Electronic Colloquium on Computational Complexity, Report No. 81 (2008)

A simple proof of Bazzi's theorem

Alexander A. Razborov *

September 11, 2008

Abstract

In 1990, Linial and Nisan asked if any polylog-wise independent distribution fools any function in AC^0. In a recent remarkable development, Bazzi solved this problem for the case of DNF formulas. The aim of this note is to present a simplified version of his proof.

In the 1990s, it was shown in a series of papers [LMN93, BRS91, ABFR94] that Boolean functions computable by constant depth polynomial size circuits can be well approximated (in various contexts) by low degree polynomials. Around the same time, Linial and Nisan [LN90] conjectured that any such function can be fooled by a polylog-wise[1] independent probability distribution. By linear duality, this conjecture is an approximation problem of precisely the kind considered in [LMN93, BRS91, ABFR94]. Therefore, it is quite remarkable that the only noticeable progress in this direction was achieved only last year by Bazzi [Baz07]. Namely, he showed that any DNF formula of polynomial size is fooled by (any) $O(\log n)^2$-independent distribution. We refer the reader to [Baz07] for motivations and applications of this result; the purpose of this note is to give a simplified version of Bazzi's proof.

For a probability distribution μ on $\{0,1\}^n$ and a function $f : \{0,1\}^n \longrightarrow \mathbb{R}$, $E_\mu(f)$ is the expected value of f w.r.t. this distribution (in particular, if f :

*University of Chicago, US and Steklov Mathematical Institute, Moscow, Russia, razborov@cs.uchicago.edu. Supported by the Russian Foundation for Basic Research.

[1]As literally stated in [LN90] the conjecture is false [LV96], so we relax the parameters appropriately.

ISSN 1433-8092

$\{0,1\}^n \longrightarrow \{0,1\}$ is a Boolean function then $E_\mu(f) = \mathbf{P}_{x\sim\mu}[f(x)=1]$ is the probability that $f(x)=1$). If μ is uniform on $\{0,1\}^n$, $E_\mu(f)$ is abbreviated to $E(f)$. The *bias* of f w.r.t. μ is defined as $|E_\mu(f) - E(f)|$, and for an integer $k \geq 0$, $\text{bias}(f;k) \stackrel{\text{def}}{=} \max_\mu |E_\mu(f) - E(f)|$, where the maximum is taken over all k-independent probability distributions on $\{0,1\}^n$.

In this note we give a simplified proof of the following theorem:

Theorem 1 (Bazzi [Baz07]) *If the Boolean function $f : \{0,1\}^n \longrightarrow \{0,1\}$ is computable by an m-term DNF formula then* $\text{bias}(f;k) \leq m^{O(1)} \exp(-\Omega(\sqrt{k}))$.

From now on we will identify a DNF formula $F = A_1 \vee \ldots \vee A_m$ and the Boolean function it represents. The first step in the proof of Theorem 1 is to reduce the problem to the case when every conjunctive term A_i has only a few variables, that is F is an s-DNF for a sufficiently small s. This simple step is borrowed from [Baz07] without any changes:

Lemma 2 ([Baz07]) *Let $k \geq s \geq 1$ be integers, and F be an m-term DNF. Then*

$$\text{bias}(F;k) \leq \max_G \text{bias}(G;k) + m2^{-s},$$

where the maximum is taken over all m-terms s-DNF G.

The next relatively simple step in Bazzi's proof that we also reproduce here without alterations is to estimate the bias of an s-DNF F in terms of a constrained version of ℓ_2-approximation by low degree polynomials called in [Baz07] *zero-energy*. Let us first recall the unconstrained version.

Definition 3 For a function $f : \{0,1\}^n \longrightarrow \mathbb{R}$ and an integer $t \geq 0$, let

$$\text{energy}(f;t) \stackrel{\text{def}}{=} \min_{\deg(g) \leq t} E((f-g)^2).$$

This quantity is equal to the sum of squares $\sum_{|S|>t} \widehat{f}(S)^2$ of high order Fourier coefficients of f. But we do *not* need this interpretation in our proof, besides making connection to the following celebrated result by Linial, Mansour and Nisan [LMN93]:

Lemma 4 ([LMN93]) *If f is a Boolean function computable by an $\{\neg, \wedge, \vee\}$-circuit of size m and depth d then for any $t > 0$,*

$$\text{energy}(f;t) \leq 2m \cdot 2^{-t^{1/d}/20}.$$

Definition 5 ([Baz07])

$$\text{zeroEnergy}(f;t) \stackrel{\text{def}}{=} \min_{\deg(g)\le t} E((f-g)^2),$$

where this time the minimum is taken over all degree $\le d$ polynomials g that satisfy one additional **zero-constraint**: $g(x) = 0$ whenever $f(x) = 0$ ($x \in \{0,1\}^n$).

Clearly, $\text{energy}(f;t) \le \text{zeroEnergy}(f;t)$. Also, bias is related to zero-energy with the following lemma:

Lemma 6 ([Baz07]) *Let F be an m-term s-DNF formula and let $k \ge s$ be an integer. Then*

$$\text{bias}(F;k) \le m \cdot \text{zeroEnergy}\left(F; \lfloor (k-s)/2 \rfloor\right).$$

In the opposite direction, bounding zero-energy in terms of energy of certain auxiliary functions is where the bulk of work is done in Bazzi's proof. And this is where our simplification comes in:

Theorem 7 *Let F be an m-term s-DNF and t be an integer. Then*

$$\text{zeroEnergy}(F;t) \le m^2 \cdot \max_G \text{energy}(G; t-s), \tag{1}$$

where the maximum is again taken over all m-term s-DNF formulas G.

Proof. Let $F = A_1 \vee \ldots \vee A_m$, where A_i are conjunctive terms of size $\le s$ each. We claim that F can be expressed in the form

$$F = \sum_{i=1}^m A_i(1 - \mathbf{E}[\boldsymbol{G_i}]), \tag{2}$$

where $\boldsymbol{G_i}$ are specially constructed random sub-DNFs of F and the expectation sign is understood pointwise: $\mathbf{E}[\boldsymbol{G}]\,(x) \stackrel{\text{def}}{=} \mathbf{E}[\boldsymbol{G}(x)]$ $(x \in \{0,1\})^n$. But before exhibiting the distributions of $\boldsymbol{G_i}$ with this property, let us see why their mere existence already implies the statement of Theorem 7.

Indeed, denoting the maximum $\max_G \text{energy}(G; t-s)$ in (1) by ϵ, we have (random) polynomials $\boldsymbol{g_i}$ of degree $\le t-s$ such that with probability one we have the bound $E((\boldsymbol{G_i} - \boldsymbol{g_i})^2) \le \epsilon$. And now we simply let

$$g \stackrel{\text{def}}{=} \sum_{i=1}^m A_i(1 - \mathbf{E}[\boldsymbol{g_i}]).$$

Since every term A_i has at most s variables, $\deg(g) \le t$. $F(x) = 0$ implies $\forall i \in [m](A_i(x) = 0)$ which in turn implies $g(x) = 0$. Therefore, g satisfies the zero-constraint. And we bound the ℓ_2-distance between F and g as follows:

$$\begin{aligned} E((F-g)^2) &= E\left(\left(\sum_{i=1}^m A_i \cdot \mathbf{E}[\boldsymbol{G_i} - \boldsymbol{g_i}]\right)^2\right) \\ &\le_{\text{Cauchy-Schwartz}} m \cdot \sum_{i=1}^m E\left((A_i \cdot \mathbf{E}[\boldsymbol{G_i} - \boldsymbol{g_i}])^2\right) \\ &\le_{\text{since } |A_i| \le 1} m \cdot \sum_{i=1}^m E\left(\mathbf{E}[\boldsymbol{G_i} - \boldsymbol{g_i}]^2\right) \\ &\le_{\text{Cauchy-Schwartz}} m \cdot \sum_{i=1}^m E\left(\mathbf{E}\left[(\boldsymbol{G_i} - \boldsymbol{g_i})^2\right]\right) \\ &= m \cdot \sum_{i=1}^m \mathbf{E}\left[E\left((\boldsymbol{G_i} - \boldsymbol{g_i})^2\right)\right] \le \epsilon m^2. \end{aligned}$$

It remains to exhibit $\boldsymbol{G_1}, \ldots, \boldsymbol{G_m}$ such that the identity (2) holds. For that purpose, we first pick $\boldsymbol{p} \in [0,1]$ uniformly at random. And then we let $\boldsymbol{G_i}$ be the sub-DNF of $(A_1 \vee \ldots \vee A_{i-1} \vee A_{i+1} \vee \ldots \vee A_m)$ in which every term is removed, independently of others, with probability $\boldsymbol{p}$ and kept alive with probability $1 - \boldsymbol{p}$.

Fix an input $x \in \{0,1\}^n$, and let $w \stackrel{\text{def}}{=} |\{ i \in [m] \mid A_i(x) = 1 \}|$. If $w = 0$ then both sides of (2) are equal to 0.

If, on the other hand, $w > 0$ then there are precisely w non-zero terms in the expression $\sum_{i=1}^m A_i(x)(1 - \mathbf{E}[\boldsymbol{G_i}](x))$. And every one of them contributes to the sum precisely

$$\int_0^1 (1 - \mathbf{E}[\boldsymbol{G_i}(x) \mid \boldsymbol{p} = p])dp = \int_0^1 \mathbf{P}[\boldsymbol{G_i}(x) = 0 \mid \boldsymbol{p} = p]\, dp = \int_0^1 p^{w-1} dp = \frac{1}{w}.$$

Thus, $\sum_{i=1}^m A_i(x)(1 - \mathbf{E}[\boldsymbol{G_i}](x)) = 1$ $(w > 0)$, and this completes the proof of (2) and of Theorem 7.■

Like in Bazzi's proof, Theorem 1 immediately follows from Lemma 2, Lemma 6, Theorem 7 and Lemma 4.

Remark. After the preliminary version of this note was disseminated, Avi Wigderson observed that the proof can be further simplified by (deterministically!) letting G_i in (2) be equal $A_1 \vee \ldots \vee A_{i-1}$. This is definitely simpler, but our version has the potential advantage of being more symmetric.

Acknowledgement. I am grateful to Scott Aaronson for useful discussions that essentially triggered off this work, to Louay Bazzi for carefully checking the correctness of the proof and to Avi Wigderson for his permission to include here the observation above.

References

[ABFR94] J. Aspnes, R. Beigel, M. Furst, and S. Rudich. The expressive power of voting polynomials. *Combinatorica*, 14(2):1–14, 1994.

[Baz07] L. Bazzi. Polylogarithmic independence can fool DNF formulas. Manuscript, available at http://www.mit.edu/~louay/recent/kwisednf2.pdf, 2007.

[BRS91] R. Beigel, N. Reingold, and D. Spielman. The perceptron strikes back. In *Proceedings of the 6th IEEE Conference on Structure in Complexity Theory*, pages 286–291, 1991.

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

[LN90] N. Linial and N. Nisan. Approximate inclusion-exclusion. *Combinatorica*, 10(4):349–365, 1990.

[LV96] M. Luby and B. Velickovic. On deterministic approximation of DNF. *Algorithmica*, 16(4/5):415–433, 1996.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de/