Boolean functions with low average sensitivity depend on few coordinates

Ehud Friedgut *

February 3, 1998

Abstract

Consider a function $f : \{0,1\}^n \to \{0,1\}$. The sensitivity of a point $v \in \{0,1\}^n$ is $|\{v' : f(v') \neq f(v),\ \mathrm{dist}(v,v') = 1\}|$, i.e. the number of neighbors of the point in the discrete cube on which the value of f differs. The *average sensitivity* of f is the average of the sensitivity of all points in $\{0,1\}^n$. (This can also be interpreted as the sum of the influences of the n variables on f, or as a measure of the edge boundary of the set which f is the characteristic function of.) We show here that if the average sensitivity of f is k then f can be approximated by a function depending on c^k coordinates where c is a constant depending only on the accuracy of the approximation but not on n. We also present a more general version of this theorem, where the sensitivity is measured with respect to a product measure which is not the uniform measure on the cube.

1 Definitions and Main Theorem

Consider a function $f : \{0,1\}^n \to \{0,1\}$. The sensitivity of a point $v \in \{0,1\}^n$ is $|\{v' : f(v') \neq f(v),\ \mathrm{dist}(v,v') = 1\}|$, i.e. the number of neighbors of the point in the discrete cube on which the value of f differs. The *average sensitivity* of f, $as(f)$ is the average of the sensitivity of all points in $\{0,1\}^n$.

* Institute of Mathematics, The Hebrew University of Jerusalem, Givat Ram, Jerusalem, Israel. Email: ehudf@math.huji.ac.il. Mathematics Subject Classification (1991): 05C80, 28A35 .

A related concept is that of *influence*: The influence of the ith variable on such a Boolean function of n variables is the probability that after assigning all the other $n-1$ variables values at random the value of the ith variable determines the value of f. It is easy to see that the sum of the influences of the variables on f is exactly the average sensitivity of f.
Another point of view is taking a set $A \subset \{0,1\}^n$ and considering *the edge boundary* of A, i.e. the number of edges between A and its complement in the graph of the cube. letting $f = \chi_A$ one has that the edge boundary of A divided by a normalizing factor of 2^{n-1} is exactly $as(f)$, and the influence of the ith variable on f is the number of edges emitting from A parallel to the corresponding direction, divided by 2^{n-1}. So the theorems in this paper can be viewed as isoperimetric inequalities.
Several papers, ([8][11][4] [5]), have been written proving theorems which all have in common the following rule of thumb: "If the influences are distributed among many variables, then their sum is large". (Using the notion of influence it is in place to remark that this paper is greatly influenced by [5].) In the same spirit we will now prove that if the sum of the influences is small then the value of the function can be determined with high probability by knowing the values of a small set of variables. For example of all Boolean functions with $\mathrm{Prob}\{f=1\} = 1/2$ the one with lowest sensitivity, $as(f) = 1$, is $f(x_1, x_2, \ldots, x_n) = x_1$, which depends on one variable only.
In the case of f being a monotone Boolean function this result has to do with the understanding of *threshold* phenomena, see [4].

Theorem 1.1 *Let f be a Boolean function of n variables with average sensitivity $as(f) = k$. Let $\epsilon > 0$. Let $M = k/\epsilon$. Then there exists a Boolean function h depending only on* $\exp\left(\left(2+\sqrt{\frac{2\log(4M)}{M}}\right)M\right)$ *variables, such that $Prob\{f \neq h\} \leq \epsilon$.*

Another way of looking at this result is that functions with low sensitivity can be approximated in the L_2 norm by polynomials of low degree. see [7] for similar results. Following this outlook our proof also gives the following theorem:

Theorem 1.2 *Let f be a Boolean function of n variables with average sensitivity $as(f) = k$. Let $\epsilon > 0$. Let $M = k/\epsilon$. Then there exists polynomial g of degree at most* $\exp\left(\left(2+\sqrt{\frac{2\log(4M)}{M}}\right)M\right)$, *such that $Prob\{f \neq g\} \leq \epsilon/2$.*

Notice that the approximation in theorem 1.2 improves that of theorem 1.1 by a constant of 2. This is the price paid in passing from the polynomial to the Boolean function.

2 Tightness of Results

The following example shows that theorem 1.1 is tight up to the value of the base of the exponent: we construct a function f such that:
(a) $\mathrm{Prob}\{f = 1\} \sim 1/2$
(b) $as(f) = k + o(1)$
(c) at least $c^{k/\epsilon}$ variables are needed for an ϵ-approximation, where $c = 2^{\frac{\epsilon}{6} \cdot \frac{k-1}{k}}$.
By more careful construction one can improve the constant in this bound, however we do not know what the best value is.

Set

$$n = \log_2(1/6\epsilon), \; T = (k-1)e/6\epsilon, \; m = T2^T.$$

Divide the integers $\{1, \ldots, m\}$ into 2^T "tribes" of size T (See [2] for this) and for Boolean variables $z_1, \ldots, z_m$ define $g(z) = 1$ iff there exists a whole "tribe" in which all variables indexed by it have value 1. Now for variables $x, y_1, \ldots, y_t, z_1, \ldots, z_m$ define

$$f(x, y_1, \ldots, y_t, z_1, \ldots, z_m) = x_0 \wedge ((1 - \wedge y_i) \vee g(z_1, \ldots, z_m)).$$

In order to approximate f up to ϵ one is forced to use $(1 - 1/e - 1/3)T2^T$ of the variables z_i in order to get a good approximation of g. This gives order of$c^{k/\epsilon}$ variables.

One may feel that the preceding example is not satisfactory since for each value of ϵ we construct a different function. Here is an example of a function f on the infinite dimensional discrete cube with bounded average sensitivity such that there exists a constant c and a series $\{\epsilon_i\}$ which tends to zero, and $c^{1/(\epsilon_i \log^2(1/\epsilon_i))}$ variables are needed for an ϵ_i-approximation.
For $i = 1, 2, \ldots$ define , $T_i = 2^i/i^2, M_i = T_i 2_i^T$.
Let $g_i(z_1^i, z_2^i, \ldots, z_{M_i}^i)$ be a "tribes" function as before.
Let $f_i = (x_1^i \wedge x_2^i \wedge ... \wedge x_i^i \wedge g_i(z_1^i, z_2^i, \ldots, z_{M_i}^i)$ where the set of variables of f_i is disjoint from that of f_j for $j \neq i$. Let

$$f = f_1 \vee f_2 \vee \ldots \vee f_n \ldots$$

To approximate f within 2^{-n} one needs c^{2^n/n^2} variables. A similar construction can give the result c^{2^n/n^τ} (with a larger constant c) for any $\tau > 1$. We do not know if this is the worst possible asymptotic behavior.

3 Fourier Analysis and Proof of Theorem.

The proof of the theorem uses Fourier analysis on the discrete cube and a hypercontractive estimate due to Bonami and Beckner ([1], [3]). This was also a key ingredient in [5]. For every $v \in \{0,1\}^n$ define $U_v : \{0,1\}^n \to \{0,1\}$ by $U_v(v') = (-1)^{<v,v'>}$.
These functions form an orthonormal basis for the space of real functions on $\{0,1\}^n$ with respect to the uniform discrete measure. For a function $f : \{0,1\}^n \to \{0,1\}$ one has the expansion of f as a linear combination with respect to this basis, the Fourier-Walsh expansion of f:

$$f = \sum \hat{f}(v) U_v$$

where $\hat{f}(v) =< f, U_v >= 2^{-n} \sum f(w) U_v(w)$.
For any function f on $\{0,1\}^n$ and $p \geq 1$ define the L_p norm of f :

$$\|f\|_p = (2^{-n} \sum |f(v)|^p)^{1/p}.$$

From the orthogonality of $\{U_v\}$ we get Parseval's identity:

$$\|f\|_2^2 = \sum \hat{f}^2(v).$$

The following lemma connects different norms of functions on the discrete cube. For any $\delta \leq 1$ define the following linear operator T_δ:

$$T_\delta(\sum \hat{f}(v) U_v) = \sum \hat{f}(v) \delta^{|v|} U_v.$$

Where $|v| = \sum v_i$.

Lemma 3.1 (Beckner, Bonami) *As an operator from $L_{1+\delta^2}$ to L_2, T_δ has norm 1.*

We will now define n functions f_i which serve to measure the influence of the ith variable on f. Define

$$f_i(v) = f(v) - f(v \oplus (i))$$

where $v \oplus (i)$ is the vector resulting by changing the value of v_i to $1 - v_i$. There is a close connection between the Fourier coefficients of f and those of f_i:

$$\hat{f}_i(v) = \begin{cases} 2\hat{f}(v) & \text{if } v_i = 1 \\ 0 & \text{if } v_i = 0 \end{cases} \tag{1}$$

The proof of this is straightforward, see [5].
From the definition of f_i we have $\|f_i\|_2^2$ is exactly the influence of the ith variable on f. From Parseval and 1 we have

$$as(f) = 4 \sum \hat{f}^2(v)|v| \tag{2}$$

Here is a quick sketch of the proof: given f first we define a set I of variables with small influence. Using the hypercontractive estimate on f_i for $i \in I$ together with (2) we deduce that most of the L_2 norm of f is concentrated on vectors v such that $v_i = 0$ for $i \in I$.
We then define a function g depending only on the coordinates not in I such that $\|f - g\|_2$ is small. Finally we replace g by a Boolean function which is still a good approximation of f.
Proof: Let f be a Boolean function with $as(f) = k$. Let $s_i = \|f_i\|_2^2$ be the influence of the ith variable on f.
Define The set I of coordinates with small influence:

$$I = \{i : s_i < \exp(-d)\}$$

where d is a parameter to be determined. (Note that the complement of I in the set $\{1, 2, \ldots, n\}$ is of size no more than $k \cdot \exp(d)$.) Note that since f_i takes on only values in $\{0, 1, -1\}$, $\|f_i\|_q^q = s_i$ for any $q \geq 1$, making it particularly convenient to use the hypercontractive estimate on it .
Applying lemma 3.1 to f_i, $i \in I$ we get

$$\|T_\delta(f_i)\|_2^2 \leq \|f_i\|_{1+\delta^2}^2 = s_i^{2/(1+\delta^2)} \tag{3}$$

For a vector $v \in \{0,1\}^n$ define $|v|_I = \sum_{i\in I} v_i$. Summing (3) on all $i \in I$ gives by Parseval:

$$4\sum \hat{f}^2(v)\delta^{2|v|}|v|_I \leq \sum_{i\in I} s_i^{2/(1+\delta^2)} \leq k \cdot \exp\left(d(1-2/(1+\delta^2))\right) \qquad (4)$$

where the last inequality uses the convexity of the function $x^{2/(1+\delta^2)}$, and the fact that $\sum s_i \leq k$. Now from (2) we have that at most $\epsilon/4$ of the L_2 norm of $\hat{f}$ is on vectors v such that

$$|v| \geq k/\epsilon \qquad (5)$$

(i.e. $\sum \hat{f}^2(v)$ on such vectors is no more than $\epsilon/4$.)
From (4) we have that at most $\epsilon/4$ of the L_2 norm is on vectors v such that

$$\delta^{2|v|}|v|_I \geq \exp\left(d(1-2/(1+\delta^2))\right) \cdot 4k/\epsilon \qquad (6)$$

So all but $\epsilon/2$ of the weight of $\hat{f}^2$ is concentrated on vectors v such that neither (5) nor (6) are true. By a suitable choice of d the negation of these two conditions can apply simultaneously only if $|v|_I = 0$, which will show that indeed most of the Fourier transform of f does not depend on I.
Denoting $k/\epsilon = M$ and $\delta^2 = x$ we have that for the vectors in question, if $|v|_I$ is not 0 (it is always a non-negative integer):

$$x^M \leq 4M \exp\left(d(1-2/(1+x))\right).$$

Optimizing on x we get that for $d \geq \left(2+\sqrt{\frac{2\log(4M)}{M}}\right) M$ the conditions can only be fulfilled if $|v|_I = 0$.

Corollary 3.2 *For a Boolean f with $as(f) = k$, $d = \left(2+\sqrt{\frac{2\log(4k/\epsilon)}{k/\epsilon}}\right) k/\epsilon$ and $I = \{i : s(i) < \exp(-d)\}$*

$$\sum_{|v|_I>0} \hat{f}^2(v) \leq \epsilon/2.$$

Noticing that U_v is in fact a polynomial of degree $|v|$, ($U_v = \prod(1-2x_i)v_i$) we see that the following definition of g completes the proof of theorem 1.2:

$$\hat{g}(v) = \begin{cases} \hat{f}(v) & \text{if } |v|_I = 0 \\ 0 & \text{if } |v|_I > 0 \end{cases}$$

(The fact that the support of $\hat{g}$ is only on v such that $|v|_I = 0$ means that g indeed depends only on the variables not indexed by I.)
Now the reason this does not prove theorem 1.1 is that g indeed depends only on the variables not in I, but is not Boolean. Let J denote the complement of I. Define a Boolean function h on $\{0,1\}^n$ depending only on the coordinates in J by taking the best such approximation of f:
let $w \in \{0,1\}^n$ and let $V_w = \{v : v$ agrees with w on the coordinates in $J\}$. For any $v \in V_w$ define :

$$h(v) = \begin{cases} 1 & \text{If Prob}(\{f(x) = 1\}|x \in V_w) \geq 1/2 \\ 0 & \text{Otherwise} \end{cases}$$

h depends only on the coordinates in J, so all that is left to show is $\|f-h\|_2^2 \leq 2\|f-g\|_2^2$ and hence $\|f-h\|_2^2 \leq \epsilon$. This follows from the following lemma:

Lemma 3.3 *Let f be a Boolean function on $\{0,1\}^n$, g the best L_2 approximation of f among functions depending only on coordinates in a set J, and h the best L_2 approximation of f among* Boolean *functions depending only on the coordinates in J. Then $\|f-h\|_2^2 \leq 2\|f-g\|_2^2$.*

Proof: For $w \in \{0,1\}^n$ define V_w as before. Let $p(w) = \text{Prob}(\{f(x) = 1\}|x \in V_w)$. From the definition of g it is easy to compute that $g(w) = p(w)$, and from the definition of h we have $h(w) = 1$ iff $p(w) \geq 1/2$. For each $v \in \{0,1\}^J$ choose a representative $w \in \{0,1\}^n$, so that w agrees with v in the appropriate coordinates. Then
$\|f-g\|_2^2 = 2^{-|J|}\sum p(w)(1-p(w))$, and
$\|f-h\|_2^2 = 2^{-|J|}\sum \min\{p(w), 1-p(w)\}$.

□

4 The Case of Non-uniform Measure

The definition of the influence of a variable on a Boolean function makes sense for any probability measure on $\{0,1\}^n$. In this section we present a general version of the previous theorems.
Let $0 < p < 1$ and $q = 1-p$ and consider the measure on the two-point space giving 0 the measure q and 1 the measure p. We will consider μ_p, the product measure on $\{0,1\}^n$, i.e. $\mu(v) = p^{|v|}q^{n-|v|}$. For $f : \{0,1\}^n \to \{0,1\}$ define as before $s_i = \text{Prob}\{f(v) \neq f(v \oplus (i))\}$, and $as(f) = \sum s_i$.

Theorem 4.1 *Let $f : \{0,1\}^n \rightarrow \{0,1\}$, and $as(f) = k$ with respect to the product measure μ_p. Let $\epsilon > 0$. Then there exists a Boolean function h depending only on $c^{(k/\epsilon)}$ variables, (where c is a constant that depends only on p), such that $Prob\{f \neq h\} \leq \epsilon$.*

One of the reasons that this theorem is of interest is the following interpretation of $as(f)$: If f is a monotone Boolean function i.e. the characteristic function of a monotone subset A of the discrete cube, then $\mu_p(A)$ is a monotone function of p for $p \in [0,1]$. In [6], [9] Russo and Margulis (using different definitions) connect $as(f)$ with the derivative of this function:

Lemma 4.2 *Let A be a monotone subset of $\{0,1\}^n$, and f its characteristic function. Then*

$$d\mu_p(A)/dp = as(f)$$

The study of $d\mu_p(A)/dp$ arises in understanding threshold phenomena in percolation theory and random graphs, see for example [4], [8].

The proof of theorem 4.1 is almost identical to the previous proof once the analogous machinery is set up. We will present below the main tools of the proof, and leave the proof itself to the interested reader. For every $i \in \{1..n\}$ define U_i, a real function defined on $\{0,1\}^n$:

$$U_i(v) = \begin{cases} -\sqrt{q/p} & \text{if } v_i = 1 \\ \sqrt{p/q} & \text{if } v_i = 0 \end{cases}$$

For every $v \in \{0,1\}^n$ define

$$U_v = \prod_{\{i : v_i = 1\}} U_i.$$

These functions are orthonormal with respect to μ_p. For any f define the Fourier expansion

$$f = \sum \hat{f}(v) U_v$$

where $\hat{f}(v) =< f, U_v >$. For a Boolean function f define the functions f_i by

$$f_i(v) = \begin{cases} q & \text{if } f(v) - f(v \oplus (i)) = 1 \\ p & \text{if } f(v) - f(v \oplus (i)) = -1 \\ 0 & \text{if } f(v) - f(v \oplus (i)) = 0 \end{cases}$$

As before there is a close connection between the Fourier coefficients of f and f_i:

$$\hat{f}_i(v) = \begin{cases} \hat{f}(v) & \text{if } v_i = 1 \\ 0 & \text{if } v_i = 0 \end{cases}$$

We have

$$s_i = qp\|f_i\|_2^2$$

and therefore

$$as(f) = qp \sum \hat{f}^2(v)|v|.$$

Notice that in the case $p = 1/2$ all the above coincides with what was described in the previous section. The functions f_i, U_i, and their properties are described in [11].

The final missing tool in order to proceed with the proof as in the case $p = 1/2$ is a lemma analogous to lemma 3.1. We will prove a slightly less general result that will suffice for our purposes.

Lemma 4.3 *Let $\{0,1\}^n$ be endowed with the product measure μ_p. Define the operator T by :*

$$T(\sum \hat{f}(v)U_v) = \sum \hat{f}(v)2^{-|v|}U_v.$$

There exists a constant $1 < \tau < 2$ such that, as an operator from L_τ to L_2, T has norm 1.

Proof: As shown in [3] and [1] the case of the product space follows from the result for a 2-point space.

On the two point space we can consider, after normalizing, functions f_x that assume the value 1 on the point 0 and x on the point 1.

The function $T(f_x)$ assumes the 2 values $(q+p/2)+xp/2$, and $q/2+x(p+q/2)$. in order to have $\|T(f)\|_2 \leq \|f\|_\tau$ we must show that for any positive x

$$4(q + px^\tau)^{2/\tau} - \big(q(q + 1 + px)^2 + p(q + (1 + p)x)^2\big) \geq 0. \qquad (7)$$

This is equivalent to finding a value of τ such that (7) holds for $0 \leq x \leq 1$ simultaneously for the given value of p, ($p = p_0$) and if the roles of p and q are interchanged (i.e. $p = 1 - p_0$.)

For $x = 1$ we have equality in (7), and for $x = 0$ we need

$$q^{2/\tau-1} \geq (1 + 3q)/4. \qquad (8)$$

Taking the derivative of (7) we will find a condition on τ so that the l.h.s. of (7) is a decreasing function of x in [0,1]. By differentiating we get the condition

$$4(q+px^\tau)^{2/\tau-1} - q(q+1+px) - (1+p)(q+(1+p)x) \leq 0. \tag{9}$$

Once again for $x = 1$ we have equality, and for $x = 0$ the condition holds. So we differentiate once again and get the condition

$$(\tau-1)x^{\tau-2}(q+px^\tau)^{2/\tau-1} + (2-\tau)px^{2(\tau-1)}(q+px^\tau)^{2/\tau-1} \geq (1+3p)/4. \tag{10}$$

This obviously holds if we demand that $q^{2/\tau-1} \geq \frac{1+3p}{4(\tau-1)}$.
In summary, keeping in mind the dual role of p and q we denote $\min\{p,q\} = r$, and all the conditions are fulfilled if τ is close enough to 2 so that

$$(\tau-1)r^{2/\tau-1} \geq (4-3r)/4.$$

$\square$

Acknowledgments

I would like to thank Nati Linial for instructive conversations on this topic.

References

[1] W. Beckner, Inequalities in Fourier analysis. Annals of Mathematics, 102 (1975), 159-182.

[2] M. Ben-Or and N. Linial, Collective coin flipping, in *Randomness and Computation* (S. Micali, ed.), Academic Press, New York, 1990, pp. 91-115. Earlier version: Collective coin flipping, robust voting games, and minima of Banzhaf value, Proc. 26th IEEE Symp. on the Foundation of Computer Science, 1985, pp. 408-416.

[3] Aline Bonami, Etude des coefficients Fourier des fonctiones de $L^p(G)$, Ann. Inst. Fourier (Grenoble) 20:2 (1970), 335-402.

[4] E. Friedgut, G. Kalai, Every monotone graph property has a sharp threshold, Proc. Amer. Math. Soc. 124 (1996), pp. 2993-3002 .

[5] J. Kahn, G. Kalai, and N. Linial, The influence of variables on Boolean functions, Proc. 29-th Ann. Symp. on Foundations of Comp. Sci., 68-80, Computer Society Press, 1988.

[6] G. Margulis, Probabilistic characteristics of graphs with large connectivity, Prob. Peredachi Inform. 10(1974), 101-108.

[7] N.Nisan, M. Szegedy, On the Degree of Boolean Functions as Real Polynomials. Computational Complexity 4 (1994), 301-313.

[8] L. Russo, On the critical percolation probabilities, Z. Wahrsch. werw. Gebiete, 43(1978), 39-48.

[9] L. Russo, An approximate zero-one law, Z. Wahrsch. werw. Gebiete, 61 (1982), 129-139.

[10] J. Schmidt-Pruzan, E. Shamir, A threshold for perfect matchings in random $d-$pure hypergraphs, Combinatorica 5 (1985) 81-94.

[11] M. Talagrand, On Russo's approximate 0-1 law, The Annals of Probability, 1994, Vol.22 No.3 1576-1587.