By Lemma 5.3.3, $\mu^{t+1} = trunc_0^1(c_t \cdot \nu^{t+1})$ for some constant c_t, and so, $\textsc{comp}_{\mathcal{F}}[\mu^{t+1}] \leq \textsc{comp}_{\mathcal{F}}[\nu^{t+1}] + 2 \leq 5 \cdot t + 5 = 5 \cdot (t + 1)$ where whether $\textsc{comp}$ includes exponentiation (i.e. $\textsc{comp}^*$) depends on the witnesses used. □

With this it follows that 6.1.1 is an explicit form of Theorem 2.2.10 ($\text{DMT}_{\text{MIN-MAX}}$):

Corollary 6.1.4. *Algorithm* **OLL** *yields an algorithmic* $\text{DMT}_{\text{MIN-MAX}}$ *(Theorem 2.2.10) with $\lambda(\epsilon, \delta) = O(\epsilon^{-2} \log \delta^{-1})$ and $c = 4$.*

6.1.2 Avoiding Exponentiation

As observed earlier, if $\mathcal{F}$ is a Boolean class, the operation of limited exponentiation is not needed. In general, when $\mathcal{F}$ is a class of $[0,1]$-bounded functions, we can use thresholding to get from $f \in \mathcal{F}$ satisfying the condition in the **if**-statement of the algorithm **OLL** a *Boolean* function $f' = th_\theta(f)$, for some $\theta \in [0,1]$, such that f' satisfies the same condition.

To see this observe that for any function $g : U \to [0,1]$ and every $x \in U$,

$$g(x) = \mathbb{E}_{\theta \in [0,1]}[th_\theta[g(x)]].$$

Hence, for any g and any distribution π over U, $\mathbb{E}_\pi[g] = \mathbb{E}_{\theta \in [0,1]}[\mathbb{E}_\pi[th_\theta[g]]]$ (refer to Lemma A.2.2 for details). Finally, using this equality, the linearity of expectation, and averaging, we conclude that if there is some f such that $f[(\mu^t)_\sigma] - f[\rho] > \epsilon$, then there is also some $\theta \in [0,1]$, such that, for $f' = th_\theta[f]$, we have $f'[(\mu^t)_\sigma] - f'[\rho] > \epsilon$.

So we have that for an arbitrary class $\mathcal{F}$, we either get a model μ^t of low complexity *without using exponentiation*, or get a universal distinguisher F that is the average of few *thresholded* functions from $\mathcal{F}$ (because now the updates must use the thresholded witnesses). Thus, we can trade the simplicity of the model for the extra complexity of the universal distinguisher. As we show next, this thresholding can be made *discrete*: the distinguisher can be made to be an average of functions from $\mathcal{F}^{O(\epsilon)}$ i.e. the thresholds needed are not arbitrary. This is accounts for use of classes $\mathcal{F}^{O(\epsilon)}$ in the hypothesis of Theorems 2.2.3, 2.2.6, 2.2.7.

6.2 Algorithmic $\mathrm{DMT}_{\text{MIN-MAX}}$ without exponentiation

Since given access to an oracle which produces a function $f \in \mathcal{F}$ witnessing that a current measure μ^t is not an $(\epsilon, \mathcal{F})$-model for ρ, the algorithm **OLL** will *efficiently construct* either a dense model μ for ρ or a universal distinguisher F. So we have an *algorithmic* $\mathrm{DMT}_{\text{MIN-MAX}}$. The idea here's to make the thresholding process algorithmic as well.

To avoid the exponentiation operation following the "thresholding" approach outlined above, one needs to find appropriate thresholds θ efficiently (hence the need to discretized the thresholds as pointed out). This can be done by random sampling, with some slight loss in parameters. The details follow.

Consider the algorithm **OLL′** which is slightly modded **OLL**:

Algorithm 2 OLL′$[\rho, \mathcal{F}, \delta, \epsilon]$ (Modified $\mathrm{DMT}_{\text{ALGORITHMIC}}$)

$t = 1$, define the measure μ^1 by $\mu^1(x) = \delta$, for each $x \in U$
$T = \frac{256}{\epsilon^2}\log(\frac{1}{\delta})$, $\alpha = \frac{\epsilon^2\delta}{256}$
while $t < T$ **do**
  if $\exists f \in \mathcal{F}$ such that $f[(\mu^t)_\sigma] - f[\rho] > \epsilon$ **then**
    let $f \in \mathcal{F}$ be any function such that $f[(\mu^t)_\sigma] - f[\rho] > \epsilon$
    for each $0 \le n \le \lceil 2\epsilon^{-1} \rceil$, set $f_n = th_{n\epsilon/2}(f)$
    let $0 \le n^* \le \lceil 2\epsilon^{-1} \rceil$ be such that, with high probability, $f_{n^*}[(\mu^t)_\sigma] - f_{n^*}[\rho] > \epsilon/4$
    $m^t = f_{n^*}$
    $\nu^{t+1}(x) = (1 - \frac{\epsilon}{16})^{m^t(x)}\mu^t(x)$
    $\mu^{t+1} = \alpha\text{-}\ P_\delta \nu^{t+1}$
    $t = t + 1$
  else
    return MODEL, μ^t
  end if
end while
return DISTINGUISHER, $F = \frac{1}{T}\sum_{t=1}^{T} m^t$

The estimation of $f_n[\rho]$ and $f_n[(\mu^t)_\sigma]$ is done by random sampling; by Chernoff bounds, we can achieve the additive error at most $\epsilon/8$ for each, with high probability, in time $poly(\epsilon^{-1})$.

Assuming that the required integer value n^* exists (and hence can be efficiently found by random sampling and the testing the $\lceil 2\epsilon^{-1} \rceil$ possible n^* threshold values), the rest of the algorithm is the same as before, and so the old analysis of Theorem 6.1.1 applies: we either get a δ-dense $(\epsilon, \mathcal{F})$-model μ^t for ρ, or a universal distinguisher F with the slightly worse distinguishing parameter $\epsilon/16$ (rather than $\epsilon/4$). The model μ^t has complexity relative to $\mathcal{F}$ at most $10 \cdot t = O(t)$ by Lemma 6.1.3 (note that the complexity does not involve exponentiation i.e. $\mu^t \in \mathcal{F}_{O(t)}$) as we use the Boolean functions m^t, obtained from some f by a single threshold operation for the updates. The complexity of F is obviously at most $2T$.

It remains to argue that n^* always exists. To this end the following basic properties of the threshold operation are required(the notation $\mathbb{E}_{\theta \in [0,1]}$ means that the expectation is taken over a uniformly random value θ from the interval $[0, 1]$):

Lemma 6.2.1. *Let $g : U \to [0, 1]$ be any function, and let ρ and τ be any distributions over U. Suppose $\mathbb{E}_\rho[g] > \mathbb{E}_\tau[g] + \epsilon$ for some $\epsilon \in [0, 1]$. Then there exist κ, θ, n satisfying:*

1. *$\kappa \in [0, 1]$ and $\mathbb{E}_\rho[th_\kappa[g]] > \mathbb{E}_\tau[th_\kappa[g]] + \epsilon$.*
2. *$\theta \in [\epsilon/2, 1]$ and $\mathbb{E}_\rho[th_\theta[g]] > \mathbb{E}_\tau[th_{\theta-\epsilon/2}[g]] + \epsilon/2$ [RTTV08a].*
3. *$n \in \mathbb{N}, n \leq \lceil 2\epsilon^{-1} \rceil$ such that $\mathbb{E}_\rho[th_{n\epsilon/2}[g]] > \mathbb{E}_\tau[th_{n\epsilon/2}[g]] + \epsilon/2$.*

Proof. From Lemma A.2.2 (item (5)), for $\theta \sim \text{UNIFORM}[0, 1]$, $\mathbb{E}_\sigma[\mathbb{E}_\theta[\text{TH}_\theta[f]] = \mathbb{E}_\sigma[f]$, so by averaging there must exists a threshold $\kappa \in [0, 1]$ such that $\mathbb{E}_\rho[th_\kappa[g]] > \mathbb{E}_\tau[th_\kappa[g]] + \epsilon$ holds.

To see item (2), assume that it fails, i.e., for every $\theta \in [\epsilon/2, 1]$, $\mathbb{E}_\rho[th_\theta[g]] \leq \mathbb{E}_\tau[th_{\theta-\epsilon/2}[g]] + \epsilon/2$. Using Lemma A.2.2 (item (5)) and the fact that $th_t[g(x)]$ as a function of t with x fixed is piece-wise constant with only discontinuity at $g(x) = t$ gives:

$$\mathbb{E}_\rho[g] = \mathbb{E}_\theta[\mathbb{E}_\rho[th_\theta[g]]] = \int_0^{\epsilon/2} \mathbb{E}_\rho[th_\theta[g]]dt + \int_{\epsilon/2}^1 \mathbb{E}_\rho[th_\theta[g]]dt \leq \epsilon/2 + \int_0^1 \left(\mathbb{E}_\tau[th_\theta[g]] + \epsilon/2\right) dt$$

yielding $\mathbb{E}_\rho[g] \leq \mathbb{E}_{\theta \in [0,1]}[\mathbb{E}_\tau[th_\theta[g]]] + \epsilon = \mathbb{E}_\tau[g] + \epsilon$: contradiction.

Finally, θ in item (2) must satisfy $\theta \in [n\epsilon/2, (n+1)\epsilon/2]$ for some $n \in \mathbb{N}$ with $n \leq 2\lceil \epsilon^{-1} \rceil$. It follows that $n\epsilon/2 \in [\theta - \epsilon/2, \theta]$. By item (2), we get $\mathbb{E}_\rho[th_{n\epsilon/2}[g]] \geq \mathbb{E}_\rho[th_\theta[g]] > \mathbb{E}_\tau[th_{\theta-\epsilon/2}[g]] + \epsilon/2 \geq \mathbb{E}_\tau[th_{n\epsilon/2}[g]] + \epsilon/2$, where we used the fact that $th_\alpha[v] \geq th_\beta[v]$ for any $v, \alpha, \beta \in [0, 1]$ such that $\alpha \leq \beta$. □

The existence of n^* required by the modified algorithm **OLL′** follows from item (3) of Lemma 6.2.1. This, together with our arguments above, yields the following analysis of the modified algorithm **OLL′**.

Theorem 6.2.2 (Analysis of **OLL′**). *For a finite probability space (U, σ), a class $\mathcal{F}$, a distribution ρ over U, and parameters $\epsilon, \delta \in (0,1)$, the algorithm* **OLL′** *finds*

1. *Either a δ-dense $(\epsilon, \mathcal{F})$-model μ^t for ρ, for some $1 \leq t \leq T$ with $\mu^t \in \mathcal{F}_{O(T)}$,*
2. *Or $F : U \to [0,1]$ such that $F[\mu_\sigma] - F[\rho] > \epsilon/16$ for every measure μ with $d_\sigma(\mu) = \delta$, where $F = \frac{1}{T}\sum_{t=1}^{T} m^t$ with $m^t \in \mathcal{F}^{\epsilon/2}$.*

6.3 Applications of the DMT Algorithm

6.3.1 Constructive Dense Model Theorems

We first prove Theorem 2.2.7.2, re-stated in the following contrapositive form.

Theorem 6.3.1 ($\textsc{Dmt}_{\textsc{Pseudodensity}}$). *Given $\epsilon, \delta \in (0,1)$, there is a $\lambda = O(\epsilon^{-2}\log\delta^{-1})$ such that the following holds. Let $\mathcal{U} = (U, \sigma)$ be any finite probability space and let $\mathcal{F}$ be any class of functions over U.*

Suppose ρ is a probability distribution over U such that, for every δ-dense measure $\mu \in \mathcal{F}_\lambda$, there is an $f \in \mathcal{F}$ such that $|f[\mu_\sigma] - f[\rho]| > \epsilon$ (witnessing that μ is not an $(\epsilon, \mathcal{F})$-model for ρ). Then there is a $[0,1]$-valued function $\Phi \in \textsc{Th}_\lambda[\mathcal{F}^{\epsilon/2}]$ such that $\Phi[\sigma] < \delta \cdot \Phi[\rho] - \epsilon\delta/16$ (witnessing that ρ does not have $(\epsilon\delta/16, \textsc{Th}_{O(\lambda)}[\mathcal{F}^{\epsilon/2}])$-pseudodensity δ).

Proof. We run Algorithm **OLL′**$[\rho, \mathcal{F}, \delta, \epsilon]$, with $T = O(\epsilon^{-2}\log\delta^{-1})$. By Theorem 6.2.2, the algorithm **OLL′** either finds a δ-dense $(\epsilon, \mathcal{F})$-model μ for ρ with $\mu \in \mathcal{F}_{O(T)}$, or finds a universal distinguisher F. For $\lambda = O(T)$ sufficiently large, ρ does not have a δ-dense $(\epsilon, \mathcal{F})$-model $\mu \in \mathcal{F}_\lambda$. For such a λ, algorithm **OLL′** will produce the distinguisher $F \in \mathcal{F}_{O(T)}$ such that, for every measure γ with $d_\sigma(\gamma) = \delta$, $F[\rho] + \frac{\epsilon}{16} < F[\gamma_\sigma]$. This implies that $\bar{F}[\rho] > \bar{F}[\gamma_\sigma] + \frac{\epsilon}{16}$, where $\bar{F} = 1 - F$.

Next we argue similarly to [RTTV08a]. Order elements of U such that $\bar{F}(x_i) \geq \bar{F}(x_{i+1})$. Let n be the largest integer such that $d_\sigma[\{x_i : i \in [n]\}] < \delta$, define measure γ by $\gamma(x_i) = 0$ for $i \geq n+2$, $\gamma(x_i) = 1$ for $i \in [n]$ and $\gamma(x_{n+1}) = c$ where $c \in (0,1]$ is such that $d_\sigma[\gamma] = \delta$.

By Lemma 6.2.1 (item 1), there exists $\kappa \in [0,1]$ such that for $\Phi = th_\kappa[\bar{F}]$,

$$\Phi[\rho] > \Phi[\gamma_\sigma] + \tfrac{\epsilon}{16}. \tag{6.8}$$

Since $1 \geq \mathbb{E}_\rho[\Phi] > \mathbb{E}_{\gamma_\sigma}[\Phi] + \frac{\epsilon}{16}$, we get $1 > \mathbb{E}_{\gamma_\sigma}[\Phi]$. Since Φ is Boolean, we conclude that there is an $x^* \in \textsc{support}[\gamma]$ such that $\Phi(x^*) = 0$, in particular $\Phi(x_{n+1}) = 0$ since $\bar{F}(x_{n+1}) = min_{x \in \in \textsc{support}[\gamma]}[\bar{F}(x)]$. So we have that for every $x \in \textsc{support}[\bar{\gamma}] \subset \{x_i : i \geq n+1\}$, $\Phi(x) = 0$ implying $\mathbb{E}_\sigma[\Phi \cdot \bar{\gamma}] = 0$.

Using this, as well as the identity $d_\sigma(\gamma) \cdot \mathbb{E}_{\gamma_\sigma}[\Phi] = \mathbb{E}_\sigma[\Phi \cdot \gamma]$, we get $\mathbb{E}_\sigma[\Phi] = \mathbb{E}_\sigma[\Phi \cdot \gamma] + \mathbb{E}_\sigma[\Phi \cdot \bar{\gamma}] = \delta \cdot \mathbb{E}_{\gamma_\sigma}[\Phi]$. By Eq. (6.12), we conclude that $\Phi[\sigma] < \delta \cdot \Phi[\rho] - \frac{\epsilon\delta}{16}$ and clearly, $\Phi \in \mathrm{TH}_\lambda[\mathcal{F}^{\epsilon/2}]$ since $F = \frac{1}{T}\sum_{t=1}^{T} m^t$ with $m^t \in \mathcal{F}^{\epsilon/2}$, therefore $\bar{F} = \frac{1}{T}\sum_{t=1}^{T} \overline{m^t}$ and by definition $m^t \in \mathcal{F}^{\epsilon/2}$ implies $\overline{m^t} \in \mathcal{F}^{\epsilon/2}$. □

Next we prove Theorem 2.2.6.2, re-stated as below. However, we first note the following remark:

Remark 6.3.2. *Making the arguments of Theorems 6.3.1, 6.3.3 but using* **OLL** *instead of* **OLL′** *yields the Theorems 2.2.6.1, 2.2.6.1 in contrapositive.*

Theorem 6.3.3 ($\mathrm{DMT}_{\mathrm{PSEUDORANDOM}}$)**.** *Given $\epsilon, \delta \in (0,1)$, there is a $\lambda = O(\epsilon^{-2}\log\delta^{-1})$ such that the following holds. Let $\mathcal{U} = (U, \sigma)$ be any finite probability space, and let $\mathcal{F}$ be any class of functions over U.*

Suppose ρ is a probability distribution over U such that, for every δ-dense measure $\mu \in \mathcal{F}_\lambda$, there is an $f \in \mathcal{F}$ such that $|f[\mu_\sigma] - f[\rho]| > \epsilon$ (witnessing that μ is not an $(\epsilon, \mathcal{F})$-model for ρ). Then there is a $[0,1]$-valued function $\Phi \in \mathrm{TH}_\lambda[\mathcal{F}^{\epsilon/2}]$ such that, for every distribution τ over U, where ρ is δ-dense inside τ, we get that $\Phi[\tau] - \Phi[\sigma] > \epsilon\delta/16$ (witnessing that τ is not $(\epsilon\delta/16, \mathrm{TH}_{O(\lambda)}[\mathcal{F}^{\epsilon/2}])$-pseudorandom).

Proof. As in the proof of Theorem 6.3.1, we get $\Phi \in \mathcal{F}_\lambda$ such that $\delta \cdot \Phi[\rho] - \Phi[\sigma] > \epsilon\delta/16$. By the δ-density of ρ inside τ, we get $\Phi[\tau] \geq \delta \cdot \Phi[\rho]$, and hence, $\Phi[\tau] - \Phi[\sigma] > \epsilon\delta/16$, as required. □

Remark 6.3.4. *The proofs of Theorems 6.3.1 and 6.3.3 yield* constructive *versions. For example, in the case of* $\mathrm{DMT}_{\mathrm{PSEUDODENSITY}}$*: if ρ has $(\frac{\epsilon\delta}{16}, \mathcal{F}_\lambda)$-pseudodensity at least δ, then it must be the case that, for some $t \leq T \in O(\epsilon^{-2}\log\delta^{-1})$, the measure $\mu^t \in \mathcal{F}_{O(T)}$ produced by the algorithm* **OLL′** *is a δ-dense $(\epsilon, \mathcal{F})$-model for ρ. The case of* $\mathrm{DMT}_{\mathrm{PSEUDORANDOM}}$ *is similar.*

Remark 6.3.5. *Making the arguments of Theorems 6.3.1, 6.3.3 but using* **OLL** *instead of* **OLL′** *yields the Theorems 2.2.6.1, 2.2.6.1 in contrapositive.*

Remark 6.3.6. *Observe that* **OLL′** *is used as a black box; any algorithm which can be used to prove 6.2.2 (or 6.1.1) with the model of low complexity and the uniform distinguisher F the an average of small number of functions from $\mathcal{F}$ would work. The same holds for the proofs of* LCAT *and* $\mathrm{HCL}_{\mathrm{STRONG}}$ *which follow.*

6.3.2 Constructive Low Complexity Approximation Theorem

Inspired by our proof of Theorem 3.2.1 (our reduction from Dense Model Theorems to LCAT), we also prove LCAT from Theorem 6.2.2. We re-state it below.

Theorem 6.3.7 (Uniform Low Complexity Approximation Theorem)**.** *Given any $\epsilon > 0$, there exists a $\lambda = O(\epsilon^{-2})$ such that the following holds. Let $\mathcal{U} = (U, \sigma)$ be any finite probability space, any let $\mathcal{F}$ be any class of functions over U. Then, for every $g : U \to [0,1]$, there exists a function $h \in \mathcal{F}_\lambda$ such that h is an $(\epsilon, \mathcal{F})$-approximation of g in $\mathcal{U}$.*

Proof. Without loss of generality assume $\alpha = d_\sigma(g)$ is known, and that $\alpha \geq \frac{1}{2}$ (since otherwise we can work with $\bar{g}$). Set $\epsilon' = c \cdot \epsilon$, for a sufficiently small constant c to be determined.

Imagine running the algorithm $\mathbf{OLL}'[g_\sigma, \mathcal{F}, \alpha, \epsilon']$. We claim that $\mathbf{OLL}'$ must produce an α-dense model μ^t for g_σ, for some $t \in O(\epsilon^{-2} \log \alpha^{-1})$; note that $1 \leq \alpha^{-1} \leq 2$ implies that $t \in O(\epsilon^{-2})$.

Indeed, suppose otherwise. Then, by Theorem 6.3.1, we have a $\Phi : U \to [0,1]$ such that $\alpha \mathbb{E}_{g_\sigma}[\Phi] - \frac{\alpha\epsilon}{16} > \mathbb{E}_\sigma[\Phi]$. This and the identity $\mathbb{E}_\sigma[g \cdot \Phi] = d_\sigma(g) \cdot \mathbb{E}_{g_\sigma}[\Phi]$ imply that $\mathbb{E}_\sigma[\Phi] \geq \mathbb{E}_\sigma[g\Phi] = \alpha \cdot \mathbb{E}_{g_\sigma}[\Phi] > \frac{\delta\epsilon}{16} + \mathbb{E}_\sigma[\Phi]$. A contradiction.

So $\mathbf{OLL}'[g_\sigma, \mathcal{F}, \alpha, \epsilon']$ constructs a α-dense measure $\mu^t \in \mathcal{F}_\lambda$ which is an $(\epsilon', \mathcal{F})$-model for g_σ, where $\lambda \in O(\epsilon^2)$. By random sampling, we can estimate $d_\sigma(\mu^t)$, and scale μ^t down, if necessary, getting a model h for g_σ such that $\alpha - \epsilon' \leq d_\sigma(h) \leq \alpha + \epsilon'$ (with high probability). The rest of the argument is exactly the same as in the proof of Theorem 3.2.1. We get that, for all $f \in \mathcal{F}$, $|\mathbb{E}_\sigma[(g-h)f]| \leq O(\epsilon')$, which can be made less than ϵ by choosing c small enough.

Note that we could have used **OLL** instead of **OLL** to get a possibly qualitatively different approximation.

□

Remark 6.3.8. *The proof of Theorem 6.3.7 yields the* constructive *version of LCAT. As in the proof, imagine running* $\mathbf{OLL}'$*. For each α-dense measure μ^t produced by* $\mathbf{OLL}'$*, form a scaled-down measure $\tilde{\mu}^t$ such that $d_\sigma(\tilde{\mu}^t) \in [\alpha - \epsilon, \alpha + \epsilon]$ with high probability. Propose this $\tilde{\mu}^t$ as a candidate for an approximation of g. If this is not an approximation yet, and we get a function $f \in \mathcal{F}$ witnessing that $\tilde{\mu}^t$ is not a $(2\epsilon, \mathcal{F})$-approximation for g:*

$$\left|\mathbb{E}_\sigma[f \cdot (\tilde{\mu}^t - g)]\right| > 2\epsilon,$$

then we conclude (as in the proof of Theorem 3.2.1) that $|f[(\tilde{\mu}^t)_\sigma] - f[g_\sigma]| > \epsilon$. Hence, $|f[(\mu^t)_\sigma] - f[g_\sigma]| > \epsilon$, witnessing that μ^t is not an $(\epsilon, \mathcal{F})$-model for g_σ yet. Therefore, we can continue running $\mathbf{OLL}'$ *with this witness f, to obtain a new measure μ^{t+1}. Since* $\mathbf{OLL}'$ *cannot run for more than $T \in O(\epsilon^{-2})$ steps (as argued in the proof of Theorem 6.3.7 above), we will obtain a $(2\epsilon, \mathcal{F})$-approximation $h \in \mathcal{F}_{O(T)}$ for g within T iterations.*

6.3.3 Constructive Strong Hardcore Lemma

Here using algorithmic $\mathrm{DMT}_{\text{MIN-MAX}}$, we prove Theorem 2.2.3 by first showing that every hardcore measure, however small, has the model of the optimal density. Recall that the motivation for trying to push this through via $\mathrm{DMT}_{\text{MIN-MAX}}$ was discussed in section 4.1.1.

Lemma 6.3.9. *For any $\epsilon, \delta \in [0,1]$, there is a $\lambda \in O(\epsilon^{-2} \log \delta^{-1})$ such that the following holds. Let $\mathcal{U} = (U, \sigma)$ be any finite probability space, and let $\mathcal{F}$ be any Boolean class of functions over U. Suppose a function $g : U \to \{0,1\}$ is $(\delta, \mathrm{TH}_\lambda[\mathcal{F}])$-hard in $\mathcal{U}$. Let μ_0 be an $(\epsilon/8, \mathcal{F})$-hardcore measure for g in $\mathcal{U}$ of arbitrary density. Then the algorithm $\mathbf{OLL}[(\mu_0)_\sigma, g \oplus \mathcal{F}, 2\delta, \epsilon]$ will produce, within λ iterations, a 2δ-dense $((9/8)\epsilon, \mathcal{F})$-hardcore measure $\mu^* \in \mathcal{F}_{O(\lambda)}$ for g in $\mathcal{U}$.*

Proof. Suppose $\mathbf{OLL}[(\mu_0)_\sigma, g \oplus \mathcal{F}, 2\delta, \epsilon]$ returns MODEL. Then there exists some $t \leq T = \lambda \in O(\epsilon^{-2} \log \delta^{-1})$ such that μ^t is a 2δ-dense $(\epsilon, \mathcal{F})$-model for $(\mu_0)_\sigma$. That is, for $\mu^* = \mu^t$, we have that, for all $f \in \mathcal{F}$, $|\mathbb{E}_{(\mu^*)_\sigma}[g \oplus f] - \mathbb{E}_{(\mu_0)_\sigma}[g \oplus f]| \leq \epsilon$. By the definition of hardcore, we have that $|\mathbb{E}_{(\mu_0)_\sigma}[g \oplus f] - \frac{1}{2}| \leq \epsilon/8$. Plugging this into the above inequality yields $|\mathbb{E}_{(\mu^*)_\sigma}[g \oplus f] - \frac{1}{2}| \leq (9/8)\epsilon$, implying that μ^* is $((9/8)\epsilon, \mathcal{F})$-hardcore measure for g w.r.t σ. Also, by Lemma 6.1.3, we have that $\mu^* \in \mathcal{F}_{O(\lambda)}$, as required.

It remains to argue that the **OLL** algorithm cannot return DISTINGUISHER. Suppose otherwise, then by Theorem 6.1.1, we get a test $F = \frac{1}{T}\sum_{t=1}^{T}(g \oplus f_t)$, with $f_t \in \mathcal{F}$ for all $1 \leq t \leq T$, such that $F[S_\sigma] - F[(\mu_0)_\sigma] > \epsilon/4$ for every subset $S \subseteq U$ with $d_\sigma(S) = 2\delta$. Since μ_0 is an $\epsilon/8$-hardcore for g, we get that $(g \oplus f_t)[(\mu_0)_\sigma] \in [\frac{1}{2} - \epsilon/8, \frac{1}{2} + \epsilon/8]$ for every $1 \leq t \leq T$, and hence, $F[S_\sigma] > \frac{1}{2} + \epsilon/8$. Using the identity $\bar{F}[\rho] = 1 - F[\rho]$, we get

$$\mathbb{E}_{S_\sigma}[\bar{F}] < \tfrac{1}{2} - \epsilon/8. \tag{6.9}$$

Using the fact that g and $\mathcal{F}$ are Boolean, it is also easy to see that

$$\bar{F} = \frac{1}{T}\sum_{t=1}^{T}(g \oplus \bar{f}_t) = \left| g - \frac{1}{T}\sum_{t=1}^{T}\bar{f}_t \right|. \tag{6.10}$$

Denoting $\Phi = \frac{1}{T}\sum_{t=1}^{T}\bar{f}_t$, we get from Eqs. (6.9) and (6.10) that $\mathbb{E}_{S_\sigma}[|g - \Phi|] < \frac{1}{2}$. The contradiction is now achieved by the following Lemma first given by Holenstein [Hol06, Hol05]. Holenstein's original argument was sets not measures with base distribution as uniform; since there are some subtleties with getting the extension we need, section 6.3.3.2 contains a detailed exposition.

Lemma 6.3.10 (Holenstein)**.** *Suppose $\Phi : U \to [0,1]$ is such that, for all $S \subseteq U$ of $d_\sigma(S) = 2\delta$, $\mathbb{E}_{S_\sigma}[|g - \bar{\Phi}|] > \frac{1}{2}$ then there is a $\kappa \in [0,1]$ such that $\mathbb{E}_\sigma[|g - th_\kappa[\Phi]|] < \delta$.*

Note that $th_\kappa[\Phi] \in \text{TH}_\lambda[\mathcal{F}]$, but since g is $(\delta, \text{TH}_\lambda[\mathcal{F}])$-hard w.r.t σ this cannot hold true. □

Remark 6.3.11. *Again the proof of Lemma 6.3.9 is constructive. Suppose a measure μ^t constructed by the algorithm* **OLL** *is not a $((9/8)\epsilon, \mathcal{F})$-hardcore for g yet, and suppose we get a function $f \in \mathcal{F}$ showing that. Then $g \oplus f$ witnesses that μ^t is not an $(\epsilon, g \oplus \mathcal{F})$-model for $(\mu_0)_\sigma$, and so we can continue with* **OLL** *to the next measure μ^{t+1}. Within $O(\epsilon^{-2} \log \delta^{-1})$ iterations, the algorithm will produce a 2δ-dense $((9/8)\epsilon, \mathcal{F})$-hardcore measure $\mu \in (g \oplus \mathcal{F})_\lambda$ for g, where $\lambda \in O(\epsilon^{-2} \log \delta^{-1})$.*

So from lemma 6.3.9, it follows that weak HCL implies strong HCL. In fact, for the process described in Remark 6.3.11 to run, we do not need to have any initial hardcore measure μ_0. We can simply run **OLL** with an imaginary μ_0 (which we do not need to have explicitly in order to get new measures μ^{t+1} from current μ^t) by hardwiring $\mathbb{E}_{(\mu^*)_\sigma}[g \oplus f] = \frac{1}{2}$. Within $O(\epsilon^{-2} \log \delta^{-1})$ iterations, we will obtain a hardcore measure of density 2δ. This latter algorithm is, in fact, exactly the algorithm used by Barak et al. [BHK09] in their proof of the Strong Hardcore Lemma. Since we want to use **OLL** as a black-box and **OLL** needs a source measure, an option is to go through Trevisan et al's reduction from Low Complexity Approximation Theorem but this has problems as discussed before, so instead we seed the domain with an artificial hardcore measure.

6.3.3.1 Using an artificial hardcore measure

Suppose g is $(\delta, \text{TH}_\lambda[\mathcal{F}])$-*hard* on (σ, U) and $\epsilon > 0$ is given. Because of Lemma 2.1.14 we can assume $2\delta + 2\epsilon < 1$.

Set $X = \{x_1, x_2\}$. Fix $x_0 \in U$ and extend $g, f \in \mathcal{F}$ to $U_* = U \cup X$ by defining $g_*, f_* \in \mathcal{F}_* : U_* \to [0,1]$, $g_*(x_1) = \bar{g}_*(x_2) = g(x_0)$, $f_*(x_1) = f_*(x_2) = f(x_0)$ and $f_* \in \text{TH}_\lambda[\mathcal{F}_*] - \mathcal{F}_*$ extended to U_* is whatever it needs to be based on the extension of $\mathcal{F}$, however, note for all $f_* \in \text{TH}_\lambda[\mathcal{F}_*], f_*(x_1) = f_*(x_2)$. Also, it follows that $\mathcal{F}_*$ is a class as well (since if f extends to f_* then $\bar{f}$ extends to $\overline{f_*}$).

Define a distribution σ_* on U_* as σ on U with probability $(1-\epsilon\delta)$ and as uniform distribution on X with probability $\epsilon\delta$.

Claim 6.3.12. *Given $\epsilon > 0$, with respect to (σ_*, U_*), g_* is $(\delta - \epsilon\delta^2 + \frac{1}{2}\epsilon\delta, \text{TH}_\lambda[\mathcal{F}_*])$-hard and the set X is $(\epsilon, \mathcal{F}_*) - hardcore$.*

Proof. Since for all $f \in \text{TH}_\lambda[\mathcal{F}_*]$, $|g_*(x_1) - f_*(x_1)| = 1 - |g_*(x_2) - f_*(x_2)|$, $\mathbb{E}_X[|g_* - f_*|] = \frac{1}{2}$. Also, for any $f_* \in \text{TH}_\lambda[\mathcal{F}_*]$, $\mathbb{E}_{\sigma_*}[|g_* - f_*|] = (1-\epsilon\delta)\mathbb{E}_\sigma[|g - f|] + \epsilon\delta\mathbb{E}_X[|g_* - f_*|] \geq \delta - \epsilon\delta^2 + \frac{1}{2}\epsilon\delta$. □

So using $\text{DMT} - \text{ALGORITHM}$ with $\mathbf{1}_X$ yields a $(2\delta + \epsilon\delta(1-2\delta))$-*dense*, $(\epsilon, \mathcal{F}_*)$-*hardcore* measure μ_* w.r.t. σ_*. Set $\mu = \mu_*|_U$, the restriction of μ_* to the original space.

Claim 6.3.13. *μ is $2\delta - 2\epsilon^2\delta^3$-dense, $(3\epsilon, \mathcal{F})$-hardcore on (U, σ).*

Proof. First we establish the density lower bound:

$$(1-\epsilon\delta)d_\sigma[\mu] + \epsilon\delta \geq (1-\epsilon\delta)\mathbb{E}_\sigma[\mu] + \epsilon\delta\mathbb{E}_X[\mu_*] = d_{\sigma_*}[\mu_*] \geq 2\delta + \epsilon\delta - 2\epsilon\delta^2$$

$$\implies d_\sigma[\mu] \geq \frac{2\delta + \epsilon\delta - 2\epsilon\delta^2 - \epsilon\delta}{1-\epsilon\delta} \geq (2\delta - 2\epsilon\delta^2)(1+\epsilon\delta) = 2\delta - 2\epsilon^2\delta^3$$

where we used lemma A.1.2 in the last inequality.

Now we show that μ is $(3\epsilon, \mathcal{F})$-*hardcore* on (U, σ). Using μ_* is $(\epsilon, \mathcal{F}_*) - hardcore$ w.r.t. σ_*, for all $f_* \in \mathcal{F}_*$ we have:

$$\tfrac{1}{2} + \epsilon \geq \mathbb{E}_{(\mu_*)_{\sigma_*}}[|g_* - f_*|] = \frac{(1-\epsilon\delta)\mathbb{E}_\sigma[|g-f|\mu] + \epsilon\delta\mathbb{E}_X[|g_*-f_*|\mu_*]}{(1-\epsilon\delta)\mathbb{E}_\sigma[\mu] + \epsilon\delta\mathbb{E}_X[\mu_*]} \geq \frac{(1-\epsilon\delta)\mathbb{E}_\sigma[|g-f|\mu]}{(1-\epsilon\delta)\mathbb{E}_\sigma[\mu] + \epsilon\delta}$$

$$\implies \tfrac{1}{2} + \epsilon \geq \frac{\mathbb{E}_\sigma[|g-f|\mu]}{\mathbb{E}_\sigma[\mu] + \epsilon\delta/(1-\epsilon\delta)} \geq \frac{\mathbb{E}_\sigma[|g-f|\mu]}{\mathbb{E}_\sigma[\mu] + 2\epsilon\delta} \tag{6.11}$$

Using lemma A.1.4, and that $d_\sigma[\mu] \geq 2\delta - 2\epsilon^2\delta^3 \geq \delta$ yields:

$$\tfrac{1}{2} + 3\epsilon \geq \frac{\mathbb{E}_\sigma[|g-f|\mu]}{\mathbb{E}_\sigma[\mu] + 2\epsilon\delta} + 2\epsilon \geq \frac{\mathbb{E}_\sigma[|g-f|\mu]}{\mathbb{E}_\sigma[\mu] + 2\epsilon\delta} + \frac{2\epsilon\delta}{\mathbb{E}_\sigma[\mu] + 2\epsilon\delta} \geq \frac{\mathbb{E}_\sigma[|g-f|\mu]}{\mathbb{E}_\sigma[\mu]} = \mathbb{E}_{\mu_\sigma}[|g-f|]$$

□

Claim 6.3.14. *$\gamma = (1-\epsilon\delta)\mu + \mathbf{1}\epsilon\delta$ is 2δ-dense, $(\frac{7}{2}\epsilon, \mathcal{F})$-hardcore on (σ, U)*

Proof. Note that $d_\sigma[\gamma] \geq d_\sigma[\mu]$. Furthermore, $d_\sigma[\gamma] = (1-\epsilon\delta)d_\sigma[\mu] + \epsilon\delta d_\sigma[\mathbf{1}] = (2\delta - 2\epsilon^2\delta^3)(1-\epsilon\delta) + \epsilon\delta = 2\delta + \epsilon\delta(1 - 2\epsilon\delta^2 - 2\delta) \geq 2\delta$. Using 6.3.13, for all $f \in \mathcal{F}$,

$$\mathbb{E}_{\gamma_\sigma}[|g-f|] = \frac{(1-\epsilon\delta)\mathbb{E}_\sigma[|g-f|\mu]}{d_\sigma[\gamma]} + \frac{\epsilon\delta\mathbb{E}_\sigma[|g-f|]}{d_\sigma[\gamma]} \leq \frac{\mathbb{E}_\sigma[|g-f|\mu]}{d_\sigma[\mu]} + \frac{\epsilon\delta}{2\delta} = \mathbb{E}_{\mu_\sigma}[|g-f|] + \frac{\epsilon}{2}$$

therefore, $\mathbb{E}_{\gamma_\sigma}[|g-f|] \leq \frac{1}{2} + \frac{7}{2}\epsilon$. □

6.3.3.2 Holenstein's Derandomizaton

The Φ obtained from equation 6.10 can be thought of as generating a randomized function Ψ_r with $\Psi_r(x) = 1$ with probability $\Phi(x)$. If Φ was $\{0,1\}$-Boolean then 6.3.10 follows trivially: if $\bar{\Phi}$ is equal to g with probability more than $1/2$ on every 2δ dense measure then it cannot be incorrect with probability more than δ on the entire space since if it was then we could pick a set of x's where $\bar{\Phi}$ does not match g along with some more x's to make the density of the set 2δ and as more than half the x's in this set are where $\bar{\Phi}$ fails to match g, the probability of $\bar{\Phi}$

matching g on this set must be less than a half which is a contradiction. This argument fails if Φ is not Boolean, however, Holenstein in [Hol06, Hol05] showed that there's way to fix the randomization of Ψ so the probability of g matching Φ can still be bounded as above.

Lemma 6.3.15 (Holenstein)**.** *Suppose $\Phi : U \to [0,1]$ is such that, for all measures μ over U with $d_\sigma(\mu) = 2\delta$, $\mathbb{E}_{\mu_\sigma}[|g - \Phi|] > \frac{1}{2}$. Then there is a $\theta \in [0,1]$ such that $\mathbb{E}_\sigma[|g - th_\theta[\bar{\Phi}]|] < \delta$.*

Proof. Define $\alpha_c(x) := 2|g - \Phi| - 1$ and $\alpha_1(x) := 2\bar{\Phi}(x) - 1$. Order elements based on $\alpha_c(x)$ from smallest to largest, inducing the ordering on the elements $x \in U$: $x_1, x_2, x_3, \ldots, x_{|U|}$ and fix $n \in \mathbb{N}$ to be largest value for which $d_\sigma[\{x_1, \ldots, x_j\}] < 2\delta$. Define the measure μ over U as follows: $\mu(x_i) = 1$ for $1 \le i \le n$, $\mu(x_j) = 0$ for $j > n+1$, and $\mu(x_{n+1}) = c$ for $0 < c \le 1$ so that $d_\sigma(\mu) = 2\delta$. Note that $\textsc{Support}[\mu_\sigma] = \{x_1, \ldots, x_{n+1}\}$ and $\textsc{Support}[\bar{\mu}_\sigma] \subset \{x_{n+1}, \ldots, x_{|U|}\}$.

Define $\kappa := \max_{x \in \textsc{Support}[\mu_\sigma]}\{\alpha_c(x)\} = \alpha_c(x_{n+1})$. Note $\mathbb{E}_{\mu_\sigma}[\alpha_c(x)] > 0$ as, by assumption, $\mathbb{E}_{\mu_\sigma}[|g - \Phi|] > \frac{1}{2}$. This also means that $\kappa > 0$.

Consider the probabilistic Boolean function Ψ (with internal randomness r) which on input $x \in U$ behaves as follows:

$$\mathbb{P}_r[\Psi(x) = 1] = trunc_0^1\left[\frac{1}{2} + \frac{\alpha_1(x)}{2\kappa}\right] = trunc_0^1\left[\frac{1}{2} - \frac{1}{2\kappa} + \frac{\bar{\Phi}(x)}{\kappa}\right] \equiv trunc_0^1[a + b\bar{\Phi}(x)] \equiv \psi(x)$$

Claim 6.3.16.

$$\mathbb{P}_r[\Psi(x) = g(x)] = trunc_0^1\left[\frac{1}{2} + \frac{\alpha_c(x)}{2\kappa}\right]$$

.

Proof. For fixed x, $\mathbb{P}_r[\Psi(x) = g(x)] = \mathbb{P}_r[\Psi = 1|g = 1]\mathbb{P}[g = 1] + \mathbb{P}_r[\Psi = 0|g = 0]\mathbb{P}[g = 0]$ (the probabilities are over the internal randomness r of Ψ).

Note $\alpha_c := 2|g - \Phi| - 1 = 2g\bar{\Phi} + 2\bar{g}\Phi - 1 = g(2\bar{\Phi} - 1) + \bar{g}(2\Phi - 1)$. Using that $2\Phi - 1 = -(2\bar{\Phi} - 1) = -\alpha_1$ gives $\alpha_c = g\alpha_1 - \bar{g}\alpha_1$ and so if $g(x) = 1$ then $\alpha_1(x) = \alpha_c(x)$ and if $g(x) = 0$, $\alpha_1(x) = -\alpha_c(x)$ which implies:

$$\begin{aligned}
\mathbb{P}_r[\Psi(x) = 1|g(x) = 1] &= trunc_0^1\left[\frac{1}{2} + \frac{\alpha_c(x)}{2\kappa}\right] \\
\mathbb{P}_r[\Psi(x) = 0|g(x) = 0] &= 1 - trunc_0^1\left[\mathbb{P}[\Psi(x) = 1|g(x) = 0]\right] = 1 - trunc_0^1\left[\frac{1}{2} + \frac{-\alpha_c}{2\kappa}\right] \\
&= trunc_0^1\left[1 - \left(\frac{1}{2} + \frac{-\alpha_c}{2\kappa}\right)\right] = trunc_0^1\left[\frac{1}{2} + \frac{\alpha_c}{2\kappa}\right]
\end{aligned}$$

Therefore,

$$\mathbb{P}_r[\Psi = g] = trunc_0^1\left[\frac{1}{2} + \frac{\alpha_c}{2\kappa}\right](\mathbb{P}[g(x) = 0] + \mathbb{P}[g(x) = 1]) = trunc_0^1\left[\frac{1}{2} + \frac{\alpha_c}{2\kappa}\right]$$

. □

Since $\alpha_c(x_i) \geq \kappa$ for all $i \geq n+1$, we get $\mathbb{P}_r[\Psi(x_i) = g(x_i)] = 1$ for all $i > n$. Hence, in particular, for every $x \in U$

$$\mathbb{E}_{(\bar{\mu})_\sigma}[\mathbb{P}_r[\Psi(x) = g(x)]] = 1. \tag{6.12}$$

Next, for $1 \leq i \leq n+1$, we have $\alpha_c(x_i) \leq \kappa$, and so $\frac{1}{2} + \frac{\alpha_c(x_i)}{2\kappa} \leq 1$. Therefore, we get

$$\mathbb{E}_{\mu_\sigma}[\mathbb{P}_r[\Psi(x) = g(x)]] \geq \mathbb{E}_{\mu_\sigma}\left[\frac{1}{2} + \frac{\alpha_c(x)}{2\kappa}\right] = \frac{1}{2} + \frac{\mathbb{E}_{\mu_\sigma}[\alpha_c(x)]}{2\kappa} > \frac{1}{2}. \tag{6.13}$$

Since $\mathbb{E}_\sigma[\mathbb{P}_r[\Psi = g]] = \mathbb{E}_\sigma[\mu]\mathbb{E}_{\mu_\sigma}[\mathbb{P}_r[\Psi = g]] + \mathbb{E}_\sigma[\bar{\mu}]\mathbb{E}_{\bar{\mu}_\sigma}[\mathbb{P}_r[\Psi = g]] = 2\delta\mathbb{E}_{\mu_\sigma}[\mathbb{P}_r[\Psi = g]] + (1-2\delta)\mathbb{E}_{\bar{\mu}_\sigma}[\mathbb{P}_r[\Psi = g]]$ using Eqs. (6.12) and (6.13), yields $\mathbb{E}_\sigma[\mathbb{P}_r[\Psi = g]] > 1 - \delta$ and as g is independent of r, $\mathbb{E}_\sigma[1 - \mathbb{P}_r[\Psi = g]] = \mathbb{E}_\sigma[\mathbb{E}_r[|\Psi - g|]] = \mathbb{E}_\sigma[g\mathbb{E}_r[\bar{\Psi}] + \bar{g}\mathbb{E}_r[\Psi]] < \delta$. This gives

$$\delta > \mathbb{E}_\sigma[|\mathbb{E}_r[\Psi] - g|] = \mathbb{E}_\sigma[|\psi - g|] = \mathbb{E}_{k\sim[0,1]}[\mathbb{E}_\sigma[|th_k[\psi] - g|]]$$

Thus, there exists $k^* \in [0,1]$ such that $\mathbb{E}_\sigma[|th_{k^*}[\psi] - g|] < \delta$. By construction for any $k \in [0,1]$, $th_k[\psi] = th_k[\text{trunc}_0^1[a + b\bar{\Phi}]] = th_k[a + b\bar{\Phi}] = th_{(k-a)b^{-1}}[\bar{\Phi}]$. Fix $\theta = (k^* - a)b^{-1}$, So $th_{k^*}[\psi] = th_\theta[\bar{\Phi}]$ giving $\mathbb{E}_\sigma[|g - th_\theta[\bar{\Phi}]|] < \delta$ as needed.

And as $\kappa \in [-1, 1]$, therefore,

$$\theta = \frac{(k^* - a)}{b} = \left(k^* - \frac{1}{2}\left(1 - \frac{1}{\kappa}\right)\right)\kappa = k^*\kappa - \frac{\kappa - 1}{2} = \kappa k^* - \frac{\kappa}{2} + \frac{1}{2} \in [0,1]$$

. □

Appendix A

Estimates and Manipulations

Technical calculations that were deferred originally in favor of focusing on the bigger scheme are detailed here.

A.1 Estimates

Lemma A.1.1 (3.2.2)**.** *For any a, b, x, y, if $|x - y| \leq \epsilon_1$ and $|a/x - b/y| \leq \epsilon_2$, then $|a - b| \leq (b/y)\epsilon_1 + x\epsilon_2$.*

Proof. Since $b/y - \epsilon_2 \leq a/x \leq b/y + \epsilon_2$, we get $(x/y)b - x\epsilon_2 \leq a \leq (x/y)b + x\epsilon_2$. Since $y - \epsilon_1 \leq x \leq y + \epsilon_1$, we get $1 - (\epsilon_1/y) \leq x/y \leq 1 + (\epsilon_1/y)$. Putting these bounds on x/y inside the earlier bounds on a yields the claim. □

Lemma A.1.2. *For $\epsilon \in (0, \frac{1}{2})$, $1/(1 - \epsilon) \in (1 + \epsilon, 1 + 2\epsilon)$.*

Proof. $1/(1 - \epsilon) = 1 + \epsilon/(1 - \epsilon)$ and $1 + \epsilon \leq 1 + \epsilon/(1 - \epsilon) \leq 1 + 2\epsilon$ for $\epsilon \in (0, \frac{1}{2})$ □

Lemma A.1.3. *For $\epsilon > 0$, $1/(1 + \epsilon) \geq 1 - \epsilon$.*

Proof. $1/(1 + \epsilon) = (1 + \epsilon)/(1 + \epsilon) - \epsilon/(1 + \epsilon) \geq (1 + \epsilon)/(1 + \epsilon) - \epsilon = 1 - \epsilon$ since $\epsilon > 0$. □

Lemma A.1.4. *Suppose $a, \alpha \in [0, 1], b \in (0, 1]$ with $a/b \leq 1$, then $\alpha/(b + \alpha) + a/(b + \alpha) \geq a/b$*

Proof. Solving for k in $k + a/(b + \alpha) = a/b$ gives $k = (a/b)\alpha/(b + \alpha)$. Using $a/b \leq 1$ implies $k \leq \alpha/(b + \alpha)$, so $\alpha/(b + \alpha) + a/(b + \alpha) \geq a/b$. □

A.2 Simple Algebraic and Averaging Manipulations

Lemma A.2.1. *Given $\epsilon \in (0, \frac{1}{2})$ for a $[0, 1]$-bounded class $\mathcal{F}$ on (U, σ) and functions $g, h : U \to [0, 1]$:*

1. *If g is $\{0,1\}$-Boolean then $2|g-f| = (1-f^\dagger g^\dagger)$ and $2|g-\bar{f}| = (1+f^\dagger g^\dagger)$.*

   Proof. $(1-f^\dagger g^\dagger) = 1-(2f-1)(2g-1) = 1-(4fg-2f-2g+1) = 2(f+g-2fg) = 2(f(1-g)+g(1-f)) = 2|g-f|$. $2|g-\bar{f}| = (1+f^\dagger g^\dagger)$ follows from the fact that $(\bar{f})^\dagger = -(f^\dagger)$. □

2. *If g is $\{0,1\}$-Boolean then $g^\dagger|g-h| = g-h$.*

   Proof. Since $g=0$ iff $g^\dagger=-1$ and $g=1$ iff $g^\dagger=1$ so $g^\dagger|g-h| = -(1-g)h+g(1-h) = g-h$. □

3. $2|\mathbb{E}_\sigma[f^\dagger(g-h)]| = |\mathbb{E}_\sigma[f^\dagger(g^\dagger-h^\dagger)]|$.

   Proof. $|\mathbb{E}_\sigma[f^\dagger(g^\dagger-h^\dagger)] = \mathbb{E}_\sigma[f^\dagger(2g-1-2h+1)]| = 2|\mathbb{E}_\sigma[f^\dagger(g-h)]|$. □

4. *For every $f^\dagger$ in class $\mathcal{F}^\dagger$, $|\mathbb{E}_\sigma[f^\dagger(g^\dagger-h^\dagger)]| = O(\epsilon)$ iff $|\mathbb{E}_\sigma[f(g-h)]| = O(\epsilon)$ for every f in class $\mathcal{F}$.*

   Proof. We only show the forward direction as the reverse is identical. By 3, $|\mathbb{E}_\sigma[f^\dagger(g^\dagger-h^\dagger)]| = O(\epsilon)$ implies $|\mathbb{E}_\sigma[f^\dagger(g-h)]| = O(\epsilon)$. Now as this holds for every $f^\dagger \in \mathcal{F}^\dagger$ and $\mathbf{1} \in \mathcal{F}^\dagger$, therefore, $|\mathbb{E}_\sigma[g-h]| = O(\epsilon)$. This yields $O(\epsilon) = |\mathbb{E}_\sigma[f^\dagger(g-h)]| \leq |\mathbb{E}_\sigma[2f(g-h)]| + |\mathbb{E}_\sigma[g-h]| = O(\epsilon)$. □

Lemma A.2.2. *Suppose $f: U \to [0,1], g: U \to \{0,1\}$ with $\mathcal{U} = (U,\sigma)$ a finite probability space then:*

1. *For every $x \in U$, $|g(x)-f(x)| = g(x)\bar{f}(x) + \bar{g}(x)f(x)$*

   Proof. By the Shannon decomposition of the Boolean g:

   $$|g-f| = g(x)(1-f(x)) + (1-g(x))f(x) = g(x)\bar{f}(x) + \bar{g}(x)f(x)$$

   □

2. *For every $x \in U$, $\overline{|g(x)-f(x)|} = |\bar{g}(x)-f(x)| = |g(x)-\bar{f}(x)|$.*

Proof. $\overline{|g-f|} = 1-|g-f| = 1-g(1-f)-(1-g)f = 1-g-f+2gf = (1-g)(1-f)+gf = \bar{g}\bar{f}+gf$ implying for every x,

$$\overline{|g(x)-f(x)|} = |\bar{g}(x)-f(x)| = |1-g(x)-f(x)| = |\bar{f}(x)-g(x)|$$

□

3. *For* $\theta \sim \text{UNIFORM}[0,1]$ *and every* $x \in U$, $g(x) = \mathbb{E}_\theta[th_\theta[g(x)]]$.

Proof. For every fixed $x \in U$, $th_\theta[g(x)]$ is a piecewise constant function and this gives:

$$\mathbb{E}_{\theta\in[0,1]}[th_\theta[g(x)]] = \int_0^1 th_\theta[g(x)]dt = \int_0^{g(x)} \mathbf{1}dt + \int_{g(x)}^1 \mathbf{0}dt = g(x)$$

□

4. *For* $\theta \sim \text{UNIFORM}[0,1]$, $\mathbb{E}_\sigma[\mathbb{E}_\theta[|g-th_\theta[f]|]] = \mathbb{E}_\theta[\mathbb{E}_\sigma[|g-th_\theta[f]|]] = \mathbb{E}_\sigma[|g-f|]$.

Proof. By finiteness of U, $\mathbb{E}_\theta[\mathbb{E}_\sigma[|g-th_\theta[f]|]] = \mathbb{E}_\sigma[\mathbb{E}_\theta[|g-th_\theta[f]|]]$. Now for every $x \in U$,

$$\begin{aligned}\mathbb{E}_\theta[|g(x)-th_\theta[f(x)]|] &= \int_0^1 |g(x)-th_\theta[f(x)]|d\theta \\ &= \int_0^{f(x)} |g(x)-1|d\theta + \int_{f(x)}^1 g(x)d\theta \\ &= (1-g(x))f(x)+g(x)(1-f(x)) = |g(x)-f(x)|\end{aligned}$$

Since it holds for all $x \in U$, so $\mathbb{E}_\theta[\mathbb{E}_\sigma[|g-th_\theta[f]|]] = \mathbb{E}_\sigma[|g-f|]$ as well. □

5. *For* $\theta \sim \text{UNIFORM}[0,1]$, $\mathbb{E}_\sigma[\mathbb{E}_\theta[th_\theta[f]] = \mathbb{E}_\theta[\mathbb{E}_\sigma[th_\theta[f]]] = \mathbb{E}_\sigma[f]$.

Proof. This follows as a special case of 4 by setting $g = \mathbf{0}$ and using $|0-f| = f$. Equivalently, it also follows from 3. □

Bibliography

[BHK09] Boaz Barak, Moritz Hardt, and Satyen Kale. The uniform hardcore lemma via approximate Bregman projections. In *SODA*, pages 1193–1200, 2009.

[Blu96] Avrim Blum. On-Line algorithms in Machine Learning. In *In Proceedings of the Workshop on On-Line Algorithms, Dagstuhl*, pages 306–325. Springer, 1996.

[Bre67] Lev M. Bregman. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR Computational Mathematics and Mathematical Physics*, 7(3):200–217, 1967.

[BV07] Andrej Bogdanov and Emanuele Viola. Pseudorandom bits for polynomials. In *In 48th Annual Symposium on Foundations of Computer Science. IEEE*, pages 41–51, 2007.

[CZ97] Yair Censor and Stavros A. Zenios. *Parallel optimization: theory, algorithms, and applications*. Numerical mathematics and scientific computation. Oxford University Press, 1997.

[Fel10] Vitaly Feldman. Distribution-specific agnostic boosting. In *Innovations in Computer Science - ICS 2010*, pages 241–250, 2010.

[FK99] Alan M. Frieze and Ravi Kannan. Quick Approximation to Matrices and Applications. *Combinatorica*, 19(2):175–220, 1999.

[FS99] Yoav Freund and Robert E. Schapire. Adaptive game playing using multiplicative weights. *Games and Economic Behavior*, 29(12):79 – 103, 1999.

[GNW95] Oded Goldreich, Noam Nisan, and Avi Wigderson. On Yao's XOR-lemma. *Electronic Colloquium on Computational Complexity (ECCC)*, 2(50), 1995.

[Gow08] W. Timothy Gowers. Decompositions, approximate structure, transference, and the Hahn-Banach theorem. *ArXiv e-prints*, 2008. arXiv:0811.3103v1 [math.CO], http://arxiv.org/abs/0811.3103v1.

[GT08] Ben Green and Terence Tao. The primes contain arbitrarily long arithmetic progressions. *Ann. of Math.*, 167(2):481–547, 2008.

[Hol05] Thomas Holenstein. Key agreement from weak bit agreement. In *37th Annual ACM Symposium on Theory of Computing*, pages 664–673, 2005.

[Hol06] Thomas Holenstein. *Strengthening Key Agreement using Hard-Core Sets.* PhD thesis, ETH Zurich, May 2006. Reprint as vol. 7 of *ETH Series in Information Security and Cryptography*, ISBN 3-86626-088-2, Hartung-Gorre Verlag, Konstanz, 2006.

[Imp95] Russell Impagliazzo. Hard-core distributions for somewhat hard problems. In *In 36th Annual Symposium on Foundations of Computer Science*, pages 538–545. IEEE, 1995.

[Imp09] Russell Impagliazzo. Algorithmic dense model theorems and weak regularity. Private communication, 2009.

[Kal07] Satyen Kale. Boosting and hard-core set constructions: a simplified approach. *Electronic Colloquium on Computational Complexity (ECCC)*, 14(131), 2007.

[KS99] Adam R. Klivans and Rocco A. Servedio. Boosting and hard-core sets. In *In Proceedings of the Fortieth Annual Symposium on Foundations of Computer Science*, pages 624–633, 1999.

[Lov09] Shachar Lovett. Unconditional pseudorandom generators for low degree polynomials. *Theory of Computing*, 5(1):69–82, 2009.

[LTW07] Chi-Jen Lu, Shi-Chun Tsai, and Hsin-Lung Wu. On the complexity of hard-core set constructions. In *Automata, Languages and Programming, 34th International Colloquium, ICALP 2007*, pages 183–194, 2007.

[NP05] Constantin Niculescu and Lars-Erik Persson. *Convex functions and their applications: A contemporary approach.* Springer, 2005.

[RTTV08a] Omer Reingold, Luca Trevisan, Madhur Tulsiani, and Salil P. Vadhan. Dense subsets of pseudorandom sets. *Electronic Colloquium on Computational Complexity (ECCC)*, 15(045), 2008.

[RTTV08b] Omer Reingold, Luca Trevisan, Madhur Tulsiani, and Salil P. Vadhan. Dense Subsets of Pseudorandom Sets [Extended Abstract]. In *Foundations of Computer*

Science, 2008. FOCS '08. IEEE 49th Annual IEEE Symposium on, pages 76 –85, oct. 2008.

[RTTV08c] Omer Reingold, Luca Trevisan, Madhur Tulsiani, and Salil P. Vadhan. New Proofs of the Green-Tao-Ziegler Dense Model Theorem: An Exposition. *ArXiv e-prints*, June 2008. arXiv:0806.0381v1 [math.CO], http://arxiv.org/abs/0806.0381v1.

[Tao07] Terence Tao. Structure and Randomness in Combinatorics. In *48th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2007)*, pages 3–15, 2007.

[TG06] Terence Tao and Ben Green. An inverse theorem for the Gowers $U^3(G)$ norm. *ArXiv e-prints*, 2006. arXiv:math/0503014v3 [math.NT], http://arxiv.org/abs/math/0503014v3.

[Tre09] Luca Trevisan. Guest column: Additive Combinatorics and Theoretical Computer Science. *SIGACT News*, 40(2):50–66, 2009.

[TTV09] Luca Trevisan, Madhur Tulsiani, and Salil P. Vadhan. Regularity, boosting, and efficiently simulating every high-entropy distribution. In *IEEE Conference on Computational Complexity*, pages 126–136, 2009.

[TZ08] Terence Tao and Tamar Ziegler. The primes contain arbitrarily long polynomial progressions. *Acta Mathematica*, 201:213–305, 2008.

[Vio08] Emanuele Viola. The sum of d small-bias generators fools polynomials of degree d. In *Proceedings of the 2008 IEEE 23rd Annual Conference on Computational Complexity*, CCC '08, pages 124–127, Washington, DC, USA, 2008. IEEE Computer Society.

[VW07] Emanuele Viola and Avi Wigderson. Norms, XOR Lemmas, and Lower Bounds for GF(2) polynomials and multiparty protocols. In *Proceedings of the Twenty-Second Annual IEEE Conference on Computational Complexity*, CCC '07, pages 141–154, Washington, DC, USA, 2007. IEEE Computer Society.

[VZ11] Salil P. Vadhan and Colin Jia Zheng. Characterizing pseudoentropy and simplifying pseudorandom generator constructions. *Electronic Colloquium on Computational Complexity (ECCC)*, 18:141, 2011.

[Wat11] Thomas Watson. Advice lower bounds for the dense model theorem. *Electronic Colloquium on Computational Complexity (ECCC)*, 18:120, 2011.

[Zha11] Jiapeng Zhang. On the query complexity for showing dense model. *Electronic Colloquium on Computational Complexity (ECCC)*, 18:38, 2011.