A slight sharpening of LMN

Johan Håstad [*]
Royal Institute of Technology, Stockholm
email:johanh@nada.kth.se

February 18, 2002

Running head: Sharpening of LMN

[*] Work done while visiting Institute for Advanced Study, supported by NSF-grant CCR-9987077.

Abstract

Linial, Mansour and Nisan [7] proved that a function computed by a small-depth circuit of limited size has most of its Fourier support on small sets. We improve their bounds. When the bottom fanin is bounded we use essentially their argument, but to reduce the general case to this case without a loss in the asymptotic bounds requires a new argument.

Keywords: Constant depth circuits, discrete Fourier Transform, learnability.

1 Introduction

The computational class of small-depth circuits has the almost unique property that we are able to prove non-conditional lower bounds, i.e. to establish that explicit functions do require large resources to be computed in this model.

One of the two main methods, initiated by Furst, Saxe and Sipser [4] is to assign random values to some of the inputs and see how much this simplifies the circuit. The procedure of doing this is now usually called "hitting the circuit with a random restriction". One useful tool for analyzing this process is given by the switching lemma [5] which says that, if parameters are suitable, such a restriction can enable you to switch a depth two circuit from CNF to DNF without getting a huge blowup.

Constant depth circuits are also studied from a learning point of view and a key result here is by Linial, Mansour, and Nisan [7] that proved that such a function can be learned fairly efficiently through the Fourier transform. The key technical result is that a function computed by a small-depth circuit of limited size has most of it's Fourier coefficients concentrated on sets of small size. To be more exact, if the circuit is of depth d and size 2^s, Linial et al. prove that the fraction of the Fourier mass that lies on sets of size at least t is bounded by $2^{s-\Omega(t^{1/d})}$. Although their analysis is rather tight, it does leave some room for improvement and the purpose of the current paper is to give such an improvement.

Our improvement is partly due to a more careful application of the switching lemma but this is a rather simple argument and one could say that it appears implicitly in previous papers. In particular, Mansour [8] uses this argument when the depth of the circuit is 2. The more difficult part of the improvement is to reduce the general case to the case when all gates at the bottom level have small fanin.

In a Boolean circuit of size 2^s it is usually not very productive to have gates on the bottom level that has fanin much larger than s. Let us consider gates of fanin at least $2s$ and see why this is the case. Assume that the gate in question is a logical and-gate. Such a gate would be false except for a fraction 2^{-2s} of the inputs and thus even given 2^s such gates, replacing them all by "false" would only change the value of the circuit on at most a fraction 2^{-s} of the inputs. For results that are robust under such perturbations one can hence assume that the bottom fanin is bounded by $2s$. In our application the results are only robust under such perturbations when s large and for small values of s the argument is not so simple and the main technical contribution is to establish that indeed the general case can be reduced to the case of bottom fanin $O(s)$.

An outline of the paper is as follows. In Section 2 we give the necessary background information, in Section 3 we give the proof for the case of small bottom fanin and in Section 4 we show how to reduce the general case to this case. Finally, in Section 5, we briefly outline the construction of circuits of small depth that, as far as we are aware, give the the strongest known correlation with parity.

2 Preliminaries

We are interested in Boolean circuits and thus we have $\wedge$-gates, $\vee$-gates and negations.

We can, using de-Morgan laws push all the negations to the inputs. If we for each gate in the original circuit introduce two gates, one computing the original value and one that computes its complement we can see that this at most doubles the size of the circuit. Since our bounds ignore multiplicative constants, we can thus assume that the circuit contains no internal negations.

The $\wedge$-gates and the $\vee$-gates are of unbounded fanin and by collapsing adjacent gates of the same type, and introducing gates of fanin 1, we can assume that the circuit consists of alternating levels of $\wedge$-gates and $\vee$-gates. We assume that the circuit is of constant depth d and hence this operation only increases the size of the circuit by at most a factor d.

Our main tool for analyzing small-depth circuits is the concept of random restrictions.

2.1 Restrictions

A restriction is a partial assignment to the inputs. A typical restriction is denoted by ρ and for a variable, x we let $\rho(x) = 1(0)$ denote that fact that x is given the value 1(0) and $\rho(x) = *$ is used to denote the fact that x is not given a value and remains a variable.

It turns out to be very useful to consider random restrictions and an important space of random restrictions was introduced by Furst, Saxe and Sipser [4].

Definition 2.1 *A restriction $\rho \in R_p$ assigns values independently to variables. It gives each variable the value $*$ with probability p and the values 0 and 1 with probability $\frac{1-p}{2}$ each.*

For a function f, $f\lceil_\rho$ denotes the function induced by making the substitutions given by ρ. If a circuit C computes f then $f\lceil_\rho$ is computed by a circuit obtained form C by making the substitutions described by ρ. We denote this circuit by $C\lceil_\rho$. At the same time as doing these substitutions we are also interested in doing some simplifications of the resulting circuit and thus $C\lceil_\rho$ is not uniquely defined, but we are content to let it be some circuit computing $f\lceil_\rho$ and hopefully no confusion arises. We have the following basic result.

Lemma 2.2 *[5] Let C be a depth 2 circuit of bottom fanin s. Then the probability that $C\lceil_\rho$ cannot be written as a decision tree of depth t is at most $(5ps)^t$.*

The lemma is usually stated in the form that $C\lceil_\rho$ can be represented as depth 2 circuit of the other type, i.e. converting an or-of-ands to an and-of-ors and the other way around. This statement clearly follows from the current version of the lemma since a function that can be represented as a decision tree of depth t can be represented as a depth 2 bottom fanin t circuit of either type. The stronger version of the lemma follows from the proof of [5], and was used, but not stated explicitly, in [6]. It is stated explicitly in many places such as [3].

Another consequence of Lemma 2.2 is that $C\lceil_\rho$ can be written as a polynomial over the reals of degree t with the same probability. In general we let $deg(f)$ denote the real degree of f.

2.2 Fourier transforms

The discrete Fourier transform is extremely useful for analyzing Boolean functions. Let f be a function mapping $\{0,1\}^n$ into the real numbers. For a subset $\alpha \subseteq [n]$ we define

$$\hat{f}_\alpha = 2^{-n} \sum_x f(x)\chi_\alpha(x).$$

The functions χ_α are the characters defined by

$$\chi_\alpha(x) = (-1)^{(\alpha,x)}$$

where (α, x) is the inner-product. We have the inversion formula

$$f(x) = \sum_\alpha \hat{f}_\alpha \chi_\alpha(x)$$

and Plancherel's equality

$$2^{-n} \sum_x f(x)^2 = \sum_\alpha \hat{f}_\alpha^2.$$

We are interested in the case when f is a Boolean function and we assume that f takes the values ± 1 and hence Plancherel's equality states that

$$\sum_\alpha \hat{f}_\alpha^2 = 1.$$

It is known already from [4] that constant-depth circuits of polynomial size could not compute parity exactly. Since the parity function equals $\chi_{[n]}$, it is not surprising that one can get strong information about the Fourier transform of functions computed by small-depth circuits of limited size. For a function f let us define

$$F(f,t) = \sum_{|\alpha| \geq t} \hat{f}_\alpha^2.$$

Linial, Mansour, and Nisan [7] proved the following theorem.

Theorem 2.3 *[7] If f is computed by a depth d circuits of size 2^s, then $F(f,t) \leq 2^{s+1-t^{1/d}/20}$.*

The main tool for proving this theorem is through analyzing the effect of restrictions on the Fourier transform. We have the following lemma, which is a combination of Lemma 5 and Lemma 6 of [7].

Lemma 2.4 *[7] Let t be an integer and p such that $pt > 8$, then*

$$F(f,t) \leq 2Pr[deg(f\lceil_\rho) \geq pt/2],$$

where ρ is a random restriction from R_p.

3 Small bottom fanin

In this section we prove the following theorem.

Theorem 3.1 *If the function f is computed by a depth d circuit with at most 2^s gates at distance at least 2 from the input and with bottom fanin bounded by s then*

1. *If $d \geq 2$ and $t \leq s^d$ then $F(f,t) \leq 2^{-\Omega(t/s^{d-1})}$.*
2. *If $d \geq 2$ and $t > s^d$ then $F(f,t) \leq 2^{-\Omega((t/s)^{1/(d-1)})}$.*

Remark 3.2 *The improvement over Theorem 2.3 is almost like replacing d by $d-1$ and hence it is largest for small d. The case when $d = 2$ is contained in the paper by Mansour [8]. Mansour attributes this result to [7] but we have not been able to find the exact statement in [7] although it follows by examining the proofs more closely, and making some optimizations.*

Proof: The case of $d = 2$ is straightforward. Simply hit the circuit with a restriction from R_p with $p = 1/(10s)$. The theorem follows from a combination of Lemma 2.4 and Lemma 2.2.

For the general case we proceed as follows. Define

$$r = 2\max(s, (t/s)^{1/(d-1)}).$$

First apply a random restriction from R_p with $p = 1/(10s)$ and then $d-2$ restrictions with $p = 1/(10r)$. We apply Lemma 2.2 to any circuit that a appears as a depth 2 subcircuit of the restrictions of our initial circuit. The hope is that the first $d-2$ restrictions enables us to convert any such circuit to a depth 2 circuit of the other type with bottom fanin at most r. Each such conversion creates two adjacent levels of gates of the same type which can be collapsed decreasing the depth of the circuit by 1. If this is successful for all subcircuits under the first $d-2$ restrictions the resulting circuit is of depth 2. The probability of the conversion failing for some depth 2 circuit is at most

$$2^s \cdot 2^{-r} \leq 2^{-r/2}.$$

If the conversion is successful we apply Lemma 2.2 once more and see that, for any k, except with probability 2^{-k} the resulting function is now of degree at most k. Thus the probability that f after the $d-1$ restrictions is of degree larger than k is bounded by

$$2^{-r/2} + 2^{-k}.$$

The combination of all the restrictions yield a restriction R_q with $q = \Omega(s^{-1}r^{2-d})$. We now apply Lemma 2.4. If $s^d > t$ we have $r = 2s$ and $k = \Omega(ts^{1-d})$ gives the result. If $s^d < t$ then $r = 2(t/s)^{1/(d-1)}$ and $k = \Omega(r)$ gives the desired result. ■

4 Unrestricted bottom fanin

We want to deal with the case where there is no bound on the bottom fanin of the circuit. In [7] such a circuit is transformed into a circuit with bounded bottom fanin by applying a restriction with $p = 1/10$. This causes a deterioration of the bounds and we avoid this by creating this first part of the restriction deterministically and greedily. The proof is inspired by the inapproximability results for parity by Boppana and Håstad given in [6].

The intuition of the approach is straightforward; we set the variables that appear in many large clauses. In the current situation we have two parameters to balance, the number of large clauses and the total mass of the Fourier coefficients on large coefficients. We have to decrease the former more than the latter and we are only able to maintain this balance in a probabilistic setting. The problem being that even though x_i appears in many large clauses, fixing it to one of its two values might not erase a single large clause. In the long run, we cannot be unlucky too many times and this is sufficient to establish the theorem.

Theorem 4.1 *If the function f is computed by a depth d circuit of size 2^s then*

1. *If $d \geq 2$ and $t \leq s^d$ then $F(f,t) \leq 2^{-\Omega(t/s^{d-1})}$.*
2. *If $d \geq 2$ and $t > s^d$ then $F(f,t) \leq 2^{-\Omega((t/s)^{1/(d-1)})}$.*

Proof: As outlined above we give values to some variables trying to eliminate gates with large fanin at the bottom. Let k be a real number and define

$$c(f,k) = \sum_{|\alpha| \geq t(1+\frac{k}{2s})} \hat{f}_\alpha^2 \min\left(|\alpha| - t(1+\frac{k}{2s}), 2t - t(1+\frac{k}{2s})\right). \tag{1}$$

One natural approach is to apply induction to prove $c(f,m)$ is small for all f computed by a circuit of depth d and size at most 2^s with at most 2^m gates at the bottom level of fanin at least $32s$. We were not able to follow this path and we use a slightly more complicated approach.

Let d_0 be the value obtained from Theorem 3.1 with values t and $32s$. We prove that

$$c(f,s) \leq td_0 + 2^{-\Omega(t)} \tag{2}$$

by analyzing a binary tree of functions obtained by fixing values of inputs to f. This is sufficient to establish the theorem since

$$F(f,2t) \leq \frac{2}{t}c(f,s) \leq 2d_0 + 2^{-\Omega(t)}$$

and, adjusting the constants, Theorem 4.1 follows.

Each node in our binary tree is associated with a function g (which is obtained from f by fixing some variables) and the two sons of a node labeled by g is found by fixing a suitable variable to both its values. Call the two resulting functions by g^0 and g^1. A node labeled by g in the tree is given a value of the form $c(g,m)$ where m is an estimate for the logarithm of the number of gates at the bottom level with fanin at least $32s$. We prove that the average value of the two sons of each node is at least the value of the father.

The function labeling the root is f and the value of the root $c(f,s)$. For most leaves it turns out that we can use Theorem 3.1 and for the rest we can use a trivial estimate.

Take a node in our tree labeled by a function g and with value $c(g,m)$. Since g has been obtained from f by substituting constants for some variables it is defined by a depth d circuit of size at most 2^s. If g has no gates of fanin $\geq 32s$ at the bottom level or $m \leq 0$ we stop and the node becomes a leaf. Otherwise we proceed as follows.

Define

$$G = \sum_{2t \geq |\alpha| \geq t(1+\frac{m}{2s})} \hat{g}_\alpha^2,$$

and

$$G_i = \sum_{2t \geq |\alpha| \geq t(1+\frac{m}{2s}) \wedge i \in \alpha} \hat{g}_\alpha^2.$$

Clearly

$$\sum_{i=1}^{n} G_i \leq 2tG, \tag{3}$$

and also $G_i \leq G$.

For each i let l_i be the fraction of the bottom gates of fanin at least $32s$ of the circuit defining g that contains the variable x_i, with or without negation. Clearly

$$\sum_i l_i \geq 32s. \tag{4}$$

By (3) and (4) it is possible to find an i such that

$$l_i \geq \frac{16sG_i}{tG}. \tag{5}$$

Let us study the two functions obtained by setting x_i to 0 and 1 respectively. These are the functions g^0 and g^1 described above that we use to label the sons in the tree. The values that label the corresponding nodes are given by (g^0, m') and (g^1, m') where

$$m' = m - \frac{4sG_i}{tG}. \tag{6}$$

We have two cases depending on whether $G_i > G/2$. Suppose first that $G_i > G/2$. In this case we have

$$\frac{t(m-m')}{2s} \geq 1 \tag{7}$$

Now look at the Fourier coefficients $\hat{g}_\alpha^0$ and $\hat{g}_\alpha^1$ for some fixed α. Let $\alpha' = \alpha \cup \{i\}$. By inspection we have

$$\hat{g}_\alpha^0 = \hat{g}_\alpha + \hat{g}_{\alpha'}$$

and

$$\hat{g}_\alpha^1 = \hat{g}_\alpha - \hat{g}_{\alpha'}$$

and hence

$$(\hat{g}^0_\alpha)^2 + (\hat{g}^1_\alpha)^2 = 2(\hat{g}^2_\alpha + \hat{g}^2_{\alpha'}). \tag{8}$$

Now because of (7) we see that

$$\begin{aligned}((\hat{g}^0_\alpha)^2 + (\hat{g}^1_\alpha)^2) \min\left(|\alpha| - t(1+\frac{m'}{2s}), 2t - t(1+\frac{m'}{2s})\right) &\geq \\ 2(\hat{g}^2_\alpha + \hat{g}^2_{\alpha'}) \min\left(|\alpha| + 1 - t(1+\frac{m}{2s}), 2t - t(1+\frac{m}{2s})\right) &\geq \\ 2\hat{g}^2_\alpha \min\left(|\alpha| - t(1+\frac{m}{2s}), 2t - t(1+\frac{m}{2s})\right) + 2\hat{g}^2_{\alpha'} \min\left(|\alpha'| - t(1+\frac{m}{2s}), 2t - t(1+\frac{m}{2s})\right)\end{aligned}$$

and summing over α we conclude that

$$c(g^0, m') + c(g^1, m') \geq 2c(g, m), \tag{9}$$

i.e. the average value of the two sons is at least the value of the father.

Let us now consider the case $G_i \leq G/2$. Using (8) and the fact that the size of each affected coefficient decreases by at most 1, we have

$$c(g^0, m) + c(g^1, m) \geq 2(c(g, m) - G_i). \tag{10}$$

For the same reasons we also have

$$\sum_{2t \geq |\alpha| \geq t(1+\frac{m}{2s})} (\hat{g}^0_\alpha)^2 + (\hat{g}^1_\alpha)^2 \geq 2(G - G_i) \geq G. \tag{11}$$

It follows from (11), the definition of $c(g, m)$, and (6) that

$$c(g^0, m') + c(g^1, m') - (c(g^0, m) + c(g^1, m)) \geq G \cdot \frac{t}{2s} \cdot \frac{4sG_i}{tG} \geq 2G_i. \tag{12}$$

From (10) and (12) we see that (9) holds also in this case.

It follows that the value of $c(f, s)$ is at least the expected value of a leaf in the tree. For the leaves where we stopped because g had no gates of fanin greater than $32s$ it follows (since $m \geq 0$) that they are labeled by a value that it at most $c(g, 0)$ and this can be bounded, by Theorem 3.1, by td_0.

For the leaves defined by $m \leq 0$, first note that $m \geq -\frac{4s}{t}$ (this follows from $G_i \leq G$, (6) and the fact that we did not stop at the previous step). This implies that we can always use the bound

$$c(g, -\frac{4s}{t}) \leq (t+2) \leq 3t.$$

This implies that the contribution to the expectation of this type of leaves is at most $3tq$ where q is the probability of reaching such a leaf when randomly walking down the tree. We now estimate q.

For a given node v labeled by g and with chosen variable x_i consider the fraction of large bottom gates that are erased. Let f_0 the the fraction of gates that remains in g^0 and f_1 the fraction that remains in g^1. By the definition of l_i and (5) we have

$$f_0 + f_1 \leq 2 - l_i \leq 2 - \frac{16sG_i}{tG}.$$

Now define a random variable X_v which takes the value $\min(1 - f_0, \frac{16sG_i}{tG})$ with probability 1/2 and the value $\min(1 - f_1, \frac{16sG_i}{tG})$ with probability 1/2. We have that

$$E[X_v] \geq \frac{8sG_i}{tG} \tag{13}$$

and

$$0 \leq X_v \leq \frac{16sG_i}{tG} \leq \frac{16s}{t}. \tag{14}$$

Now consider any path that leads to leaf with a function which still has large-fanin gates at the bottom and $m \leq 0$.

We know that along such a path we must have

$$\sum_v X_v \leq s \tag{15}$$

since otherwise less than $2^s e^{-\sum_v X_v} < 1$ gates at the bottom with fanin $\geq 32s$ would remain.

By the update rule on m and the fact that its value is at most 0 at the leaf we know that

$$\sum_v \frac{4sG_i}{tG} \geq s$$

But this, by (13), implies

$$\sum_v E[X_v] \geq 2s. \tag{16}$$

Now, since the X_v are independent and we are in a good position to apply Chernoff bounds and in particular consider the following theorem which is Theorem A.1.19 of [2].

Theorem 4.2 *For every $C > 0$ and $\epsilon > 0$ there exists $\delta > 0$ so that the following holds: Let Y_i, $1 \leq i \leq n$, n arbitrary, be independent random variables with $E[Y_i] = 0$, $|Y_i| \leq C$ and $Var[Y_i] = \sigma_i^2$. Set $Y = \sum_{i=i}^n Y_i$ and $\sigma^2 = \sum_{i=1}^n \sigma_i^2$ so that $Var[Y] = \sigma^2$. Then for $0 < a \leq \delta\sigma$,*

$$Pr[Y \geq a\sigma] < e^{-\frac{a^2}{2}(1-\epsilon)}.$$

To apply this theorem set $Y_v = \frac{t}{16s}(E[X_v] - X_v)$. We have $E[Y_v] = 0$, $|Y_v| \leq 1$, and

$$Var[Y_v] = \left(\frac{t}{16s}\right)^2 Var[X_v] \leq \left(\frac{t}{16s}\right)^2 E[X_v^2] \leq \frac{t}{16s} E[X_v]. \tag{17}$$

If we let μ denote $\sum_v E[X_v]$ then (15) and (16) implies that

$$\sum_v Y_v \geq \frac{t}{16s}(\mu - s) \geq \frac{t\mu}{32s}.$$

By (17) $\sigma^2 \leq \frac{t\mu}{16s}$ and applying Theorem 4.2 with $C = 1$, $\epsilon = \frac{1}{2}$, and $a = \min(\delta\sigma, \sigma/2)$, we see that the probability of (15) happening is bounded by

$$2^{-\Omega(\frac{t\mu}{s})} \leq 2^{-\Omega(t)}.$$

The proof of the inequality (2) is complete and hence the theorem follows. ■

5 Circuits approximating parity

It is interesting to investigate constructions showing to what extent our bounds are tight. We have nothing new to report but take the opportunity to recall known circuits approximating parity and hence giving a large value of $\hat{f}_{[n]}$. The first construction of these circuits is unknown to us and we consider them as folklore.

Divide the inputs into groups of s^{d-1} variables and we want to compute the parity of each group in depth d and size around 2^s. The parity of such a group can be computed by a "parity-tree" of depth $d-1$ and fanout s where each node computes the parity of its inputs. Write each such parity as a depth 2 circuit of size 2^s, but use CNF on even levels and DNF on odd levels. The creates a depth $2(d-1)$ circuit of size $O(ds2^s)$ but closer inspection shows that we have adjacent levels of gates of the same type and thus it can be collapsed to a depth d circuit. We thus have ns^{1-d} circuits each computing the parity of a subset. Assume that the output gates of these circuits are all $\vee$-gates.

Take the $\vee$ of these circuits. This does not increase the depth and gives a correlation with parity that is at least $2^{-ns^{1-d}}$, to see that latter, note that when the circuit outputs 0 then it is always correct. The size of the constructed circuits is only marginally larger than 2^s.

This implies that our bounds are close to tight when $d=2$ and for general d when $t \leq s^d$.

Our bounds are probably not tight for other values of s and d. In particular for small values of s better bounds are known. In fact, Ajtai [1] proved that $\hat{f}_{[n]}$ is of size at most

$$2^{-\Omega(n^{1-\epsilon})}$$

for any $\epsilon > 0$ as long as d is any constant and $s = O(\log n)$.

Acknowledgment We are grateful to Gil Kalai for asking the question whether the bounds of [7] were sharp. We are also grateful for comments on the paper by an anonymous referee.

References

[1] M. Ajtai. Σ_1^1 formulae on finite structures. Annals of Pure and Applied Logic, 24:1-48, 1983.

[2] N. Alon and J. Spencer, The probabilistic Method, 2nd edition, 2000, Wiley, New York.

[3] Paul Beame. A switching lemma primer. Technical Report UW-CSE-95-07-01, Department of Computer Science and Engineering, University of Washington, November 1994.

[4] M. Furst, J. Saxe and M. Sipser, Parity, circuits, and the polynomial-time hierarchy, Math. Systems Theory 17 (1984), 13– 27.

[5] J. Håstad. Almost Optimal Lower Bounds for Small Depth Circuits, in *Randomness and Computation*, Advances in Computing Research, Vol 5, ed. S. Micali, 1989, JAI Press Inc, pp 143–170.

[6] J. Håstad Computational limitations for small depth circuits. MIT Press, 1986. Ph.D Thesis.

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

[8] Y.Mansour An $O(n^{\log\log n})$ learning algorithm for DNF under the uniform distribution. Journal of Computer and Systems Sciences 50(3):543-550 (1995).