The Influence of Variables on Boolean Functions (extended abstract)

Jeff Kahn* Gil Kalai† Nathan Linial‡

1 Introduction

This paper applies methods from harmonic analysis to prove some general theorems on boolean functions. The result that is easiest to describe says that "Boolean functions always have small dominant sets of variables." The exact definitions will be given shortly, but let us be more specific: Let f be an $n-$variable boolean function taking the value zero for half of the 2^n variable assignments. Then there is a set of o(n) variables such that almost surely the value of f is undetermined as long as these variables are not assigned values. This proves some of the conjectures made in [BL].

These new connections with harmonic analysis are very promising. Besides the results on boolean functions they enable us to prove new theorems on the rapid mixing of the random walk on the cube, as well as new theorems in the extremal theory of finite sets.

We begin by reviewing some definitions from [BL]. Let f be a boolean function on n variables, and let S be some set of variables. The *influence of* S *over* f, denoted by $I_f(S)$ is defined as follows. Assign values to the variables not in S at random, that is, variables are set independently of each other and the probability of a zero assignment is one half. This partial assignment may already suffice to set the value of f. The probability that f remains undetermined is defined as the influence of S over f.

For concreteness let us temporarily restrict ourselves to functions f for which $Pr(f=0) = 1/2$. (The probability space consists of all binary $n-$strings with uniform distribution.) It was observed in [BL] that the average influence of a single variable over f is at least $1/n$. This is a consequence of a standard fact in combinatorics, the edge isoperimetric inequality for the cube. (e.g. [Bo, Theorem 16.2]). One also notices that for the function $f(x_1, ..., x_n) = x_1$ the influence of x_1 is one while all the other x_i have zero influence. So in this case the average is indeed as

*Department of Mathematics and Center for OR, Rutgers University, New Brunswick, New Jersey 08903.
Research supported in part by grants from NSF and AFOSR and by a Sloan Research Fellowship.
†Institute of Mathematics and Computer Science, Hebrew University, Jerusalem 91904, Israel.
‡IBM Research Almaden, 650 Harry Rd. San Jose, CA 95120 Computer Science Department, Stanford University and Institute of Mathematics and Computer Science, Hebrew University, Jerusalem 91904, Israel.
Research supported in part by grant NSF IRI - 8814953 at the Institute for Mathematical Studies in the Social Sciences, Stanford University, Stanford CA.

small as $1/n$. However in all the examples that were examined in that paper there was at least one variable whose influence was as big as $\Omega(\log n/n)$. In fact, Ben-Or and Linial construct a function f for which each variable has influence $\Theta(\log n/n)$. This prompted the conjecture that for every f with $Pr(f = 0) = 1/2$ there is a variable whose influence is $\Omega(\log n/n)$. This conjecture is proved in the present article. Moreover we show that the sum of squares of the individual influences is $\Omega(\log^2 n/n)$.

What can be said about the influence of larger sets of variables? We should first point out that in dealing with the influence of a set of variables there are several different quantities to be considered. We have already encountered one of them, viz., $I_f(S)$. Besides this, there is the influence of S towards zero which we now define, and the influence towards one, which is defined analogously. Let $p := Pr(f = 0)$. Assign values to the variables outside S at random, and denote by p' the probability that given the values assigned to the variables not in S, it is possible to assign values to the variables in S so as to make f equal to zero. The difference $p' - p$ is defined to be $I_f^0(S)$, the *influence of S toward zero.* The question then arises of the existence of small sets of variables with large influence, where the meaning of the question depends, of course, on the notion of influence intended. It is worth mentioning (and easy to check) that for all f and S :

$$I_f(S) = I_f^0(S) + I_f^1(S).$$

In the case we are most interested in, when $Pr(f = 0) = 1/2$, it is clear that influence towards either zero or one cannot exceed $1/2$. We look for sets which get close to this bound. Now in the above mentioned construction from [BL] a set of variables whose influence towards zero is $\frac{1}{2} - o(1)$ must have cardinality $\Omega(n/\log n)$. This was conjectured (ibid.) to be best possible. We prove a slightly weaker result showing that there always is a set of $O(n\omega(n)/\log n)$ variables whose influence towards zero is $\frac{1}{2} - o(1)$, where $\omega(n)$ is any function which tends to infinity with n. Clearly the same holds with zero replaced by one. Let us mention the following closely related problem: There is a construction (again from [BL]) of a function f where sets of $o(n^\alpha)$ variables have $o(1)$ influence both towards zero and towards one, $\alpha = \log 2/\log 3 = 0.63\ldots$. It may well be that there is a constant $\beta < 1$ such that there is always a set of $O(n^\beta)$ variables where at least one of $I_f^0(S), I_f^1(S)$ is $\frac{1}{2} - o(1)$. We are unable to settle this question at the time of writing.

Let us also remark that our results extend beyond the case where $Pr(f = 0) = 1/2$. We have stated all of our results in this case, since it is the most interesting one for computer science applications and to avoid more technical statements.

A word is in order now about our methods. We use ideas from harmonic analysis. This circle of problems turns out to be best viewed in terms of the Fourier analysis of the $n-$dimensional cube, thought of as the abelian group $\mathbf{Z_2^n}$. We assume familiarity with the most basic facts of harmonic analysis which can be found in essentially any text in the area. (For example Dym and McKean [DM] is an excellent introduction to the subject which contains numerous interesting applications.) We need only the most basic notions of this theory viz., *characters, dual group,* and *Fourier transform.* The only fact we use is Parseval's identity. The harmonic analysis of $\mathbf{Z_2^n}$, will be reviewed as needed. We make substantial use of Beckner's [B] elegant inequalities in (classical) Fourier analysis.

Our method enables us to prove new results on the rapid mixing of the random walk on the cube. While many of the properties of this walk are well studied and the speed at which it converges to the (uniform) limit distribution is known, not so much is known if we start from a distribution which is not concentrated at one point. In particular what if the initial distribution is uniform on a set of vertices of a given size? We are able to give estimates for this problem which turn out to be asymptotically correct for a large range of sizes. However, the problem in general is still far from solved.

Although it is tempting to conjecture that in worst (=slowest) case the initial distribution is supported on some simple set such as a Hamming ball or a subcube such an exact result seems well beyond the reach of present methods. In fact, the present analytic methods seem, for the most part, ill-suited to exact results, while combinatorial techniques which have proved quite powerful for extremal problems with more obvious candidates for extrema have to date been surprisingly ineffective for problems of the type we are considering. There are many natural problems in this area for which exact determination of extrema seems unlikely, and it may be that the correct approach to some of these involves blending combinatorial and analytic methods.

There is a close connection between the problems we mentioned on influence and some aspects of the following very general question: Let F be a family of m binary $n-$vectors. What can be said about the distribution of Hamming distances between the vectors in F ? At this level of generality this question is completely hopeless. In particular it contains all of the theory of error correcting codes. On the other hand, many special cases of this problem which may be tractable are very far from being understood. Our methods allow us to derive some new results on the following, narrower class of problems: How densely packed may F be? For example: given n, $|F|$ and an integer $b \leq n$, what is the largest possible number of pairs of vectors in F whose Hamming distance is at most b ?

For $b = 1$ this is answered by the edge isoperimetric inequality for the cube, mentioned before. The answer is basically that subcubes of the cube are the best families. Already for $b = 2$ this is not true. There seems to be a more complicated dependency on the relationship between $|F|$ and n. Our methods allow us to get estimates for this "dense packing" problem, which in certain cases are exact, and in other ranges can be shown to be fairly tight.

It is interesting to compare the outcomes of this method with what can be achieved using eigenvalues. Many of the questions addressed in this paper can be formulated as dealing with the expansion factors of various graphs. It is often possible to derive some estimates for the expansion factor from eigenvalues. However, this method is known to break down completely when applied to small sets of vertices. Our method succeeds in getting nontrivial estimates in some cases where the eigenvalue method fails. If this phenomenon can be extended to other graphs as well there could be extremely interesting consequences in theoretical computer science, but we were so far unable to make much further progress in this direction.

One more word about the literature: Many of the questions we consider here come from [BL], an earlier version of which is [BL']. A survey of this area including the connection with various problems in computer science, can be found in [BLS].

2 Harmonic Analysis of $\mathbf{Z_2^n}$ and influences.

Assuming some familiarity with basic harmonic analysis, we can explain its connection with the problems on influence described above. The group we deal with is $\mathbf{Z_2^n}$. As a set this is just the $n-$dimensional cube $\mathbf{C^n}$ and the group structure allows us to make use of the tools of harmonic analysis. We think of the elements in this group in a number of equivalent ways: as group elements, as binary vectors or as characteristic vectors of sets which we also identify with the sets themselves. All these terminologies will be used throughout. First we need to find all the characters. This is well known and easy to check, so we state this fact without proof:

Proposition 2.1: *Associate with every $A \subseteq [n]$, a real function $u = u_A$ defined on $\mathbf{Z_2^n}$ by:*

$$u_A(B) = (-1)^{|A \bigcap B|}.$$

Then u_A is a character for $\mathbf{Z_2^n}$ and moreover all irreducible characters of $\mathbf{C^n}$ are obtained in this way.

(Here and throughout $[n]$ stands for $\{1, ..., n\}$.)

Note that the isomorphism between $\mathbf{Z_2^n}$ and its dual is explicitly given by this proposition. We think of A both as an element of $\mathbf{C^n}$ and as the character associated with it. Throughout this paper we will deal with functions f defined on $\mathbf{C^n}$, typically expanded as $\sum_A \alpha_A u_A$. Note that the α's are the usual Fourier transform of f and are preferred to the traditional $\hat{f}$ only for typographic convenience. We also think of $\mathbf{C^n}$ as a probability space with uniform distribution. This allows us to take inner products of functions on the cube:

$$< f, g >:= \sum_A f(A)g(A)2^{-n}.$$

The Fourier coefficients for f are given by:

$$\alpha_A =< f, u_A >= \sum_B (-1)^{|A \bigcap B|} f(B).$$

We need the following fact from [BL] which is an easy consequence of the *shifting* technique of the extremal theory of finite sets (e.g. [F]):

Proposition 2.2: *For any boolean function f there exists a monotone boolean function g on the same set of variables such that*

$$(i) Pr(f = 0) = Pr(g = 0).$$

$$(ii) \forall A \subseteq [n],\ I_f(A) \geq I_g(A).$$

Inequality (ii) holds also for influences towards zero and one.

A consequence of this proposition is that there is no loss of generality in assuming that f is monotone. Now for a monotone f and any variable x_i the influence of x_i on f is easily seen to be given by:

$$2^{-n} \sum_{i \notin S} f(S \bigcup \{i\}) - f(S)$$

But this is exactly the same as $\alpha_{\{i\}}$. This fact creates the link between our problems and harmonic analysis. We are looking for bounds on the Fourier coefficients of certain real functions defined on $\mathbf{C^n}$. Our first difficulty is how to exploit the condition that our functions take only the values $0, 1$, which is not particularly natural from the standpoint of mathematical analysis. Roughly, this is accomplished as follows. Some initial combinatorial manipulations reduce the problem to another, similar problem involving functions taking values $0, 1, -1$, but having relatively small support. In this case it is not so much the precise range of the functions as the fact that we have good control of their various norms which becomes useful, and we are able to complete the proof using some inequalities of Beckner [B] relating the norms of a function and those of its images under certain linear operators.

First we prove our main new theorems on boolean functions and then we go on to sketch some sample new results on random walks on the cube and in the extremal theory of finite sets.

3 Lower bounds on influences.

We begin with a statement of our result on the influence of single variables. The result is given now in its more general form and not only for the case where f is equally often zero and one. This more general form can then be applied repeatedly to derive the lower bounds on the influence of sets of variables. Let us recall that $\mathbf{C^n}$, the $n-$dimensional cube is equipped with the uniform probability distribution, so we can speak, for example of the probability that f is zero.

Theorem 3.1: *Let f be a boolean function on n variables, which equals one with probability p and assume $p \leq 1/2$. Then*

$$\sum (I_f(x_i))^2 \geq Cp^2 \log^2 n/n$$

where C is an absolute positive constant, (for example $C = 1/5$ suffices.) Consequently there exists at least one variable whose influence is at least $Cp \log n/n$. These bounds are tight except for the value of C.
Also,

$$\sum (I_f(x_i)) \geq p \log \frac{1}{p}.$$

This bound is tight.

So these are the best lower bounds for the vector of influences of the individual variables in euclidean (L^2) norm, max (L^∞) norm and sum (L^1) norm. The L^∞ estimate is, of course, an easy consequence of the L^2 estimate. In fact with a little more care we can get essentially the same bound even for much smaller norms. For example, let $p = 1/2$, and set $\epsilon := C \log \log n / \log n$, and $q = (1+\epsilon)$. Then the L^q-norm of the vector of influences is $\Omega(\frac{(C-1)\log n}{n})$ which is particular in $\Omega(\frac{\log n}{n})$ if C-1 is greater than some positive constant. This is very nearly best possible, since for $C < 1$ the upper bound of (approximately) $\log^C n/n$ obtained from the half-cube is $o(\log n/n)$. It does seem reasonable to conjecture that for every q the correct answer is the minimum of $\log n/n$ and the value obtained from the half-cube.

The L^1 estimate mentioned in the theorem is nothing but a restatement of the edge isoperimetric inequality for the cube. It is quoted here only for completeness' sake. In the case $p = 1/2$ it implies the existence of a variable with an influence of at least $1/n$, as noted already in [BL]. Some improvements on this latter bound were made by Noga Alon [A] who, with eigenvalue arguments increased the bound to $(2-\epsilon)/n$, and by B. Chor and M. Gereb [CG] who proved $(3-\epsilon)/n$. It is interesting that the three approaches (all arrived at independently) have essentially the same point of departure (though not all in the same language). To date no bound better than the rather trivial $1/n$ has been obtained by what could be considered a purely combinatorial argument. We should also mention here that for very small p (up to $2^{-n/2}$) the best possible results on the max norm are available. Frankl [Fr] solved the problem using the Kruskal-Katona Theorem. We do not see how to extend this to larger p. For the more restricted class of f's corresponding to intersecting families of subsets (in the language of game theory, symmetric games), an equivalent version of the problem of minimizing the maximum influence had been raised earlier (as the first case of a more general question) by Daykin and Frankl [DF], who also observed the $1/n$ lower bound.

We now turn to the proof.

First we define a set of n functions on $\mathbf{C^n}$ whose range is $\{-1, 0, 1\}$. The i-th of those $(1 \le i \le n)$ is denoted by f^i and is defined by

$$f^i(T) := f(T) - f(T \oplus \{i\})$$

where $\oplus$ stands for symmetric difference, or equivalently in terms of binary vectors the mod 2 sum which is the same, of course. The shorthand $T \oplus i$ is used below.

Return to the Fourier expansion of f:

$$f = \sum_S \alpha_S u_S.$$

The expansion of f^i is written as:

$$f^i = \sum_S \alpha_S^i u_S.$$

To evaluate the α_S^i we write:

$$\alpha_S^i = < f^i, u_S > = 2^{-n} \sum_T f^i(T) u_S(T) =$$

$$2^{-n} \sum (f(T) - f(T \oplus i))(-1)^{|S \bigcap T|} =$$

$$2^{-n} \sum f(T)((-1)^{|S \bigcap T|} - (-1)^{|S \bigcap (T \oplus i)|}).$$

Now if $i \notin S$, the expression in the last brackets vanishes, and so does α^i_S. On the other hand if $i \in S$, then the term in the brackets becomes $2(-1)^{|S \bigcap T|}$ and $\alpha^i_S = 2\alpha_S$.

Parseval's Theorem now gives the euclidean norm of f^i.

$$\|f^i\|_2^2 = 4 \sum_{i \in S} \alpha_S^2.$$

Now we want to relate this to influence. Let β_i denote the influence of the $i-$th variable. From the definition it follows that this is the same as the fraction of sets S not containing i for which $f(S) \neq f(S \bigcup \{i\})$. Whenever this happens both $f^i(S)$ and $f^i(S \bigcup i)$ are in $\{-1, 1\}$. Consequently $\beta_i = \|f^i\|_2^2$. In other words

$$\beta_i = I_f(x_i) = 4 \sum_{i \in S} \alpha_S^2.$$

Summing this over all $1 \leq i \leq n$ we obtain:

$$\sum \beta_i = 4 \sum |S| \alpha_S^2$$

These equations suggest the following approach. Assume for a contradiction, that the β_i are small. Since

$$\sum \alpha_S^2 = \|f\|_2^2 = p$$

is given this can only happen if $\sum \alpha_S^2$ comes mainly from sets S of small cardinality. Our goal is to show this is impossible. This is achieved by proving such a result for the functions f^i. The point, to some extent, is that the f^i are assumed to have relatively small support, which should prevent their Fourier transforms from being concentrated on very small sets. The implementation of this idea is based on some elegant inequalities of Beckner [B] which we now describe.

As we have already indicated, an important feature of Beckner's method is the use of estimates for f in various L^p norms. The *two point space* X consists of the two real numbers $-1, 1$ and is equipped with the uniform probability distribution. Notice that $\mathbf{C^n} = X^n$ both as sets and as probability spaces. Consider the linear space of real functions defined on X. Every function on X is the restriction of a linear function, say, $h(x) = a + bx$.

Introduce the linear operator T_1 which maps $h(x)$ into the function $a + \epsilon bx$. It may appear mysterious at this time that such an operator should be relevant, but this will hopefully be clarified later on. We think of T_1 as operating on L^p functions and carrying them to L^2 functions.

Lemma 3.2: *The operator T_1 from L^p to L^2 has norm 1, for $p = 1 + \epsilon^2$.*

The second of Beckner's lemmas deals with the product of operators of the type considered in the previous lemma.

Lemma 3.3: *For $i = 1, 2$ let (X_i, ρ_i) and (Y_i, λ_i) be normed measure spaces and let T_i be an operator from $L^p(X_i)$ to $L^q(Y_i)$ which is an integral operator defined by a kernel, i.e.,*

$$T_i f(y) = \int_X f(x) K(x, y) d\rho_i(x).$$

Now let T be the product of these operators, mapping $L^p(X_1 \times X_2)$ to $L^q(Y_1 \times Y_2)$. If both T_1 and T_2 have norm at most 1, then so does T.

This last lemma can clearly be applied also to products of more than two spaces. In particular, multiplying the two point space by itself n times we arrive at the space $\mathbf{C^n}$. Let us evaluate the product T of n copies of the one-dimensional operators T_1.

Since the characters on $\mathbf{C^n}$ span the space of real functions on $\mathbf{C^n}$ it clearly is enough to determine their images under T. It is not hard to see that u_S is carried to $\epsilon^{|S|} u_S$. So one nice feature of the operator T is that the characters of $\mathbf{C^n}$ form a complete set of its eigenfunctions and moreover we know the corresponding eigenvalues.

Now we are in a position to (partly) demystify the connection between Beckner's work and our problems. As we explained before, our goal would be reached if we could prove theorems saying that it is impossible for most of the L^2 norm of α to be concentrated on those α_S with small $|S|$. In other words we look for upper bounds on sums such as

$$\sum_{|S| \leq b} \alpha_S^2$$

for some bound b.

Unfortunately, sums of this kind are not too convenient to work with. Alternatively, one may try and look at sums of the form

$$\sum w_S \alpha_S^2,$$

where w is an appropriately chosen weight function. Ideally, w should be 1 on sets S of cardinality at most b and 0 on larger sets. However, even a weight function which only approximates this behavior may enable us to obtain some interesting estimates. For $f = \sum \alpha_S u_S$ we know that Tf is given by:

$$Tf = \sum \epsilon^{|S|} \alpha_S u_S.$$

Denoting ϵ^2 by δ we have:

$$\|Tf\|_2^2 = \sum \delta^{|S|} \alpha_S^2.$$

This yields an estimate for sums as discussed earlier, with weight function $w_S = \delta^{|S|}$.

The other nice feature (for us) of Beckner's results is that since we are dealing with functions into $\{-1, 0, 1\}$, it is very easy to calculate their L^p norms exactly.

We apply Beckner's lemmas to our problem and arrive at the following fact which is a key to all that follows.

Lemma 3.4: *Let g be a function from $\mathbf{C^n}$ into $\{-1,0,1\}$ (for example the characteristic function of a set). Let t be the probability that $g \neq 0$ and let*

$$g = \sum \alpha_S u_S$$

be the Fourier expansion of g. Then,

$$t^{\frac{2}{1+\delta}} \geq \sum \delta^{|S|} \alpha_S^2$$

for every $0 \leq \delta \leq 1$.

We apply this Lemma with $g = f^i$. The probability that $f^i \neq 0$ is exacly β_i, and so

$$\beta_i^{\frac{2}{1+\delta}} \geq \sum \delta^{|S|} (\alpha_S^i)^2.$$

Summing this over $1 \leq i \leq n$, we have

$$\sum \beta_i^{\frac{2}{1+\delta}} \geq \sum \delta^{|S|} |S| \alpha_S^2.$$

Now ignoring the portion of the sum contributed by the sets S of cardinality exceeding b (a parameter which we shortly select), we obtain:

$$\sum \beta_i^{\frac{2}{1+\delta}} \geq \delta^b \sum_{|S| \leq b} |S| \alpha_S^2.$$

We also keep in mind that

$$p = \sum \alpha_S^2 = \alpha_\emptyset$$

which comes from $\alpha_\emptyset =< f, u_\emptyset >$ and the fact that $u_\emptyset$ is identically one. So also

$$\sum \beta_i^{\frac{2}{1+\delta}} \geq \delta^b (\sum_{|S| \leq b} \alpha_S^2 - p^2).$$

At the same time, since

$$\sum \beta_i = 4 \sum |S| \alpha_S^2$$

we also have

$$\sum \beta_i \geq b \sum_{|S| > b} \alpha_S^2.$$

Now we combine these inequalities to obtain:

$$\delta^{-b} \sum \beta_i^{\frac{2}{1+\delta}} + b^{-1} \sum \beta_i \geq \sum \alpha_S^2 - p^2$$

(3.4.1) $$= p - p^2 \geq p/2.$$

Denote $\sum \beta_i^2$ by λ^2/n where we assume

(3.4.2) $$\lambda < \frac{p \log n}{40}.$$

¿From Cauchy–Schwartz we have:

$$\sum \beta_i < \lambda.$$

Since $\frac{2}{1+\delta} < 2$ we can use the monotonicity of $r-$th power averages (e.g. [HLP p. 26]) to estimate:

$$\sum \beta_i^{\frac{2}{1+\delta}} \leq \lambda^{\frac{2}{1+\delta}} n^{-\frac{1-\delta}{1+\delta}}.$$

Choose b to be $4\lambda/p$. The second term in (3.4.1) cannot exceed $p/4$ and so we remain with:

$$\delta^{-\frac{4\lambda}{p}} \lambda^{\frac{2}{1+\delta}} n^{-\frac{1-\delta}{1+\delta}} \geq \frac{p}{4}.$$

It is now a routine matter to check that for $\delta = 1/2$, λ as in (3.4.2), any $p \leq 1/2$ and for large enough n, this inequality fails. This contradiction proves our theorem. ▮

By repeated use of this theorem we arrive at the existence of a small set of variables which dominates the function f.

Corollary 3.5: *Let f be a boolean function on n variables, let $p = \Theta(1)$ be the probability that $f = 1$ and let $\omega = \omega(n)$ be any function tending to infinity with n. Then there is a set of $\frac{n}{\log n}\omega(n) = o(n)$ variables S whose influence towards one is $p - o(1)$. This bound is tight, except for the ω term.*

4 Consequences for random walk on the cube.

The present method provides new information on the speed of convergence of random walks on the cube. We just give some indication of what we can say in this vein, leaving details and more comprehensive statements to the full paper. For simplicity (mainly to ensure ergodicity) we consider walks which on a given step move to any of the n neighbors of the current vertex v with probability $1/2n$ and otherwise remain at v. Write $f^{(t)}$ for the distribution after t steps of such a walk with initial distribution $f = f^{(0)}$, and U for the limiting (uniform) distribution.

We will be interested in convergence in the sense of L^2, rather than the more usual L^1. That is, we would like to know how slowly $\|f^{(t)} - U\|_2$ can tend to zero given various restrictions on the initial distribution $f = f^{(0)}$. (Of course, $\|f - U\|_2^2 = \|f\|_2^2 - 2^{-2n}$, and we often find it more convenient to deal with $\|f\|_2^2$.)

When $f = f_F$ is the uniform distribution on some $F \subset \mathbf{C^n}$, this question is very close to the considerations of section 3. For example, if the Fourier coefficients of $\mathbf{1}_F$ are α_S, then it is easily seen that

$$\|f^{(t)}\|_2^2 = |F|^{-2} \sum (1 - \frac{|S|}{n})^{2t} \alpha_S^2,$$

implying

$$\sum_{|S| \leq k} \alpha_S^2 \leq (1 - k/n)^{-2t} |F|^2 \|f^{(t)}\|_2^2.$$

Thus upper bounds on $\|f^{(t)}\|_2$ give upper bounds on "initial segments" of $\sum \alpha_S^2$ as needed earlier. This, in fact, was our starting point, though as it turned out the results on random walks were eventually obtained only through the above attack on the Fourier coefficients.

The most natural problem for such "semiuniform" distributions f_F is to estimate how slowly a function $t = t(n,m)$ can grow if it satisfies

$$\|f_F^{(t)}\|_2 = (1 + o(1))2^{-n} \tag{4.1}$$

for every $m-$subset F of $\mathbf{C^n}$. Intuitively, this convergence should be slower the more concentrated F is, and it is natural to expect the worst F (for given m) to be something like a ball or subcube. For example, letting $B(n,m)$ (resp. $C(n,m)$) denote the first m binary $n-$vectors in the lexicographic (resp. reverse lexicographic) order (i.e., identifying a set with its characteristic vector, $S <_L T$ if $|S| < |T|$ or $[|S| = |T|$ and $max(S \oplus T) \in T]$, while $S <_{RL} T$ if $|S| < |T|$ or $[|S| = |T|$ and $min(S \oplus T) \in S]$), we have

Theorem 4.1: *Suppose $m = m(n)$ is at least 2^{n-d}, with $d = o((n/\log n)^{1/2})$, and let $g = g^{(0)}$ be the uniform distribution on $C(n,m)$. Suppose further that $t = t(m)$ is such that*

$$\|g^{(t)}\|_2 = (1 + o(1))2^{-n.}$$

Then the same is true with g replaced by f_F for any F of size m, and in fact for any such F,

$$\frac{f_F^{(t)} - 2^{-n}}{g^{(t)} - 2^{-n}} < (1 + o(1)) \ln 4.$$

Remark 4.2: Theorem 4.1 holds for any initial f satisfying $\|f\|_2 \le (m/2^n)^{1/2}$ (i.e. the L^2-norm of f_F when $|F| = m$). ▮

Remark 4.3: For m as in the Theorem, the condition on t amounts to $t = [(1 + \epsilon)/2]n \ln d$ with $\epsilon = \omega(\frac{1}{\log d})$. ▮

Again considering $f = f_F$ we have the following natural interpretation for $\|f^{(t)}\|_2^2$. Denote by $\Psi_s(F)$ the probability that a walk starting from a randomly (uniformly) chosen point of F is again in F after the $s-$th step.

Proposition 4.4: *For F and f as above,*

$$\|f^{(t)}\|_2^2 = \Psi_{2t}(F)/(2^n|F|).$$

Given n, m and s, one may ask for (but surely not receive) the maximum of $\Psi_s(F)$ as F ranges over $m-$subsets F of $\mathbf{C^n}$. More realistically, one may hope to give bounds on this maximum which are of the correct order of magnitude. (Note that these questions are more general than that of the rate of growth of t for (4.1).) As above, one expects that balls and subcubes (the usual suspects) should come close to maximizing Ψ_s. For example, it might be true that for every F of size m

$$\Psi_s(F) = O(max\{\Psi_s(B(n,m)), \Psi_s(C(n,m))\}).$$

We can in fact show this for various ranges of the parameters (some of which, for example, will be evident from Theorem 4.3), but are apparently far from showing it in general. Although it is probably too much to expect that one of these two values always *is* the maximum, this is at least true at the outset:

Proposition 4.5: *For $s = 1, 2$ and for every n and m*

$$max\{\Psi_s(F) : |F| = m, F \subseteq \mathbf{C^n}\} = \Psi_s(C(n, m)).$$

(For $s = 1$ this is essentially the edge-isoperimetric inequality. For $s = 2$ it is a little harder, but still elementary.) It is not true that $C(n, m)$ is best for all s. (It's instructive to consider, for instance, the comparison between $B(n, n+1)$ and $C(n, n+1)$ as s grows.) What does seem possible (though for now this is little more than a guess) is that for a given n and m, $C(n, m)$ is (roughly?) optimal for s up to a certain point, after which something like $B(n, m)$ takes over.

5 Distribution of Hamming distances

A fundamental problem in Discrete Mathematics is: Given a family of binary $n-$vectors F of a given cardinality, what can be said about the distribution of Hamming distances between pairs of vectors in F? In such generality the question is, of course, quite hopeless (The whole theory of error correcting codes revolves around the more limited question of how large can the minimum distance be made.) Still, one may fruitfully study portions of the problem.

The observation which connects this problem with harmonic analysis is that if $f = \mathbf{1}_F$, the characteristic function of the family F, then the distribution of distances in F can be easily determined from $f * f$, the *convolution* of f with itself. Letting $g := f * f$, the frequency with which the vector S appears as the *mod 2* sum of pairs of vectors in F is given by $g(S)$. (The information encoded in the convolution g, is of course far more detailed then the distribution of distances.) This is a good point of view for a number of results in error correcting codes (see [MS], the standard text in this field), for instance MacWillams' formula for weight distributions of dual codes, or the inequalities underlying the Linear Programming bound. This issue will be elaborated on in the later version of this article.

We denote by $d_j = d_j(F)$, the number of ordered pairs of vectors (=sets) in F whose Hamming distance (=size of their symmetric difference) equals j. As we mentioned above this is the same as the sum of $g(S)$ over all sets S of cardinality j. We also define $\bar{d}_j$ as the number of ordered pairs $X \in F, Y \notin F$ whose distance is j, and set $d_{\leq b} := \sum_{j \leq b} d_j$, and $\bar{d}_{\leq b} = \sum_{j \leq b} \bar{d}_j$. Obviously,

$$d_j(F) + \bar{d}_j(F) = \binom{n}{j}|F|.$$

We are interested in a question which is at the other extreme from that studied in coding theory, namely we want to understand how densely packed F can be. Specifically, given the dimension n, the cardinality $m = |F|$ and a bound $b < n$ we want to determine (or estimate)

$$D(m, n, b) := max\ d_{\leq b}(F),$$

where, again, the maximum is over all families F of m binary $n-$vectors. The case $b = 1$ is again covered by the edge isoperimetric inequality, but even for $b = 2$ the question is open and an exact solution appears to be hard. Parts of this section deal with D while in others we study $\bar{D}(m, n, b) := min\ \bar{d}_{\leq b}$. This may seem strange, as the exact determination of D and $\bar{D}$ are equivalent questions (since their sum is known.) However, in most ranges of the parameters we only aim at asymptotic results, which are only of interest for the smaller of the two quantities.

We give here only a partial account of our results on this problem. It is a problem which for different ranges of the (three) parameters exhibits different optimal behavior. The results described here have been chosen to convey, according to our current understanding, some of the characteristic behavior of the quantity $D(m, n, b)$.

A large portion of the extremal theory of finite sets is devoted to proving various inequalities for which the extreme cases are well-defined. Most typically one shows the extremity of families such as cubes (e.g. for the edge isoperimetric problem), Hamming balls (for the vertex isoperimetric problem), Projective Spaces and various substructures of them. This is the case with many of the fundamental theorems in this area, for instance the Erdös–Ko–Rado and Kruskal–Katona Theorems. (Good sources for this subject are [Bo] and [Fra].) It has often proved more difficult to obtain good *estimates* in cases where there do not appear to be natural guesses as to extreme cases. We consider one of the more appealing aspects of the present work to be the fact that we do have some success in this direction.

We start with the following easy observations on $D(m, n, b)$:

- For fixed n and b, and m small enough the optimum for D is attained by making F a subset of a Hamming ball of the least possible radius.
- When $|F| = 2^{n-1}$, the optimum is attained by a subcube of dimension $n-1$. Uniqueness depends on the parity of b : for odd b the cube is the only optimum, while for even b the set of all vectors of even Hamming weight is also optimal.
- It clearly suffices to consider the range $m \leq 2^{n-1}$. We show an upper bound on $D = D(m, n, b)$, which is close to optimal when m is close enough to this upper bound.

In other words, for fixed n and b if m is small enough the optimal family is a ball in which no distance exceeds b. When m is at its maximum 2^{n-1} the cube is the best family, and near the upper bound cubes are known to be at least close to optimal.

Remark 5.1: The near optimality just mentioned is established by comparing our lower bound with the corresponding quantity for F a cube of the appropriate dimension. The question again arises as to which are the optimal families. It would be very interesting to decide whether there exist extreme families which are essentially different from both cubes and Hamming balls. At this stage we cannot even show that the optimal family is always a "weighted majority" family (i.e., the intersection of the cube and a halfspace.) See [HLL] for a case where such a result is (easily) established in a closely related situation. A standard combinatorial argument implies that there is no loss of generality in assuming F to be a shifted ideal. (See [F] for a survey of shifting.) ∎

The more substantial result of this section is a lower bound for $\bar{D}$ when n, b are fixed and m is large enough. This lower bound is shown to be near-optimal by a comparison with the case where F is a cube. This is more interesting than the third of the previously mentioned results, as in this range $\bar{D} = o(D)$.

Our first observation is that a result of Kleitman [K] settles the problem for small $m = |F|$.

Proposition 5.2: *If*

$$m \leq \sum_{j \leq \frac{b}{2}} \binom{n}{j}$$

then

$$D(m, n, b) = \binom{|F|}{2}$$

and F is optimal iff it is contained in a Hamming ball of radius $b/2$.

To proceed with our next two results we first develop a formula for $d_k = d_k(F)$ in terms of the Fourier coefficients of $f = \mathbf{1}_F$. Let $f = \sum \alpha_S u_S$, then:

$$d_k(F) = 2^n \sum \alpha_S^2 P_k(|S|)$$

where P_k is the $k-$th Kraouchuk polynomial, (e.g. [MS]) given by:

$$P_k(x) = \sum_j (-1)^j \binom{x}{j} \binom{n-x}{k-j}.$$

This formula readily supplies an answer to our problem when $|F| = 2^{n-1}$. Sum the expression for d_k over $k \leq b$ to derive a formula for $d_{\leq b}$ of the form $\sum w_{|S|} \alpha_S^2$. But $\sum_{S \neq \emptyset} \alpha_S^2$ equals $p - p^2$ (where $p = |F|/2^n$) by Parseval. Given p (and hence $\alpha_\emptyset$,) the maximum of $d_{\leq b} = \sum_{j \leq b} d_j$ is attained by making all the α_S vanish except when $w_{|S|}$ is maximal. It turns out that w_k is largest only for $k = 1$ if b is odd, and for $k = 1, n$ when b is even.

Theorem 5.3:

$$D(2^{n-1}, n, b) = 2^{n-1} \sum_{j \leq b} \binom{n-1}{j}$$

For odd b this is attained only by the $(n-1)-$dimensional cube. For even b the same holds also for the set of vectors of even Hamming weight.

The third observation is quantified as:

Proposition 5.4: *Let $p := m/2^n$. If*

$$b \log \frac{1}{p} = o(n),$$

then,

$$D(m, n, b) = (1 - o(1)) m \sum_{j \leq b} \binom{n}{j}$$

A more interesting result is that in the range considered in the previous proposition the cube is within a constant factor away of a lower bound for $\bar{D}$. Following are some remarks on the proof (which is omitted.) The previous expression for $\bar{d}_j$ may be summed to yield

$$\bar{d}_{\leq b} = \sum_S \alpha_S^2 Q_b(|S|).$$

Where $Q_b(x)$ is a polynomial of degree b. By analyzing its behavior and employing Lemma 3.4 much in the same way it was used to prove Theorem 3.1 we derive

$$\bar{d}_{\leq b} = \Omega((\log \frac{1}{p})\binom{n}{b-1}p).$$

Where the Ω expression refers to some specific absolute constant, for example $1/2$. Even values of b turn out to create some extra complication. Standard estimates for F a cube prove the complementary inequality in

Theorem 5.5: *Under the assumptions of the previous proposition*

$$\bar{D}(m,n,b) = \Theta(m(\binom{n}{b} - \binom{\log m}{b})).$$

The existence of densely packed families of sets can also be studied using the eigenvalue method. Consider the graph whose vertices are all binary strings of length n where two strings are adjacent, if their Hamming distance does not exceed b. We are studying the edge isoperimetric inequality for this graph. Again the eigenvalues of this graph may be computed using Kraouchuk Polynomials (see [MS]).

While the eigenvalue technique is in some cases quite powerful, it yields most of the time estimates much inferior to those given by the present approach. To give just one concrete example, suppose $|F| = 2^n/n$ and $b = n(1/2 - 1/\log n)$. (Appearances notwithstanding, this choice is not at all arbitrary: any "decent" upper bound on $d_{\leq b}(F)$ for such values would give the results on influence described earlier.) In this case the trivial upper bound

$$d_{\leq b}(F) \leq |F| \sum_{k \leq b} \binom{n}{k}$$

and the actual value for a subcube, viz. $|F| \sum_{k \leq b} \binom{\log |F|}{k}$, differ by a factor of about n. The bound given by the eigenvalue method differs from the trivial bound by a factor of about $\log^2 n$ (this is not "decent"), whereas our approach gets about half way to the cube, beating the trivial bound by a factor of about $\sqrt{n}$. (We still don't know the answer in this range, but believe the lower bound to be close to, if not equal to, the truth.)

6 Acknowlegements

We have had many interesting conversations on this subject with colleagues too numerous to mention, but we particularly want to express our gratitude to Dick Gundy for expanding our harmonic analytic horizons and to Benji Weiss for sharing with us his broad knowledge of mathematics.

References

[A] Alon, N.,: Private communication.

[B] Beckner, W.,: Inequalities in Fourier Analysis. *Annals of Mathematics* 102(1975), pp. 159–182.

[Bo] Bollobas, B.,: **Combinatorics** Cambridge University Press, 1986.

[BL'] Ben-Or, M., Linial, N.: Collective coin flipping, robust voting games and minima of Banzhaf values. *Proc. 26th IEEE Symp. on the Foundations of Computer Science* Portland, 1985, pp. 408–416.

[BL] Ben-Or, M., Linial, N.: Collective coin flipping. In **Randomness and Computation**, (S. Micali, ed.) Academic Press, to appear. (Also available as TR 87–2, The Leibniz Center, Computer Science Department, Hebrew University, March 1987.)

[BLS] Ben-Or, M., Linial, N., Saks, M.: Collective coin flipping and other models of imperfect information. *Rutcor Research Report 44-87, RUTCOR, Rutgers University*, December, 1987.

[CG] Chor, B., Gereb–Graus, M.: On the influence of single participants in coin flipping schemes. *Harvard University Technical report TR-06-87*, 1987.

[DF] Daykin D. E., Frankl P.: Sets of finite sets satisfying union conditions. *Mathematika* 29(1982) pp. 128-134.

[DM] Dym, H., McKean, H. P.: **Fourier Series and Integrals**, Academic Press, 1972.

[F] Frankl, P.: The shifting technique in extremal set theory, In **Surveys in Combinatorics 1987**, (C. Whitehead, ed.) London Mathematical Society Notes in Mathematics 123, Cambridge University Press, 1987.

[Fr] Frankl, P.: On the trace of finite sets. *Jour. Comb. Th. ser. A* 34(1983) pp. 41-45.

[Fra] Frankl, P.: Extremal problems for finite sets, To appear in **Handbook of Combiatorics**, (R. Graham, M. Grötschel, L. Lovász eds.)

[HLP] Hardy G. H., Littlewood, J. E, Polya, G.: **Inequalities** (2nd edition), Cambridge University Press, 1952.

[HLL] Holzman R., Lehrer E., Linial N.,: Some bounds for the Banzhaf value and other semivalues. *Math. Oper. Res.* 13(1988) pp. 358-363.

[K] Kleitman D. J.,: On a combinatorial conjecture of Erdös. *Jour. Comb. Th.* 1(1966) pp. 209-214.

[MS] MacWilliams, J., Sloane, N.,: **The Theory of Error Correcting Codes** North-Holland, 1977.