An Introduction to Analysis of Boolean functions

Mohammad Bavarian
Massachusetts Institute of Technology

1 Introduction

The focus of much of harmonic analysis is the study of functions, and operators on functions, in the setting of the Euclidean space $\mathbb{R}^n$. Although because of its relevance to physics and theory of PDEs the setting of Euclidean harmonic analysis has traditionally been the most well-studied setting of the subject, in many other situations it is important to consider operators and functions over other domains. Specifically our goal in this note is to introduce the reader to an area usually called *analysis of Boolean functions* which concerns itself with the analysis of functions over the discrete hypercube $\{0,1\}^n$. The harmonic analysis of functions $f : \{0,1\}^n \to \{-1,1\}$ has in fact many applications to various branches of theoretical computer science and combinatorics and it is already a rich theory with many non-trivial results, techniques and open problems. (See [8, 10] for a short discussion of the applications to theoretical computer science as well as social choice theory and threshold phenomenon in combinatorics. See [9] for a more extensive exposition.)

In this note however instead of trying to survey the applications, we focus our attention on the proof of one of the classical and most influential theorems in this area obtained by Kahn-Kalai-Linial (KKL) [6] in the 80's. To do so, we start by giving some basic definitions and notations in Section 2. In Section 3, we state the KKL theorem and the Friedgut's junta theorem. In Section 4, we introduce the important *noise operator*[1] and explain some of its properties. In Section 5, we finally use the properties of the noise operator (specifically the hypercontractivity theorem) to prove Friedgut's main lemma which gives us the KKL theorem and Friedgut's junta theorem as corollaries.

2 Basic Facts and Notations

Consider the vector space of functions $f : \{0,1\}^n \to \mathbb{R}$. For each $S \subseteq [n]$ we have a Fourier character $\chi_S : \{0,1\}^n \to \{-1,1\}$ given by

$$\chi_S(x) = (-1)^{\sum_{i \in S} x_i}.$$

So for example $\chi_\varnothing$ is the constant 1 function and $\chi_{[n]}$ is the PARITY function which specifies whether a binary string $x \in \{0,1\}^n$ has an odd or an even hamming weight.

[1] This operator is essentially equivalent to the heat evolution operator—or random walk operator—on the hypercube with the usual hammming graph structure.

We turn the space of real-valued functions over $\{0,1\}^n$ into an *inner product space* by equipping it with the uniform measure:

$$\langle f, g\rangle = \mathop{\mathbf{E}}_{x\in\{0,1\}^n} f(x)g(x) = \frac{1}{2^n}\sum_x f(x)g(x).$$

It is easy to see that the Fourier characters $\{\chi_S\}_{S\subseteq[n]}$ form an orthonormal basis for this space and we have the Fourier inversion formula,

$$\widehat{f}(S) = \mathop{\mathbf{E}}_x f(x)\,\chi_S(x) \quad\iff\quad f(x) = \sum_{S\subseteq[n]} \widehat{f}(S)\,\chi_S(x),$$

and the Parseval's identity,

$$\mathop{\mathbf{E}}_x f(x)^2 = \sum_S \widehat{f}(S)^2.$$

Directional derivatives and influences. The geometry of discrete hypercube which allow us to define the notion of derivatives is given by the usual of Hamming structure of $\{0,1\}^n$ in which each point $x\in\{0,1\}^n$ has n *neighbors* in directions 1 to n given by the n points which differ with x in only one single coordinate.

Given $x\in\{0,1\}^n$ we denote by x^i the neighbor of x in direction i. We define

$$\partial_i f(x) = \frac{f(x)-f(x^i)}{2},$$

and so we have

$$\partial_i f(x) = \sum_{S\ni i} \widehat{f}(S)\,\chi_S(x).$$

Definition 2.1 (The i^{th} influence)**.** *The influence of direction i in the value of a Boolean function $f:\{0,1\}^n\to\{-1,1\}$ is the probability that for a random $x\in\{0,1\}^n$ we have $f(x)\neq f(x^i)$. More generally, for a function $f:\{0,1\}^n\to\mathbb{R}$ we define*

$$\mathrm{Inf}_i(f) := \|\partial_i f(x)\|_2^2 = \sum_{S\ni i}\widehat{f}(S)^2\,, \tag{1}$$

where the last equality follows by Parseval.

Lemma 2.2 (Poincare's inequality)**.** *For any $f:\{0,1\}^n\to\mathbb{R}$ we have*

$$\mathrm{var}(f) \leqslant \sum_{i=1}^n \mathrm{Inf}_i(f).$$

The quantity in the RHS of the above equation is usually referred to as *the total influence* of f and is analogous to the L^2 norm of ∇f from the Euclidean harmonic analysis.

Proof of Lemma 2.2. We have

$$\mathrm{var}(f) = \mathbf{E}[f^2]-\mathbf{E}[f]^2 = \sum_S \widehat{f}(S)^2 - \widehat{f}(\varnothing)^2 = \sum_{S\neq\varnothing}\widehat{f}(S)^2.$$

On the other hand,

$$\sum_{i=1}^n \mathrm{Inf}_i(f) = \sum_{S\subseteq[n]} |S|\,\widehat{f}(S)^2 \geqslant \sum_{S\neq\varnothing}\widehat{f}(S)^2. \tag{2}$$

□

3 The KKL Theorem and Friedgut's Junta Theorem

Theorem 3.1 (KKL)**.** *For any function $f : \{0,1\}^n \to \{-1,1\}$ there exists $i \in [n]$ such that*

$$\mathrm{Inf}_i(f) = \Omega\left(\frac{\mathrm{var}(f)\log n}{n}\right).$$

To get some perspective on the above theorem, note that from Poincare's inequality it follows that for a *typical direction* $i \in [n]$ we have

$$\mathrm{Inf}_i(f) \geqslant \Omega\left(\frac{\mathrm{var}(f)}{n}\right).$$

The above theorem says that the above estimate can be improved substantially for some direction i.

Remark 3.2. *The Boolean valuedness hypothesis in the statement of the Theorem 3.1 is essential as the example of the function $g = \frac{1}{\sqrt{n}}\sum_{i=1}^n \chi_i$ shows.* [2]

We prove the KKL theorem as a corollary of the following powerful lemma of Friedgut [4].

Lemma 3.3 (Friedgut's main lemma)**.** *Assume $f : \{0,1\}^n \to \{0,1\}$ has total influence $I[f] = \sum_{i=1}^n \mathrm{Inf}_i(f)$. Let $d = I[f]/\epsilon$. Let*

$$J = \{i \in [n] : \mathrm{Inf}_i(f) \geqslant d^{-2}9^{-d}\}.$$

Then if one defines

$$f_{junta} = \sum_{\substack{S\subseteq J\\ |S|\leqslant d}} \widehat{f}(S)\,\chi_S(x)$$

we $\|f - f_{junta}\|_2^2 \leqslant 2\epsilon$.

The interesting setting of parameters for Friedgut's main lemma is when $d = I[f]/\epsilon \ll \log n$ in which case the set J is not too large, i.e. $|J| \leqslant I[f]d^2 9^d \leqslant d^3 9^d \ll n$. This means that f_{junta} (and by proxy f) is determined by a small number (much smaller than n) of coordinates.

There is potentially one small downside to the above lemma; as stated the lemma sends us out of the category of Boolean functions. However note that this can be easily fixed since $\|f - \mathrm{sgn}(f_{junta})\|_2^2 \leqslant 8\epsilon$. The fact that any Boolean function with low influence is close to a *Boolean function* depending on few coordinates is the main Theorem of [4] proved as an immediate consequence of Lemma 3.3.

3.1. From Friedgut's main lemma to the KKL theorem

To prove the KKL theorem via the Friedgut's main lemma we consider two cases:

(i) Either $\mathrm{Inf}(f) \geqslant \frac{\mathrm{var}(f)\log n}{100}$

[2] An important feature of a large number of interesting theorems in analysis of Boolean functions is that they manage to extract non-trivial information from the seemingly hard to use hypothesis of Boolean valuedness.

(ii) Or else $\mathrm{Inf}(f) < \frac{\mathrm{var}(f)\log n}{100}$.

In the first case the result holds for a typical $i \in [n]$ by Lemma 2.2. In the **second case** we apply Friedgut's main lemma. Let $\epsilon = \frac{\mathrm{var}(f)}{4}$ and $d = \frac{\mathrm{Inf}(f)}{\epsilon}$ as in Lemma 3.3. We see

$$\mathrm{var}(f_{junta}) \geqslant \mathrm{var}(f) - 2\epsilon = \frac{\mathrm{var}(f)}{2}.$$

Applying Poincare's inequality to f_{junta} and noting that $\mathrm{Inf}_i(f_{junta}) = 0$ for all $i \notin J$ we see

$$\sum_{i\in J} \mathrm{Inf}_i(f_{junta}) \geqslant \mathrm{var}(f_{junta}) \geqslant \frac{\mathrm{var}(f)}{2}$$

which implies that there exists $i \in J$ such that $\mathrm{Inf}_i(f_{junta}) \geqslant \frac{\mathrm{var}(f)}{2|J|}$. Moreover, we know that

$$|J|d^{-2}9^{-d} \leqslant \sum_{i=1}^{n} \mathrm{Inf}_i(f) < \frac{\mathrm{var}(f)\log n}{100}$$

which mean that $\frac{\mathrm{var}(f)}{2|J|} > \frac{50}{d^2 9^d \log n}$. Since $d = \frac{4\mathrm{Inf}(f)}{\mathrm{var}(f)} < \frac{\log n}{25}$ some computation implies that $\frac{\mathrm{var}(f)}{2|J|} \geqslant \frac{1}{\sqrt{n}}$. So we have established that

$$\exists i \in J \ : \mathrm{Inf}_i(f_{junta}) \geqslant \frac{1}{\sqrt{n}}.$$

However, since

$$\mathrm{Inf}_i(f_{junta}) = \sum_{\substack{i\in S\subseteq J\\ |S|\leqslant d}} \widehat{f}(S)^2 \leqslant \sum_{S\ni i} \widehat{f}(S)^2 = \mathrm{Inf}_i(f)$$

this finishes the proof in the second case as well.

4 Noise Operator and Hypercontractivity

A crucial instrument in our analysis is the bound on $L^{q\to p}$ norm of the following Fourier multiplier operator

$$T_\rho f = \sum_{S\subseteq[n]} \rho^{|S|}\widehat{f}(S)\chi_S(x).$$

Since the above is a Fourier multiplier, in the physical space it corresponds to a convolution operator. In other words we have

$$T_\rho f(x) = f * K_\rho(x) = \mathop{\mathbf{E}}_{y\sim_\rho x}[f(y)],$$

where $y \sim_\rho x$ means that y is chosen by taking $y_i = x_i$ with probability $\frac{1+\rho}{2}$ and $y_i = 1 - x_i$ with probability $\frac{1-\rho}{2}$ independently for each coordinate (K_ρ is the kernel associated to the above process).

Hence, we see that the noise operator is really an averaging operator and hence we expect it to have an smoothing effect on functions. This is captured by the following theorem which we state here without proof.

Theorem 4.1 (Hypercontractivity theorem, see [9]). *Let $1 \leqslant p \leqslant q$ such that $(q-1) \leqslant (p-1)\rho^{-2}$ (i.e. q is not too huge) then for all functions $f : \{0,1\}^n \to \mathbb{R}$,*

$$\|T_\rho f\|_q \leqslant \|f\|_p.$$

Let us make a couple of remarks regarding the proof of Theorem 4.1. First of all, we note that in our application we do not need the above theorem in full generality. In fact the special choice of $(q = 2, p = \frac{4}{3}, \rho = \frac{1}{\sqrt{3}})$ will suffice for us. Secondly, the proof of Theorem 4.1 even in full generality is by no means difficult. The standard proof goes by induction on n establishing the case n from the cases $n-1$ and 1 using Minkowski's inequality. The case $n = 1$ itself is elementary and follows by a careful calculus argument.

In the next section we apply the hypercontractivity theorem to prove Lemma 3.3.

5 Proof of Friedgut's main lemma

We have the following claim about the truncated sum of the L^2 Fourier mass of $\partial_j f$.

Proposition 5.1. *Let $f : \{0,1\}^n \to \{-1,1\}$ be a Boolean function and let d be a positive integer. For any $j \in [n]$ we have*

$$\sum_{\substack{S \ni j \\ |S| \leqslant d}} \widehat{f}(S)^2 \leqslant 3^d \mathrm{Inf}_i(f)^{3/2}.$$

Proof. Using $(q = 2, p = \frac{4}{3}, \rho = \frac{1}{\sqrt{3}})$ hypercontractivity we have

$$\sum_{\substack{S \ni j \\ |S| \leqslant d}} \widehat{f}(S)^2 \leqslant 3^d \sum_{\substack{S \ni j \\ |S| \leqslant d}} \left(\frac{1}{3}\right)^{|S|} \widehat{f}(S)^2 = 3^d \|T_{\frac{1}{\sqrt{3}}} \partial_j f\|_2^2 \leqslant 3^d \|\partial_j f\|_{4/3}^2.$$

Since f is Boolean, $\partial_j f$ is $\{-1,0,1\}$ valued and hence

$$\|\partial_j f\|_{4/3}^2 = \left(\mathbf{E}(\partial_j f)^{4/3}\right)^{3/2} = \left(\mathbf{E}(\partial_j f)^2\right)^{3/2} = \mathrm{Inf}_j(f)^{3/2}.$$

This finishes the proof. $\square$

From this Friedgut's lemma follows quickly.

Proof of Lemma 3.3. We decompose the Fourier expression for f into three orthogonal parts,

$$f = f_{high} + f_{junta} + f_{psr},$$

with $f_{high} = \sum_{|S|>d} \widehat{f}(S)\chi_S$, $f_{psr} = \sum_{\substack{|S| \leqslant d \\ S \nsubseteq J}} \widehat{f}(S)\chi_S$, and $f_{junta} = \sum_{\substack{|S| \leqslant d \\ S \subseteq J}} \widehat{f}(S)\chi_S$. We show $\|f_{high}\|_2^2 \leqslant \epsilon$ and $\|f_{psr}\|_2^2 \leqslant \epsilon$ to prove the main result $\|f - f_{junta}\|_2^2 \leqslant 2\epsilon$. We have

$$d\|f_{high}\|_2^2 = d \sum_{|S|>d} \widehat{f}(S)^2 \leqslant \sum_S |S|\widehat{f}(S)^2 = \sum_{i=1}^n \mathrm{Inf}_i(f) = I[f].$$

Since $d = I[f]/\epsilon$ by definition, the first claim follows. For the second claim we have

$$\|f_{psr}\|_2^2 = \sum_{\substack{S \not\subset J \\ |S| \leqslant d}} \widehat{f}(S)^2 \leqslant \sum_{|S| \leqslant d} |S \backslash J| \cdot \widehat{f}(S)^2 = \sum_{j \notin J} \sum_{\substack{S \ni j \\ |S| \leqslant d}} \widehat{f}(S)^2.$$

The last term above expression is bounded by $3^d \mathrm{Inf}_j(f)^{3/2}$ by Proposition 5.1. Hence,

$$\|f_{psr}\|_2^2 \leqslant 3^d \sum_{j \notin J} \mathrm{Inf}_j(f)^{3/2} \leqslant 3^d \sum_{j=1}^{n} \mathrm{Inf}_j(f) \cdot \max_{j \notin J} \sqrt{\mathrm{Inf}_j(f)}.$$

By definition of J we have $\max_{j \notin J} \sqrt{\mathrm{Inf}_j(f)} \leqslant d^{-1} 3^{-d}$. Since $d = \frac{I[f]}{\epsilon}$ the result follows. □

It is interesting to reflect why the hypercontractivity theorem has any bearing on the KKL theorem—though the latter has no natural dynamic or operation associated to it. The main point I believe is that the hypercontractivity theorem really works as a tool to allow us to extract some useful information from the Boolean-valuedness of the function f by connecting the high L^p norms with the low L^p norms. In fact in the above argument, the only place that the Boolean-valuedness of f was used in any essential way was in Proposition 5.1 where we used the equality $\|\partial_i f\|_{4/3}^2 = \|\partial_i f\|_2^2$.

6 Concluding Remarks

In this short note our aim was to convey the flavor of some the techniques and ideas that are widely used in the analysis of Boolean functions. As such we focused on perhaps the earliest (and in our view beautiful) application of the hypercontractivity theorem—that is the KKL theorem.

There are variety of open problems in the area (some surveyed at [7]) that however seem beyond the reach of the current techniques. In my view, ideas from the Euclidean harmonic analysis, especially the properties of operators other than T_ρ which like T_ρ have some physical significance (note that T_ρ is essentially the random walk or heat diffusion operator over the hypercube), could be useful to make progress on some of these problems.

To conclude I finish by recounting some of my favorite open problem from the area.

Conjecture 6.1 (Aaronson-Ambainis)**.** *Let $f : \{0,1\}^n \to [-1,1]$ be a function with Fourier degree $\leqslant k$, i.e. $\widehat{f}(S) = 0$ for all $|S| > k$. Prove that there exists $i \in [n]$ such that*

$$\mathrm{Inf}_i(f) \geqslant \left(\frac{\mathrm{var}\, f}{k}\right)^{O(1)}.$$

An exponentially weaker bound of the form $\left(\frac{\mathrm{var}\, f}{\exp(k)}\right)^{O(1)}$ is known [3].

Conjecture 6.2 (Sensitivity Conjecture)**.** *Define the maximum sensitivity of a function $f : \{0,1\}^n \to \{-1,1\}$ as the L^∞ norm of its total derivative, i.e.*

$$s(f,x) = \sum_{i=1}^{n} |\partial_i f(x)|, \qquad s(f) = \max_{x \in \{0,1\}^n} s(f,x).$$

Prove that

$$\deg(f) = O(s(f)^{O(1)}).$$

The conjecture is almost 20 years old and is considered difficult; see [5] for a survey. The best current bound is of the form $\deg(f) \leqslant 2^{s(f)+O(\sqrt{s(f)})}$ which follows from the recent work [1].

References

[1] Ambainis, A., M. Bavarian, Y. Gao , J. Mao, X. Sun, and S. Zuo. *Tighter relations between sensitivity and other complexity measures.* In Proceedings of Automata, Languages, and Programming Conference (ICALP), 2014.

[2] Buhrman, H. and R. De Wolf. *Complexity measures and decision tree complexity: a survey.* Theoretical Computer Science 288.1 (2002): 21-43.

[3] Dinur, I., E. Friedgut, G. Kindler, and R. O'Donnell. *On the Fourier tails of bounded functions over the discrete cube.* Israel Journal of Mathematics, 160(1), 389-412 (2007).

[4] Friedgut, E. *Boolean functions with low average sensitivity depend on few coordinates.* Combinatorica, 18(1), 27-35, 1998.

[5] Hatami, P., R. Kulkarni and D. Pankratov. *Variations on the Sensitivity Conjecture.* Theory of Computing, Graduate Surveys, 2, 1-27 (2011).

[6] Kahn, J., G. Kalai, and N. Linial. *The influence of variables on Boolean functions.* IEEE 54th Annual Symposium on Foundations of Computer Science (FOCS), 1988.

[7] Mossel, E., K. Oleszkiewicz, and E. O'Donnel (editors). *Open problems in Analysis of Boolean Functions.* The first Simons Symposium on Analysis of Boolean functions. Available at `http://www.cs.cmu.edu/~odonnell/papers/aobf-open-problems.pdf`.

[8] O'Donnell, R. *Some topics in analysis of Boolean functions.* Proceedings of the fortieth annual ACM symposium on Theory of computing (STOC), 2008.

[9] O'Donnell, R. *Analysis of boolean functions.* Cambridge University Press, 2014.

[10] De Wolf, R. *A Brief Introduction to Fourier Analysis on the Boolean Cube.* Theory of Computing, Graduate Surveys 1 (2008): 1-20.