Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 23 (2010)

Mansour's Conjecture is True for Random DNF Formulas

Adam Klivans
University of Texas at Austin
klivans@cs.utexas.edu

Homin K. Lee
University of Texas at Austin
homin@cs.utexas.edu

Andrew Wan
Columbia University
atw12@cs.columbia.edu

March 4, 2010

Abstract

In 1994, Y. Mansour conjectured that for every DNF formula on n variables with t terms there exists a polynomial p with $t^{O(\log(1/\epsilon))}$ non-zero coefficients such that $\mathbf{E}_{x\in\{0,1\}^n}[(p(x)-f(x))^2] \leq \epsilon$. We make the first progress on this conjecture and show that it is true for several natural subclasses of DNF formulas including randomly chosen DNF formulas and read-k DNF formulas for constant k.

Our result yields the first polynomial-time query algorithm for agnostically learning these subclasses of DNF formulas with respect to the uniform distribution on $\{0,1\}^n$ (for any constant error parameter).

Applying recent work on sandwiching polynomials, our results imply that a $t^{-O(\log 1/\epsilon)}$-biased distribution fools the above subclasses of DNF formulas. This gives pseudorandom generators for these subclasses with shorter seed length than all previous work.

1 Introduction

Let $f : \{0,1\}^n \to \{0,1\}$ be a DNF formula, *i.e.*, a function of the form $T_1 \vee \cdots \vee T_t$ where each T_i is a conjunction of at most n literals. In this paper we are concerned with the following question: how well can a real-valued polynomial p approximate the Boolean function f? This is an important problem in computational learning theory, as real-valued polynomials play a critical role in developing learning algorithms for DNF formulas.

Over the last twenty years, considerable work has gone into finding polynomials p with certain properties (*e.g.*, low-degree, sparse) such that

$$\mathop{\mathbf{E}}_{x\in\{0,1\}^n}[(p(x)-f(x))^2] \leq \epsilon.$$

In 1989, Linial *et al.* [LMN93] were the first to prove that for any t-term DNF formula f, there exists a polynomial $p : \{0,1\}^n \to \{0,1\}$ of degree $O(\log(t/\epsilon)^2)$ such that $\mathbf{E}_{x\in\{0,1\}^n}[(p(x)-f(x))^2] \leq \epsilon$. They showed that this type of approximation implies a quasipolynomial-time algorithm for PAC learning DNF formulas with respect to the uniform distribution. Kalai *et al.* [KKMS08] observed that this fact actually implies something stronger, namely a quasipolynomial-time agnostic learning algorithm for learning DNF formulas (with respect to the uniform distribution). Additionally, the above approximation was used in recent work due to Bazzi [Baz07] and Razborov [Raz08] to show that bounded independence fools DNF formulas.

Three years later, building on the work of Linial *et al.* Mansour [Man95] proved that for any DNF formula with t terms, there exists a polynomial p defined over $\{0,1\}^n$ with *sparsity* $t^{O(\log\log t \log(1/\epsilon))}$ such that $\mathbf{E}_{x\in\{0,1\}^n}[(p(x)-f(x))^2] \leq \epsilon$. By sparsity we mean the number of nonzero coefficients of p. This

ISSN 1433-8092

result implied a nearly polynomial-time *query* algorithm for PAC learning DNF formulas with respect to the uniform distribution.

Mansour conjectured [Man94] that the bound above could improved to $t^{O(\log 1/\epsilon)}$. Such an improvement would imply a polynomial-time query algorithm for learning DNF formulas with respect to the uniform distribution (to within any constant accuracy), and learning DNF formulas in this model was a major open problem at that time.

In a celebrated work from 1994, Jeff Jackson proved that DNF formulas were learnable in polynomial time (with queries, uniform distribution) *without* proving the Mansour conjecture. His "Harmonic Sieve" algorithm [Jac97] used boosting in combination with some weak approximation properties of polynomials. As such, for several years, Mansour's conjecture remained open and attracted considerable interest, but its resolution did not imply any new results in learning theory.

In 2008, Gopalan *et al.* [GKK08b] proved that a positive resolution to the Mansour conjecture also implies an efficient query algorithm for *agnostically* learning DNF formulas (to within any constant error parameter). The agnostic model of learning is a challenging learning scenario that requires the learner to succeed in the presence of adversarial noise. Roughly, Gopalan *et al.* showed that if a class of Boolean functions $\mathcal{C}$ can be ϵ-approximated by polynomials of sparsity s, then there is a query algorithm for agnostically learning $\mathcal{C}$ in time $\mathrm{poly}(s, 1/\epsilon)$ (since decision trees are approximated by sparse polynomials, they obtained the first query algorithm for agnostically learning decision trees with respect to the uniform distribution on $\{0,1\}^n$). Whether DNF formulas can be agnostically learned (queries, uniform distribution) still remains a difficult open problem [GKK08a].

1.1 Our Results

We prove that the Mansour conjecture is true for several well-studied subclasses of DNF formulas. As far as we know, prior to this work, the Mansour conjecture was not known to be true for any interesting class of DNF formulas.

Our first result shows that the Mansour conjecture is true for the class of randomly chosen DNF formulas:

Theorem 1. *Let $f : \{0,1\}^n \to \{0,1\}$ be a DNF formula with t terms where each term is chosen independently from the set of all terms of length $\log t$. Then with probability $1 - n^{\Omega(\log t)}$ (over the choice of the DNF formula). there exists a p with $\|p\|_1 = t^{O(\log 1/\epsilon)}$ such that $\mathbf{E}[(p(x) - f(x))^2] \le \epsilon$.*

By $\|p\|_1$ we mean the sum of the absolute value of the coefficients of p. It is easy to see that this implies that f is ϵ-approximated by a polynomial of sparsity at most $(\|p\|_1/\epsilon)^2$. We choose the terms to be of length $\Theta(\log t)$ so that the expected value of f is bounded away from either 0 or 1.

Our second result is that the Mansour conjecture is true for the class of read-k DNF formulas:

Theorem 2. *Let $f : \{0,1\}^n \to \{0,1\}$ be a DNF formula with t terms where each literal appears at most k times. Then there exists a p with $\|p\|_1 = t^{O(k \log 1/\epsilon)}$ such that $\mathbf{E}[(p(x) - f(x))^2] \le \epsilon$.*

Even for the case $k = 1$, Mansour's conjecture was not known to be true. Mansour [Man95] proves that any polynomial that approximates read-once DNF formulas to ϵ accuracy must have *degree* at least $\Omega(\log t \log(1/\epsilon)/\log\log(1/\epsilon))$. He further shows that a "low-degree" strategy of selecting all of a DNF's Fourier coefficients of monomials up to degree d results in a polynomial p with $\|p\|_1 = t^{O(\log\log t \log 1/\epsilon)}$. It is not clear, however, how to improve this to the desired $t^{O(\log 1/\epsilon)}$ bound.

As mentioned earlier, by applying the result of Gopalan *et al.* [GKK08b], we obtain the first polynomial-time query algorithms for agnostically learning the above classes of DNF formulas to within any constant accuracy parameter. We consider this an important step towards agnostically learning all DNF formulas.

Corollary 3. *Let $\mathcal{C}$ be the class of DNF formulas with t terms where each term is randomly chosen from the set of all terms of length $\log t$. Then there is a query-algorithm for agnostically learning $\mathcal{C}$ with respect to the uniform distribution on $\{0,1\}^n$ to accuracy ϵ in time $\mathrm{poly}(n) \cdot t^{O(\log 1/\epsilon)}$ with probability $1 - n^{\Omega(\log t)}$ (over the choice of the DNF formula).*

We define the notion of agnostic learning with respect to randomly chosen concept classes in Section 2. We also obtain a corresponding agnostic learning algorithm for read-k DNF formulas:

Corollary 4. *Let $\mathcal{C}$ be the class of read-k DNF formulas with t terms. Then there is a query-algorithm for agnostically learning $\mathcal{C}$ with respect to the uniform distribution on $\{0,1\}^n$ to accuracy ϵ in time $\mathrm{poly}(n) \cdot t^{O(k \log 1/\epsilon)}$.*

Our sparse polynomial approximators can also be used in conjunction with recent work due to De. *et al.* to show that for any randomly chosen or read-k DNF f, a $1/t^{O(\log 1/\epsilon)}$-biased distribution fools f (for $k = O(1)$):

Theorem 5. *Let f be a randomly chosen DNF formula or a read-k DNF formula. Then there exists a pseudorandom generator G such that*

$$\left| \Pr_{x \in \{0,1\}^s}[f(G(x)) = 1] - \Pr_{z \in \{0,1\}^n}[f(z) = 1] \right| \leq \epsilon$$

with $s = O(\log n + \log t \cdot \log(1/\epsilon))$.

Previously it was only known that these types of biased distributions fool read-once DNF formulas [DETT09].

1.2 Related Work

As mentioned earlier, Mansour, using the random restriction machinery of Håstad and Linial *et al.* [Hås86, LMN93] had shown that for any DNF formula f, there exists a p of sparsity $t^{O(\log \log t \log 1/\epsilon)}$ that approximates f.

The subclasses of DNF formulas that we show are agnostically learnable have been well-studied in the PAC model of learning. Read-k DNF formulas were shown to be PAC-learnable with respect to the uniform distribution by Hancock and Mansour [HM91], and random DNF formulas were recently shown to be learnable on average with respect to the uniform distribution in the following sequence of work [JS05, JLSW08, Sel08, Sel09].

Recently (and independently) De *et al.* proved that for any read-once DNF formula f, there exists an approximating polynomial p of sparsity $t^{O(\log 1/\epsilon)}$. More specifically, De *et al.* showed that for any class of functions $\mathcal{C}$ fooled by δ-biased sets, there exist sparse, sandwiching polynomials for $\mathcal{C}$ where the sparsity depends on δ. Since they show that $1/t^{O(\log 1/\epsilon)}$-biased sets fool read-once DNF formulas, the existence of a sparse approximator for the read-once case is implicit in their work.

1.3 Our Approach

As stated above, our proof does not analyze the Fourier coefficients of DNF formulas, and our approach is considerably simpler than the random-restriction method taken by Mansour (we consider the lack of Fourier analysis a feature of the proof, given that all previous work on this problem has been Fourier-based). Instead, we use polynomial interpolation.

A Basic Example. Consider a DNF formula $f = T_1 \vee \cdots \vee T_t$ where each T_i is on a disjoint set of exactly $\log t$ variables. For simplicity assume none of the literals are negated. Then the probability that each term is satisfied is $1/t$, and the expected number of satisfied terms is one. Further, since the terms are disjoint, with high probability over the choice of random input, only a few—say d—terms will be satisfied. As such, we construct a univariate polynomial p with $p(0) = 0$ and $p(i) = 1$ for $1 \le i \le d$. Then $p(T_1 + \cdots + T_t)$ will be exactly equal to f as long as at most d terms are satisfied. A careful calculation shows that the inputs where p is incorrect will not contribute too much to $\mathbf{E}[(f-p)^2]$, as there are few of them. Setting parameters appropriately yields a polynomial p that is both sparse and an ϵ-approximator of f.

Random and read-once DNF formulas. More generally, we adopt the following strategy: given a DNF formula f (randomly chosen or read-once) either (1) with sufficiently high probability a random input does not satisfy too many terms of f or (2) f is highly biased. In the former case we can use polynomial interpolation to construct a sparse approximator and in the latter case we can simply use the constant 0 or 1 function.

The probability calculations are a bit delicate, as we must ensure that the probability of many terms being satisfied decays faster than the growth rate of our polynomial approximators. For the case of random DNF formulas, we make use of some recent work due to Jackson *et al.* on learning random monotone DNF formulas [JLSW08].

Read-k DNF formulas. Read-k DNF formulas do not fit into the above dichotomy, so we do not use the sum $T_1 + \cdots + T_t$ inside the univariate polynomial. Instead, we use a sum of *formulas* (rather than terms) based on a construction from [Raz08]. We modify Razborov's construction to exploit the fact that terms in a read-k DNF formula do not share variables with many other terms. Our analysis shows that we can then employ the previous strategy: either (1) with sufficiently high probability a random input does not satisfy too many formulas in the sum or (2) f is highly biased.

2 Preliminaries

In this paper, we will primarily be concerned with Boolean functions $f : \{0,1\}^n \to \{0,1\}$. Let $x_1, \ldots, x_n$ be Boolean variables. A *literal* is either a variable x_i of its negation $\bar{x}_i$, and a *term* is a conjunction of literals. Any Boolean function can be expressed as a disjunction of terms, and such a formula is said to be a *disjunctive normal form* (or DNF) formula. A read-k DNF formula is a DNF formula in which the maximum number of occurrences of each variable is bounded by k.

2.1 Sparse Polynomials

Every function $f : \{0,1\}^n \to \mathbb{R}$ can be expressed by its Fourier expansion: $f(x) = \sum_S \hat{f}(S)\chi_S(x)$ where $\chi_S(x) = \prod_{i \in S}(-1)^{x_i}$ for $S \subseteq [n]$, and $\hat{f}(S) = \mathbf{E}[f \cdot \chi_S]$. The Fourier expansion of f can be thought of as the unique polynomial representation of f over $\{+1,-1\}^n$ under the map $x_i \mapsto \frac{1-x_i}{2}$.

Definition 6. *The* Fourier ℓ_1-norm *(also called the* spectral norm*) of f is defined to be $\|f\|_1 := \sum_S |\hat{f}(S)|$. We will also use the following minor variant, $\|f\|_1^{\neq\emptyset} := \sum_{S \neq \emptyset} |\hat{f}(S)|$.*

We will use the same notation for polynomials $p : \{0,1\}^n \to \mathbb{R}$ (so $\|p\|_1$ here is the sum of the absolute value of the coefficients of the polynomial p), which is justified by the following lemma.

Fact 7. *Given a polynomial $p : \{0,1\}^n \to \mathbb{R}$ with $\|p\|_1 = L$, the spectral norm of p is also L.*

Proof. We can obtain the Fourier representation of p by replacing each term of the polynomial by the Fourier expansion for the term. It is easy to see that the spectral norm of a conjunction is 1. ■

We are interested in the spectral norm of functions because functions with small spectral norm can be approximated by sparse polynomials over $\{+1,-1\}^n$.

Fact 8 ([KM93]). *Given any function with $\mathbf{E}[f^2] \le 1$ and $\epsilon > 0$, let $\mathcal{S} = \{S \subseteq [n] : |\hat{f}(S)| \ge \epsilon/\|f\|_1\}$, and $g(x) = \sum_{S\in\mathcal{S}} \hat{f}(S)\chi_s(x)$. Then $\mathbf{E}[(f-g)^2] \le \epsilon$, and $|\mathcal{S}| \le (\|f\|_1/\epsilon)^2$*

In Mansour's setting, Boolean functions output $+1$ for FALSE and -1 for TRUE. Mansour conjectured that polynomial-size DNF formulas could be approximated by sparse polynomials over $\{+1,-1\}^n$.

Conjecture 9 ([Man94]). *Let $f : \{+1,-1\}^n \to \{+1,-1\}$ be any function computable by a t-term DNF formula. Then there exists a polynomial $p : \{+1,-1\}^n \to \mathbb{R}$ with $t^{O(\log 1/\epsilon)}$ terms such that $\mathbf{E}[(f-p)^2] \le \epsilon$.*

We will prove the conjecture to be true for various subclasses of polynomial-size DNF formulas. In our setting, Boolean functions will outputs 0 for FALSE and 1 for TRUE. However, we can easily change the range by setting $f^{\pm} := 1 - 2\cdot f$. Thus, given a DNF formula $f : \{0,1\}^n \to \{0,1\}$, we will exhibit a polynomial $p : \{0,1\} \to \mathbb{R}$ with $\|p\|_1 = t^{O(\log 1/\epsilon)}$ that approximates f in the ℓ_2-norm to within $\epsilon/8$. Changing the range to $\{+1,-1\}$ will change the accuracy of the approximation by at most a factor of 4: $\mathbf{E}[((1-2f)-(1-2p))]^2 = 4\,\mathbf{E}[(f-p)^2] \le \epsilon/2$, and it will change the ℓ_1-norm of p by at most a factor of 2. Finally, changing the domain of $2p-1$ to $\{+1,-1\}^n$ won't change the ℓ_1-norm of p (Fact 7), and by Fact 8 there is a polynomial with $(2t^{O(\log 1/\epsilon)})^2(2/\epsilon)^2 = t^{O(\log 1/\epsilon)}$ terms over $\{+1,-1\}^n$ that approximates $1-2p$ to within $\epsilon/2$, thus proving Mansour's conjecture.

2.2 Agnostic learning

We first describe the traditional framework for agnostically learning concept classes with respect to the uniform distribution and then give a slightly modified definition for an "average-case" version of agnostic learning where the unknown concept (in this case a DNF formula) is randomly chosen.

Definition 10 (Standard agnostic model). *Let $\mathcal{D}$ be the uniform distribution on $\{0,1\}^n$. Let $f : \{0,1\}^n \to \{0,1\}$ be an arbitrary function. Define*

$$\mathsf{opt} = \min_{c\in\mathcal{C}} \Pr_{x\sim\mathcal{D}}[c(x) \ne f(x)].$$

That is, opt *is the error of the best fitting concept in $\mathcal{C}$ with respect to $\mathcal{D}$. We say that an algorithm A agnostically learns $\mathcal{C}$ with respect to $\mathcal{D}$ if the following holds for any f: if A is given black-box access to f then with high probability A outputs a hypothesis h such that $\Pr_{x\sim\mathcal{D}}[h(x) \ne f(x)] \le \mathsf{opt} + \epsilon$.*

The intuition behind the above definition is that a learner—given access to a concept $c \in \mathcal{C}$ where an η fraction of c's inputs have been adversarially corrupted—should still be able to output a hypothesis with accuracy $\eta + \epsilon$ (achieving error better than η may not be possible, as the adversary could embed a completely random function on an η fraction of c's inputs). Here η plays the role of opt.

This motivates the following definition for agnostically learning a randomly chosen concept from some class $\mathcal{C}$:

Definition 11 (Agnostically learning random concepts)**.** *Let $\mathcal{C}$ be a concept class and choose c randomly from $\mathcal{C}$. We say that an algorithm A agnostically learns random concepts from $\mathcal{C}$ if with probability at least $1-\delta$ over the choice of c the following holds: if the learner is given black-box access to c' and $\Pr_{x\in\{0,1\}^n}[c(x) \neq c'(x)] \leq \eta$, then A outputs a hypothesis h such that $\Pr_{x\in\{0,1\}^n}[h(x) \neq c'(x)] \leq \eta + \epsilon$.*

We are unaware of any prior work defining an agnostic framework for learning randomly chosen concepts.

The main result we use to connect the approximation of DNF formulas by sparse polynomials with agnostic learning is due to Gopalan *et al.* [GKK08b]:

Theorem 12 ([GKK08b])**.** *Let $\mathcal{C}$ be a concept class such that for every $c \in \mathcal{C}$ there exists a polynomial p such that $\|p\|_1 \leq s$ and $\mathbf{E}_{x\in\{+1,-1\}^n}[|p(x)-c(x)|^2] \leq \epsilon^2/2$. Then there exists an algorithm B such that the following holds: given black-box access to any Boolean function $f : \{+1,-1\}^n \to \{+1,-1\}$, B runs in time $\mathrm{poly}(n, s, 1/\epsilon)$ and outputs a hypothesis $h : \{+1,-1\}^n \to \{+1,-1\}$ with*

$$\Pr_{x\in\{+1,-1\}^n}[h(x) \neq f(x)] \leq \mathsf{opt} + \epsilon.$$

3 Approximating DNFs using univariate polynomial interpolation

Let $f = T_1 \vee T_2 \vee \cdots \vee T_t$ be any DNF formula. We say $T_i(x) = 1$ if x satisfies the term T_i, and 0 otherwise. Let $y_f : \{0,1\}^n \to \{0,\ldots,t\}$ be the function that outputs the number of terms of f satisfied by x, *i.e.*, $y_f(x) = T_1(x) + T_2(x) + \cdots + T_t(x)$.

Our constructions will use the following univariate polynomial P_d to interpolate the values of f on inputs $\{x : y_f(x) \leq d\}$.

Fact 13. *Let*

$$P_d(y) := (-1)^{d+1}\frac{(y-1)(y-2)\cdots(y-d)}{d!} + 1. \tag{1}$$

Then, (1) the polynomial P_d is a degree-d polynomial in y; (2) $P_d(0) = 0$, $P_d(y) = 1$ for $y \in [d]$, and for $y \in [t] \setminus [d]$, $P_d(y) = -\binom{y-1}{d} + 1 \leq 0$ if d is even and $P_d(y) = \binom{y-1}{d} + 1 > 1$ if d is odd; and (3) $\|P_d\|_1 = d$.

Proof. Properties (1) and (2) can be easily verified by inspection. Expanding the falling factorial, we get that $(y-1)(y-2)\cdots(y-d) = \sum_{j=0}^{d}(-1)^{d-j}\left[{d+1 \atop j+1}\right]y^j$, where $\left[{a \atop b}\right]$ denotes a Stirling number of the first kind. The Stirling numbers of the first kind count the number of permutations of a elements with b disjoint cycles. Therefore, $\sum_{j=0}^{d}\left[{d+1 \atop j+1}\right] = (d+1)!$ [GKP94]. The constant coefficient of P_d is 0 by Property (2), thus the sum of the absolute values of the other coefficients is $((d+1)! - d!)/d! = d$. ■

For any t-term DNF formula f, we can construct a polynomial $p_{f,d} : \{0,1\}^n \to \mathbb{R}$ defined as $p_{f,d} := P_d \circ y_f$. A simple calculation, given below, shows that the ℓ_1-norm of $p_{f,d}$ is polynomial in t and exponential in d.

Lemma 14. *Let f be a t-term DNF formula, then $\|p_{f,d}\|_1 = t^{O(d)}$.*

Proof. By Fact 13, P_d is a univariate polynomial with d-terms and coefficients of magnitude at most d. We can view the polynomial $p_{f,d}$ as the polynomial $P'_d(T_1,\ldots,T_t) := P_d(T_1 + \cdots + T_t)$ over variables $T_i \in \{0,1\}$. Expanding out (but not recombining!) P'_d gives us at most dt^d terms with coefficients of magnitude at most d. Now each term of P'_d is a product of at most d T_i's, thus $p_{f,d}$ is a polynomial over $\{0,1\}^n$ with at most dt^d terms with coefficients of magnitude at most d, and thus $\|p_{f,d}\|_1 = t^{O(d)}$. ■

The following lemma shows that we can set d to be fairly small, $\Theta(\log 1/\epsilon)$, and the polynomial $p_{f,d}$ will be a good approximation for any DNF formula f, as long as f is unlikely to have many terms satisfied simultaneously.

Lemma 15. *Let f be any DNF formula with $t = \mathrm{poly}(n)$ terms, $d = 4e^3 \ln 1/\epsilon$, and $\ell = (1/8)\log n$. If*

$$\Pr[y_f(x) = j] \leq \left(\frac{e \ln 1/\epsilon}{j}\right)^j,$$

for every $d \leq j \leq \ell$, then the polynomial $p_{f,d}$ satisfies $\mathbf{E}[(f - p_{f,d})^2] \leq \epsilon$.

Proof. We condition on the values of $y_f(x)$, controlling the magnitude of $p_{f,d}$ by the unlikelihood of y_f being large. By Fact 13, $p_{f,d}(x)$ will output 0 if x does not satisfy f, $p_{f,d}(x)$ will output 1 if $y_f(x) \in [d]$, and $|p_{f,d}(x)| < \binom{j}{d}$ for $y_f(x) \in [t] \setminus [d]$. Hence:

$$\begin{aligned}
\|f - p_{f,d}\|^2 &< \sum_{j=d+1}^{\ell-1} \binom{j}{d}^2 \left(\frac{e \ln 1/\epsilon}{j}\right)^j + \binom{t}{d}^2 \cdot \Pr[y_f \geq \ell] \\
&< \sum_{j=d+1}^{\ell-1} 2^{2j} \left(\frac{e \ln 1/\epsilon}{4e^3 \ln 1/\epsilon}\right)^j + n^{-\Omega(\log\log n)} \\
&< \epsilon \sum_{j=d+1}^{\ell-1} \frac{1}{e^j} + n^{-\Omega(\log\log n)} < \epsilon.
\end{aligned}$$

■

For f satisfying the condition in Lemma 15, we may take $d = \Theta(\log 1/\epsilon)$ and apply Lemma 14 to obtain an ϵ-approximating polynomial for f with spectral norm $t^{O(\log 1/\epsilon)}$.

4 Mansour's Conjecture for Random DNF Formulas

In this section, we establish various properties of random DNF formulas and use these properties to show that for almost all f, Mansour's conjecture holds. Roughly speaking, we will show that a random DNF formula behaves like the "tribes" function, in that any "large" set of terms is unlikely to be satisfied by a random assignment. This notion is formalized in Lemma 18. For such DNF formulas, we may use the construction from Section 3 to obtain a good approximating polynomial for f with small spectral norm (Theorem 19).

Throughout the rest of this section we assume that $n^a \leq t(n) \leq n^b$ for any constants $a, b > 0$. For brevity we write t for $t(n)$. Let $\mathcal{D}_n^t$ be the probability distribution over t-term DNF formulas induced by the following process: each term is independently and uniformly chosen at random from all $t\binom{n}{\log t}$ possible terms of size exactly $\log t$ over $\{x_1, \ldots, x_n\}$.

If t grows very slowly relative to n, say $t = O(n^{1/4})$, then with high probability a random f drawn from $\mathcal{D}_n^t$ will be a read-once DNF formula, in which case the results of Section 5 hold. If the terms are not of size $\Theta(\log n)$, then the DNF will be biased, and thus be easy to learn. We refer the reader to [JS05] for a full discussion of the model.

To prove Lemma 18, we require two Lemmas, which are inspired by the results of [JS05] and [JLSW08]. Lemma 16 shows that with high probability the terms of a random DNF formula are close to being disjoint, and thus cover close to $j \log t$ variables.

Lemma 16. *With probability at least* $1 - t^j e^{j\log t}(j\log t)^{\log t}/n^{\log t}$ *over the random draw of* f *from* $\mathcal{D}_n^t$, *at least* $j\log t - (\log t)/4$ *variables occur in every set of* j *distinct terms of* f. *The failure probability is at most* $1/n^{\Omega(\log t)}$ *for any* $j < (\ln(2)/4)\log n$.

Proof. Let $k := \log t$. Fix a set of j terms, and let $v \le jk$ be the number of distinct variables (negated or not) that occur in these terms. We will bound the probability that $v > w := jk - k/4$. Consider any particular fixed set of w variables. The probability that none of the j terms include any variable outside of the w variables is precisely $\left(\binom{w}{k}/\binom{n}{k}\right)^j$. Thus, the probability that $v \le w$ is by the union bound:

$$\binom{n}{w}\left(\frac{\binom{w}{k}}{\binom{n}{k}}\right)^j < \left(\frac{en}{w}\right)^w \left(\frac{w}{n}\right)^{jk} = \frac{e^{jk-k/4}(jk-k/4)^{k/4}}{n^{k/4}} < \frac{e^{jk}(jk)^{k/4}}{n^{k/4}}.$$

Taking a union bound over all (at most t^j) sets, we have that with the correct probability every set of j terms contains at least w distinct variables. ∎

We will use the method of bounded differences (a.k.a., McDiarmid's inequality) to prove Lemma 18.

Proposition 17 (McDiarmid's inequality)**.** *Let* $X_1, \ldots, X_m$ *be independent random variables taking values in a set* $\mathcal{X}$, *and let* $f : \mathcal{X}^m \to \mathbb{R}$ *be such that for all* $i \in [m]$, $|f(a) - f(a')| \le d_i$, *whenever* $a, a' \in \mathcal{X}^m$ *differ in just the* i*th coordinate. Then for all* $\tau > 0$,

$$\Pr\left[f > \mathbf{E}\, f + \tau\right] \le \exp\left(-\frac{2\tau^2}{\sum_i d_i^2}\right) \text{ and } \Pr\left[f < \mathbf{E}\, f - \tau\right] \le \exp\left(-\frac{2\tau^2}{\sum_i d_i^2}\right).$$

The following lemma shows that with high probability over the choice of random DNF formula, the probability that exactly j terms are satisfied is close to that for the "tribes" function: $\binom{t}{j}t^{-j}(1-1/t)^{t-j}$.

Lemma 18. *For any* $j < (\ln(2)/4)\log n$, *with probability at least* $1 - 1/n^{\Omega(\log t)}$ *over the random draw of* f *from* $\mathcal{D}_n^t$, *the probability over the uniform distribution on* $\{0,1\}^n$ *that an input satisfies exactly* j *distinct terms of* f *is at most* $2\binom{t}{j}t^{-j}(1-1/t)^{t-j}$.

Proof. Let $f = T_1 \vee \cdots \vee T_t$, and let $\beta := t^{-j}(1-1/t)^{t-j}$. Fix any $J \subset [t]$ of size j, and let U_J be the probability over $x \in \{0,1\}^n$ that the terms T_i for $i \in J$ are satisfied and no other terms are satisfied. We will show that $U_J < 2\beta$ with high probability; a union bound over all possible sets J of size j in $[t]$ gives that $U_J \le 2\beta$ for every J with high probability. Finally, a union bound over all $\binom{t}{j}$ possible sets of j terms (where the probability is taken over x) proves the lemma.

Without loss of generality, we may assume that $J = [j]$. For any fixed x, we have:

$$\Pr_{f \in \mathcal{D}_n^t}\left[x \text{ satisfies exactly the terms in } J\right] = \beta,$$

and thus by linearity of expectation, we have $\mathbf{E}_{f \in \mathcal{D}_n^t}\left[U_J\right] = \beta$. Now we show that with high probability that the deviation of U_J from its expected value is low.

Applying Lemma 16, we may assume that the terms $T_1, \cdots, T_j$ contain at least $j\log t - (\log t)/4$ many variables, and that $J \cup T_i$ for all $i = j+1, \cdots, t$ includes at least $(j+1)\log t - (\log t)/4$ many unique variables, while increasing the failure probability by only $1/n^{\Omega(\log t)}$. Note that conditioning on this event can change the value of U_J by at most $1/n^{\Omega(\log t)} < \frac{1}{2}\beta$, so under this conditioning we have $\mathbf{E}[P_j] \ge \frac{1}{2}\beta$. Conditioning on this event, fix the terms $T_1, \cdots, T_j$. Then the terms $T_{j+1}, \cdots, T_t$ are chosen uniformly and independently from the set of all terms T of length $\log t$ such that the union of the variables in J and T includes at least $(j+1)\log t - (\log t)/4$ unique variables. Call this set $\mathcal{X}$.

We now use McDiarmid's inequality where the random variables are the terms $T_{j+1}, \ldots, T_t$ randomly selected from $\mathcal{X}$, letting $g(T_{j+1}, \cdots, T_t) = U_J$ and $g(T_{j+1}, \cdots, T_{i-1}, T_i', T_{i+1}, \cdots, T_t) = U_J'$ for all $i = j+1, \ldots, t$. We claim that:

$$\left|U_J - U_J'\right| \le d_i := \frac{t^{1/4}}{t^{j+1}}.$$

This is because U_J' can only be larger than U_J by assignments which satisfy $T_1, \cdots, T_J$ and T_i. Similarly, U_J' can only be smaller than U_J by assignments which satisfy $T_1, \cdots, T_J$ and T_i'. Since T_i and T_i' come from $\mathcal{X}$, we know that at least $(j+1)t - (\log t)/4$ variables must be satisfied.

Thus we may apply McDiarmid's inequality with $\tau = \frac{3}{2}\beta$, which gives that $\Pr_f[U_J > 2\beta]$ is at most

$$\exp\left(\frac{-2\frac{9}{4}\beta^2}{t^{3/2}/t^{2j+2}}\right) \le \exp\left(\frac{-9\sqrt{t}(1-1/t)^{2(t-j)}}{2}\right).$$

Combining the failure probabilities over all the $\binom{t}{j}$ possible sets, we get that with probability at least

$$\binom{t}{j}\left(\frac{1}{n^{\Omega(\log t)}} + e^{-9\sqrt{t}(1-1/t)^{2(t-j)}/2}\right) = \frac{1}{n^{\Omega(\log t)}},$$

over the random draw of f from $\mathcal{D}_n^t$, U_J for all $J \subseteq [t]$ of size j is at most 2β. Thus, the probability that a random input satisfies exactly some j distinct terms of f is at most $2\binom{t}{j}\beta$. ■

Using these properties of random DNF formulas we can now show that Mansour's conjecture [Man94] is true with high probability over the choice of f from $\mathcal{D}_n^t$.

Theorem 19. *Let $f : \{0,1\}^n \to \{0,1\}$ be a t-term DNF formula where each term is chosen independently from the set of all terms of length $\log t$. Then with probability at least $1 - n^{-\Omega(\log t)}$ over the choice of f, there exists a polynomial p with $\|p\|_1 = t^{O(\log 1/\epsilon)}$ such that $\mathbf{E}[(p(x) - f(x))^2] \le \epsilon$.*

Proof. Let $d := 4e^3 \ln(1/\epsilon)$ and $p_{f,d}$ be as defined in Section 3. Lemma 14 tells us that $\|p_{f,d}\|_1 = t^{O(\log 1/\epsilon)}$. We show that with probability at least $1 - n^{-\omega(1)}$ over the random draw of f from $\mathcal{D}_n^t$, $p_{f,d}$ will be a good approximator for f. This follows by Lemma 18; with probability at least $1 - (\ell - d - 1)/n^{\Omega(\log t)} = 1 - n^{-\Omega(\log t)}$, we have $\Pr[y = j]$ for all $d < j \le \ell$. Thus for such f Lemma 15 tells us that $\mathbf{E}[(f - p_{f,d})^2] \le \epsilon$. ■

5 Mansour's Conjecture for Read-k DNF Formulas

In this section, we give an ϵ-approximating polynomial for any read-k DNF formula and show that its spectral norm is at most $t^{O(k \log 1/\epsilon)}$. This implies that Mansour's conjecture holds for all read-k DNF formulas where k is any constant.

We first illustrate the intuition behind the general case by proving Mansour's conjecture for read-once DNF formulas. To do this we show that, like random DNF formulas, read-once DNF formulas satisfy the conditions of Lemma 15, and thus we can use our construction from Section 3.

Read-k DNF formulas, in contrast, may not satisfy the conditions of Lemma 15, so we must change our approach. Instead of using $\sum_{i=1}^t T_i$ inside our univariate polynomial, we use a different sum, which is based on a construction from [Raz08] of representing any DNF formula. We modify this representation to exploit the fact that for read-k DNF formula, the variables in a term can not share variables with too many other terms. The details are given in Section 5.2.

5.1 Read-once DNF Formulas

For a read-once DNF formula, the probability that a term is satisfied is independent of whether or not any of the other terms are satisfied, and we take advantage of this fact.

Lemma 20. *Let $f = T_1 \vee, \cdots, \vee T_t$ be a read-once DNF formula of size t such that $\Pr[f] < 1 - \epsilon$. Then the probability over the uniform distribution on $\{0,1\}^n$ that some set of $j > e \ln 1/\epsilon$ terms is satisfied is at most $\left(\frac{e \ln 1/\epsilon}{j}\right)^j$.*

Proof. For any assignment x to the variables of f, let $y_f(x)$ be the number terms satisfied in f. By linearity of expectation, we have that $\mathbf{E}_x[y_f(x)] = \sum_{i=1}^t \Pr[T_i = 1]$. Note that $\Pr[\neg f] = \prod_{i=1}^t (1 - \Pr[T_i])$, which is maximized when each $\Pr[T_i] = \mathbf{E}[y_f]/t$, hence $\Pr[\neg f] \leq (1 - \mathbf{E}[y_f]/t)^t \leq e^{-\mathbf{E}[y_f]}$. Thus we may assume that $\mathbf{E}[y_f] \leq \ln 1/\epsilon$, otherwise $\Pr[f] \geq 1 - \epsilon$.

Assuming $\mathbf{E}[y_f] \leq \ln 1/\epsilon$, we now bound the probability that some set of $j > e \ln 1/\epsilon$ terms of f is satisfied. Since all the terms are disjoint, this probability is $\sum_{S \subseteq [n], |S| = j} \prod_{i \in S} \Pr[T_i]$, and the arithmetic-geometric mean inequality gives that this is maximized when every $\Pr[T_i] = \mathbf{E}[y_f]/t$. Then the probability of satisfying some set of j terms is at most:

$$\binom{t}{j}\left(\frac{\ln 1/\epsilon}{t}\right)^j \leq \left(\frac{et}{j}\right)^j \left(\frac{\ln 1/\epsilon}{t}\right)^j = \left(\frac{e \ln 1/\epsilon}{j}\right)^j,$$

which concludes the proof of the lemma. ■

Theorem 21. *Let f be any read-once DNF formula with t terms. Then there is a polynomial $p_{f,d}$ with $\|p_{f,d}\|_1 = t^{O(\log 1/\epsilon)}$ and $\mathbf{E}[(f - p_{f,d})^2] \leq \epsilon$ for all $\epsilon > 0$.*

Proof. This follows immediately by combining Lemmas 20, 14, and 15. ■

5.2 Read-k DNF Formulas

Unlike for read-once DNF formulas, it is not clear that the number of terms satisfied in a read-k DNF formula will be extremely concentrated on a small range. In this section we show how to modify our construction so that a concentration result does hold.

Let f be any t-term read-k DNF formula. For any term T_i of f, let ϕ_i be the DNF formula consisting of those terms in $T_1, \cdots, T_{i-1}$ that overlap with T_i, *i.e.*,

$$\phi_i := \bigvee_{j \in \mathcal{C}_i} T_j, \text{ for } \mathcal{C}_i = \{j < i \mid T_j \cap T_i \neq \emptyset\}.$$

We define $A_i := T_i \wedge \neg\phi_i$ and $z_f := \sum_{i=1}^t A_i$. The function $z_f : \{0,1\}^n \rightarrow \{0, \ldots, t\}$ outputs the number of disjoint terms of f satisfied by x (greedily starting from T_1). Note that if f is a read-once DNF formula $z_f = y_f$

As we did in Section 3, we can construct a polynomial $q_{f,d} : \{0,1\}^n \rightarrow \mathbb{R}$ defined as $q_{f,d} := P_d \circ z_f$ for any t-term read-k DNF formula f. The following lemma shows that the spectral norm of $q_{f,d} := P_d \circ z_f$ remains small.

Lemma 22. *Let f be a t-term read-k DNF formula with terms of length at most w. Then $\|q_{f,d}\|_1 = 2^{O(d(\log t + kw))}$.*

Proof. By Fact 13, P_d is a degree-d univariate polynomial with d terms and coefficients of magnitude at most d. We can view the polynomial $q_{f,d}$ as the polynomial $P'_d(A_1, \dots, A_t) := P_d(A_1 + \cdots + A_t)$ over variables $A_i \in \{0,1\}$. Expanding out (but not recombining) P'_d gives us at most dt^d terms with coefficients of magnitude at most d.

Because f is a read-k DNF, each ϕ_i has at most kw terms. We can represent each A_i as $T_i \cdot \prod_{T_j \in \mathcal{C}_i}(1 - T_j)$, which expanded out has at most 2^{kw} terms each with coefficient 1. Now each term of P'_d is a product of at most d A_i's. Therefore $q_{f,d}$ is a polynomial over $\{x_1, \dots, x_n\}$ with at most $2^{kwd}dt^d$ terms with coefficients of magnitude at most d, and $\|q_{f,d}\|_1 = 2^{kwd}dt^d = 2^{O(d(\log t + kw))}$. ∎

In order to show that Mansour's conjecture holds for read-k DNF formulas, we show that $z_f = \sum_{i=1}^t A_i$ behaves much like $y_f = \sum_{i=1}^t T_i$ would if f were a read-once DNF formula, and thus we can use our polynomial P_d (Equation 1) to approximate f. The following claim formalizes the intuition that z_f behaves like y_f would if the $T_i's$ were independent.

Claim 23. *For any $S \subseteq [t]$, $\Pr[\wedge_{i \in S} A_i] \le \prod_{i \in S} \Pr[T_i \mid \neg\phi_i]$.*

Proof. If there is a pair $j, k \in S$ such that $T_j \cap T_k \neq \emptyset$ for some $j < k$, then ϕ_k contains T_j and both $T_j \wedge \neg\phi_j$ and $T_k \wedge \neg\phi_k$ cannot be satisfied simultaneously. Hence, $\Pr[\wedge_{i\in S} A_i] = 0$.

We proceed to bound $\Pr[\wedge_{i\in S} A_i]$ assuming no such pair exists, which means all the terms indexed by S are disjoint. Observe that in this case, the event that T_i is satisfied is independent of all T_j for $i \neq j \in S$. Let T_s be the last term (according to the order the terms appear in f) of S, and condition on the event that A_s is satisfied:

$$\Pr[\wedge_{i\in S} A_i] = \Pr[\wedge_{i \in S\setminus\{s\}} A_i \mid T_s \wedge \neg\phi_s] \Pr[T_s \mid \neg\phi_s] \Pr[\neg\phi_s]. \tag{2}$$

Now, consider any $T_i \wedge \neg\phi_i$ for $i \in S \setminus \{s\}$. The term T_i is independent of T_s, and any term in ϕ_i that overlaps with T_s is also contained in ϕ_s. Hence, conditioning on A_s is equivalent to conditioning on $\neg\phi_s$, *i.e.,*

$$\Pr[\wedge_{i \in S\setminus\{s\}} A_i \mid A_s] = \Pr[\wedge_{i \in S\setminus\{s\}} A_i \mid \neg\phi_s].$$

Substituting this into Equation 2 and observing that $\Pr[\wedge_{i \in S\setminus\{s\}} A_i \mid \neg\phi_s] \Pr[\neg\phi_s] \le \Pr[\wedge_{i \in S\setminus\{s\}} A_i]$, we have:

$$\Pr[\wedge_{i\in S} A_i] \le \Pr[\wedge_{i \in S\setminus\{s\}} A_i] \Pr[T_s \mid \neg\phi_s].$$

Repeating this argument for the remaining terms in $S \setminus s$ yields the claim. ∎

Using Claim 23, we can prove a lemma analogous to Lemma 20 by a case analysis of $\sum_{i=1}^t \Pr[T_i \mid \neg\phi_i]$; either it is large and f must be biased toward one, or it is small so z_f is usually small.

Lemma 24. *Let $f = T_1 \vee, \cdots, \vee T_t$ be a read-k DNF formula of size t such that $\Pr[f] < 1 - \epsilon$. Then the probability over the uniform distribution on $\{0,1\}^n$ that $z_f \ge j$ (for any $j > e \ln 1/\epsilon$) is at most $\left(\frac{e \ln 1/\epsilon}{j}\right)^j$.*

Proof. First, we show that if $T_A := \sum_{i=1}^t \Pr[T_i | \neg\phi_i] > \ln 1/\epsilon$, then $\Pr[f] > 1 - \epsilon$. To do this, we bound $\Pr[\neg f]$ using the fact that the $T_i's$ are independent of any terms $T_j \notin \phi_i$ with $j < i$:

$$\Pr[\neg f] = \Pr\left[\wedge_{i=1}^t \neg T_i\right] = \prod_{i=1}^t \left(1 - \Pr\left[T_i | \wedge_{j=1}^{i-1} \neg T_j\right]\right) = \prod_{i=1}^t \left(1 - \Pr\left[T_i | \neg\phi_i\right]\right).$$

This quantity is maximized when each $\Pr[T_i \mid \neg\phi_i] = T_A/t$, hence:

$$\Pr[\neg f] < \left(1 - \frac{\ln 1/\epsilon}{t}\right)^t \leq \epsilon.$$

We now proceed assuming that $T_A \leq \ln 1/\epsilon$. The probability that some set of j A_i's is satisfied is at most $\sum_{S \subseteq [t], |S|=j} \Pr[\wedge_{i \in S} A_i]$. Applying Claim 23, we have:

$$\sum_{S \subseteq [t], |S|=j} \Pr[\wedge_{i \in S} A_i] \leq \sum_{S \subseteq [t], |S|=j} \prod_{i \in S} \Pr[T_i \mid \neg\phi_i].$$

The arithmetic-geometric mean inequality shows that this quantity is maximized when all $\Pr[T_i \mid \neg\phi_i]$ are equal, hence:

$$\sum_{S \subseteq [t], |S|=j} \prod_{i \in S} \Pr[T_i | \neg\phi_i] \leq \binom{t}{j} \left(\frac{T_A}{t}\right)^j \leq \left(\frac{eT_A}{j}\right)^j \leq \left(\frac{e \ln 1/\epsilon}{j}\right)^j$$

■

We can now show that Mansour's conjecture holds for read-k DNF formulas with any constant k.

Theorem 25. *Let $f : \{0,1\}^n \to \{0,1\}$ be any read-k DNF formula with t terms. Then there is a polynomial $q_{f,d}$ with $\|q_{f,d}\|_1 = t^{O(k \log 1/\epsilon)}$ and $\mathbf{E}[(f - q_{f,d})^2] \leq \epsilon$ for all $\epsilon > 0$.*

Proof. If $\Pr[f = 1] > 1 - \epsilon$, the constant 1 is a suitable polynomial. Let g be the DNF formula f after dropping terms of length greater than $w := \log(2t/\epsilon)$. (This only changes the probability by $\epsilon/2$.) Let $d := 4e^3 \ln(2/\epsilon)$ and $q_{g,d}$ be as defined at the beginning of Section 5.2. Lemma 22 tells us that $\|q_{g,d}\|_1 = t^{O(kd)}$, and Lemma 24 combined with Lemma 15 tells us that $\mathbf{E}[(g - q_{g,d})^2] \leq \epsilon/2$. ■

6 Pseudorandomness

De *et al.* [DETT09] recently improved long-standing pseudorandom generators against DNF formulas.

Definition 26. *A probability distribution X over $\{0,1\}^n$* ϵ-fools *a real function $f : \{0,1\}^n \to \mathbb{R}$ if*

$$|\mathbf{E}[f(X)] - \mathbf{E}[f(U_n)]| \leq \epsilon.$$

If $\mathcal{C}$ is a class of functions, then we say that X ϵ-fools $\mathcal{C}$ if X ϵ-fools every function $f \in \mathcal{C}$.

We say a probability distribution X over $\{0,1\}^n$ is ϵ-biased *if it ϵ-fools the character function χ_S for every $S \subseteq [n]$.*

De *et al.* observed that the result of Bazzi [Baz07] implied a pseudorandom generator that ϵ-fools t-term DNF formulas over n variables with seed length $O(\log n \cdot \log^2(t/\beta))$, which already improves the long-standing upper bound of $O(\log^4(tn/\epsilon))$ of Luby *et al.* [LVW93]. They go on to show a pseudorandom generator with seed length $O(\log n + \log^2(t/\epsilon) \log\log(t/\epsilon))$.

They prove that a sufficient condition for a function f to be ϵ-fooled by an ϵ-biased distribution is that the function be "sandwiched" between two bounded real-valued functions whose Fourier transform has small ℓ_1 norm:

Lemma 27 (Sandwich Bound [DETT09]). *Suppose $f, f_\ell, f_u : \{0,1\}^n \to \mathbb{R}$ are three functions such that for every $x \in \{0,1\}^n$, $f_\ell(x) \le f(x) \le f_u(x)$, $\mathbf{E}[f_u(U_n)] - \mathbf{E}[f(U_n)] \le \epsilon$, and $\mathbf{E}[f(U_n)] - \mathbf{E}[f_\ell(U_n)] \le \epsilon$. Let $L = \max(\|f_\ell\|_1^{\neq\emptyset}, \|f_u\|_1^{\neq\emptyset})$. Then any β-biased probability distribution $(\epsilon + \beta L)$-fools f.*

Naor and Naor [NN93] prove that an ϵ-biased distribution over n bits can be sampled using a seed of $O(\log(n/\epsilon))$ bits. Using our construction from Section 4, we show that random DNF formulas are ϵ-fooled by a pseudorandom generator with seed length $O(\log n + \log(t)\log(1/\epsilon))$:

Theorem 28. *Let $f = T_1 \vee \cdots \vee T_t$ be a random DNF formula chosen from $\mathcal{D}_n^t$. For $1 \le d \le t$, with probability $1 - 1/n^{\Omega(\log t)}$ over the choice of f, β-biased distributions $O(2^{-\Omega(d)} + \beta t^d)$-fool f. In particular, we can ϵ-fool most $f \in \mathcal{D}_n^t$ by a $t^{-O(\log(1/\epsilon))}$-biased distribution.*

Proof. Let d^+ be the first odd integer greater than d, and let d^- be the first even integer greater than d. Let $f_u = p_{f,d^+}$ and $f_\ell = p_{f,d^-}$ (where $p_{f,d}$ is defined as in Section 3). By Lemma 14, the ℓ_1-norms of f_u and f_ℓ are $t^{O(d)}$. By Fact 13, we know that $f_u(y) = \binom{y-1}{d} + 1 > 1$ and $f_\ell(y) = -\binom{y-1}{d} + 1 \le 0$ for $y \in [t] \setminus [d]$, hence:

$$\mathbf{E}[f_u(U_n)] - \mathbf{E}[f(U_n)] = \sum_{j=d+1}^{t} \left(\binom{j-1}{d} + 1 - 1 \right) \Pr[y = j],$$

which with probability $1 - 1/n^{\Omega(\log t)}$ over the choice of f is at most $2^{-\Omega(d)}$ by the analysis in Lemma 15. The same analysis applies for f_ℓ, thus applying Lemma 27 gives us the theorem. ∎

De *et al.* match our bound for random DNF formulas for the special case of read-once DNF formulas. Using our construction from Section 5 and a similar proof as the one above, we can show that read-k formulas are ϵ-fooled by a pseudorandom generator with seed length $O(\log n + \log(t)\log(1/\epsilon))$.

Theorem 29. *Let $f = T_1 \vee \cdots \vee T_t$ be a read-k DNF formula for constant k. For $1 \le d \le t$, β-biased distributions $O(2^{-\Omega(d)} + \beta t^d)$-fool f. In particular, we can ϵ-fool read-k DNF formulas by a $t^{-O(\log(1/\epsilon))}$-biased distribution.*

7 Discussion

On the relationship between Mansour's Conjecture and the Entropy-Influence Conjecture. As a final note, we would like to make a remark on the relationship between Mansour's conjecture and the entropy-influence conjecture. The *spectral entropy* of a function is defined to be $E(f) := \sum_S -\hat{f}(S)^2 \log(\hat{f}(S)^2)$ and the *total influence* to be $I(f) := \sum_S |S|\hat{f}(S)^2$. The *entropy-influence* conjecture is that $E(f) = O(I(f))$ [FK96].[1] Boppana showed that the total influence of t-term DNF formulas is $O(\log t)$ [Bop97]. From this it follows that Mansour's conjecture is implied by the entropy-influence conjecture.

It can be shown that for $\mathrm{poly}(n)$-size DNF formulas Mansour's conjecture implies an upper bound on the spectral entropy of $O(\log n)$. Thus, for the class of DNF formulas we consider in Section 4 (which have total influence $\Omega(\log n)$), our results imply that the entropy-influence conjecture is true.

Acknowledgments. Thanks to Sasha Sherstov for important contributions at an early stage of this work, and Omid Etesami for pointing out an error in an earlier version of this paper.

[1] http://terrytao.wordpress.com/2007/08/16/gil-kalai-the-entropyinfluence-conjecture/

References

[Baz07] Louay Bazzi. Polylogarithmic independence can fool DNF formulas. In *Proc. 48th IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 63–73, 2007.

[Bop97] Ravi B. Boppana. The average sensitivity of bounded-depth circuits. *Information Processing Letters*, 63(5):257–261, 1997.

[DETT09] Anindya De, Omid Etesami, Luca Trevisan, and Madhur Tulsiani. Improved pseudorandom generators for depth 2 circuits. Technical Report 141, Electronic Colloquium on Computational Complexity (ECCC), 2009.

[FK96] Ehud Friedgut and Gil Kalai. Every monotone graph property has a sharp threshold. *Proceedings of the American Mathematical Society*, 124(10), 1996.

[GKK08a] Parikshit Gopalan, Adam Kalai, and Adam R. Klivans. A query algorithm for agnostically learning DNF? In *21st Annual Conference on Learning Theory - COLT 2008, Helsinki, Finland, July 9-12, 2008*, pages 515–516. Omnipress, 2008.

[GKK08b] Parikshit Gopalan, Adam Tauman Kalai, and Adam R. Klivans. Agnostically learning decision trees. In *Proceedings of the 40th Annual ACM Symposium on Theory of Computing, Victoria, British Columbia, Canada, May 17-20, 2008*, pages 527–536. ACM, 2008.

[GKP94] Ronald L. Graham, Donald E. Knuth, and Oren Patashnik. *Concrete Mathematics: A Foundation for Computer Science*. Addison-Wesley, 1994.

[Hås86] Johan Håstad. *Computational Limitations for Small Depth Circuits*. MIT Press, 1986.

[HM91] Thomas Hancock and Yishay Mansour. Learning monotone k-μ DNF formulas on product distributions. In *Proc. of the 4th Annual Conference on Computational Learning Theory (COLT)*, pages 179–183, 1991.

[Jac97] Jeffrey C. Jackson. An efficient membership-query algorithm for learning DNF with respect to the uniform distribution. *Journal of Computer and System Sciences*, 55(3):414–440, 1997.

[JLSW08] Jeffrey C. Jackson, Homin K. Lee, Rocco A. Servedio, and Andrew Wan. Learning random monotone DNF. In *11th International Workshop on Approximation Algorithms for Combinatorial Optimization Problems and 12th International Workshop on Randomization and Computation (RANDOM-APPROX)*, pages 483–497. Springer-Verlag, 2008.

[JS05] Jeffrey C. Jackson and Rocco A. Servedio. On learning random DNF formulas under the uniform distribution. In *8th International Workshop on Approximation Algorithms for Combinatorial Optimization Problems and 9th International Workshop on Randomization and Computation (RANDOM-APPROX)*, volume 3624 of *Lecture Notes in Computer Science*, pages 342–353. Springer-Verlag, 2005.

[KKMS08] A. Kalai, A. Klivans, Y. Mansour, and R. Servedio. Agnostically learning halfspaces. *SIAM Journal on Computing*, 37(6):1777–1805, 2008.

[KM93] Eyal Kushilevitz and Yishay Mansour. Learning decision trees using the Fourier spectrum. *SIAM Journal on Computing*, 22(6):1331–1348, December 1993. Prelim. ver. in *Proc. of STOC'91*.

[LMN93] N. Linial, Y. Mansour, and N. Nisan. Constant depth circuits, Fourier transform and learnability. *Journal of the ACM*, 40(3):607–620, 1993.

[LVW93] Michael Luby, Boban Velickovic, and Avi Wigderson. Deterministic approximate counting of depth-2 circuits. In *ISTCS 1993*, pages 18–24, 1993.

[Man94] Y. Mansour. *Learning Boolean functions via the Fourier transform*, pages 391–424. Kluwer Academic Publishers, 1994.

[Man95] Yishay Mansour. An $O(n^{\log \log n})$ learning algorithm for DNF under the uniform distribution. *Journal of Computer and System Sciences*, 50:543–550, 1995. Prelim. ver. in *Proc. of COLT'92*.

[NN93] Joseph Naor and Moni Naor. Small-bias probability spaces: Efficient constructions and applications. *SIAM Journal on Computing*, 22(4):838–856, 1993.

[Raz08] A. Razborov. A simple proof of Bazzi's theorem. Technical Report 81, Electronic Colloquium on Computational Complexity (ECCC), 2008.

[Sel08] Linda Sellie. Learning random monotone DNF under the uniform distribution. In *Proc. of the 21th Annual Conference on Computational Learning Theory (COLT)*, pages 181–192, 2008.

[Sel09] Linda Sellie. Exact learning of random DNF over the uniform distribution. In *Proc. 41st Annual ACM Symposium on Theory of Computing (STOC)*, pages 45–54, 2009.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de