Analysis of Boolean Functions

Yuval Filmus

February 22, 2015

1 Plan of Talk

- What is Analysis of Boolean Functions?
- Two examples (or only one example?).
- My work.

2 What is Analysis of Boolean Functions?

A *Boolean function* is a function $f\colon D \to \{0,1\}$. Usually the domain D is finite, and has an underlying probability measure (often the uniform measure over D). We can identify a Boolean function on D with a subset of D, a connection underlying combinatorial applications of the area.

Analysis of Boolean functions is a body of techniques and results aimed at studying (usually) Boolean functions using spectral techniques. The original motivation comes from extremal combinatorics and voting theory, but the area has become crucial for parts of theoretical computer science, and has important applications elsewhere (for example, Friedgut's sharp threshold theorem is important for random graph theory).

There are two basic ideas in the area.

Extending the range. We can view a Boolean function $f\colon D \to \{0,1\}$ as a real-valued (or sometimes complex-valued) function $f\colon D \to \mathbb{R}$ whose range is $\{0,1\}$. This allows us to use linear algebra to reason about f. Classically, the domain D is an Abelian group (most often the cube $\mathbb{Z}_2^n$), and f is analyzed through its Fourier expansion.

Extending the domain. One of the reasons finite combinatorics is difficult is that analysis is harder to apply on discrete domains. Sometimes it is possible to find a continuous domain D' extending D, and a lift $f'\colon D' \to \mathbb{R}$ of f, such that properties of f' correspond to properties of f. The classical example is studying the binomial distribution via the central limit theorem. A vast generalization of this result, called the *invariance principle*, has proved influential in the field. When $D = \mathbb{Z}_2^n = \{\pm 1\}^n$, the domain D' is $\mathbb{R}^n$ equipped with a Gaussian distribution.

3 Two examples

3.1 Erdős–Ko–Rado theory

Lovász in his paper on the Shannon capacity of the pentagon introduced the θ function, a spectral tool which can be used to solve problems in extremal combinatorics (an even earlier paper is due to Hoffman, working on the chromatic number). He gave a spectral proof of the Erdős–Ko–Rado theorem, a classical result in extremal combinatorics:

Suppose F is a collection of subsets of $\{1,\ldots,n\}$ of size k in which any two subsets intersect. If $k < n/2$ then $|F| \leq \binom{n-1}{k-1}$, with equality only for the families S_i ("stars") consisting of all sets containing i.

Lovász defined a matrix A whose rows and columns are indexed by k-subsets of $\{1,\ldots,n\}$ such that $A_{ST} = 1$ whenever S, T intersect, and the spectral radius of A is $\binom{n-1}{k-1}$. Let f be the characteristic vector of an intersecting family. On the one hand, $f'Af = |F|^2$. On the other hand, $f'Af \leq \binom{n-1}{k-1} f'f = \binom{n-1}{k-1}|F|$. This shows that $|F| \leq \binom{n-1}{k-1}$. Moreover, if $|F| = \binom{n-1}{k-1}$ then f must lie in the eigenspace of $\binom{n-1}{k-1}$. This eigenspace is spanned by all stars $S_1, \ldots, S_n$, and this implies that F is a star.

Later on, Friedgut showed how to prove a stronger version of the same theorem using the same argument. The stronger version states that if $|F| \approx \binom{n-1}{k-1}$ then F is close to a star. Indeed, if $|F| \approx \binom{n-1}{k-1}$ then f is close to a function in the eigenspace of $\binom{n-1}{k-1}$, and this implies that F is close to a star. We call this property "stability".

While the classical Erdős–Ko–Rado theorem itself has completely combinatorial proofs (even the stability result had been proved by Frankl), the only known proof of some similar results involves a spectral argument. Such results include versions of the Erdős–Ko–Rado theorem for vector spaces, for (multiply intersecting) permutations, and for (triangle-intersecting) graphs (due to Ellis, Friedgut and myself). In all these cases, stability comes for free.

3.2 Majority is the Stablest [Remove]

Consider an elections between two candidates. If there are n voters, we can think of the rules of the elections as a function $f\colon \{0,1\}^n \to \{0,1\}$ satisfying $f(1-x) = 1 - f(x)$. Some noise is inherent in the process of eliciting votes, and we would like to minimize this effect. What voting rule should we use?

One way to formalize this question is through the notion of *noise sensitivity*. Let x be a uniformly random vote, and let y be obtained from x by flipping the vote of each voter with probability p. The p-sensitivity of f is $S_p(f) = \Pr[f(x) \neq f(y)]$.

It turns out that the voting rules minimizing $S_p(f)$ are $f(x) = x_i$, which have sensitivity p ("Dictator is Stablest"). This rule is problematic, of course. What happens if we insist that no voter is too "influential"? In that case the best voting rule is majority vote, which has sensitivity $\frac{1}{\pi}\arccos(1-2p) \approx \frac{2}{\pi}\sqrt{p}$ ("Majority is Stablest").

This result was the motivation behind the invariance principle. The principle allows us to deduce Majority is Stablest from a corresponding result in Gaussian space ($\mathbb{R}^n$ equipped with an independent multivariate Gaussian distribution). The corresponding result, Borell's isoperimetric theorem, states that the measure-1/2 sets in Gaussian space which are least noise sensitive are halfspaces. In contrast to the discrete case, in this result there is no need to explicitly rule out dictatorships. Borell's theorem can be proved using symmetrization, a technique which cannot be applied in this case directly on $\{0,1\}^n$. (I should mention that more recent proofs of Majority is Stablest are direct and don't involve the invariance principle.)

While the conjecture Majority is Stablest arose in voting theory, what motivated Mossel, O'Donnell and Oleszkiewicz to solve it was an application to theoretical computer science, specifically to hardness of approximation of MAX-CUT. The field of hardness of approximation has seen many applications of analysis of Boolean functions. *Should I mention anything more?*

4 Non-product Domains

Traditionally, the functions being considered come from product domains, in fact usually the Boolean cube $\{0,1\}^n$. The product nature of the domain and of the corresponding probability measures often plays a key role in the proof of theorems in the area.

Some problems arising in extremal combinatorics and in theoretical computer science involve functions over non-product domains, the two main examples being the symmetric group and the "slice" (the set of all elements of $\{0,1\}^n$ of some fixed weight k). My recent research has focused on generalizing classical theorems in the field to these exotic domains.

4.1 Key Theorems

Here are some results one would like to generalize.

Friedgut–Kalai–Naor Easy version: Suppose $f\colon \{0,1\}^n \to \{0,1\}$ is of the form $c_0 + \sum_i c_i x_i$. Then f depends on at most one coordinate.

Hard version: If f is *close* to a function of the form $c_0 + \sum_i c_i x_i$ (in an L2 sense) then f is *close* to a function depending on at most one coordinate.

The FKN theorem, or rather its version for the domain $\binom{[n]}{k}$, allows us to prove the stability version of the Erdős–Ko–Rado theorem mentioned above.

Kindler–Safra Easy version: If $f\colon \{0,1\}^n \to \{0,1\}$ is a degree d polynomial then f depends on C_d coordinates.

Hard version: If f is close to a degree d polynomial then f is close to a function depending on C_d coordinates.

This allows us to prove stability in the context of triangle-intersecting families of graphs.

Friedgut's theorem Suppose A is a subset of $\{0,1\}^n$. The *edge boundary* of A is the set of edges (x,y) such that $x \in A$ and $y \notin A$. We usually normalize this quantity by dividing by 2^n. One way to obtain a small edge boundary is by having A depend only on a small number of coordinates: if A depends on c coordinates, then the (normalized) edge boundary is at most c. Friedgut's theorem states that if the edge boundary of A is $O(1)$ then A is close to a set B depending on $O(1)$ coordinates.

Friedgut's theorem is useful in theoretical computer science. In some situations, it allows us to show that *every* monotone function is a junta when looked at from the correct angle.

Invariance principle Let $X_1, \ldots, X_n \sim \{-1,1\}$. Effective versions of the central limit theorem show that $\sum_i c_i X_i$ behaves like a Gaussian with mean zero and variance $\sum_i c_i^2$, as long as no c_i is too "prominent". The invariance principle generalizes this to low degree polynomials. If P is a low degree multilinear polynomial in which no variable is "prominent" then $P(X_1, \ldots, X_n)$ behaves similarly to $P(G_1, \ldots, G_n)$, where $G_1, \ldots, G_n \sim N(0,1)$.

4.2 Our results

We can summarize our results neatly in a table:

What	Slice	Symmetric group
FKN	Filmus	Ellis, Filmus, Friedgut
KS	?	Ellis, Filmus, Friedgut
Junta	Wimmer; Filmus	
Invariance	Filmus, Kindler, Mossel, Wimmer	

The empty spaces indicate results which might be false. The question mark indicates a result which is certainly true, but we are still working on a proof.

4.3 Challenges and how to overcome them

The main challenge in proving theorems in non-product domains is the fact that there is no canonical spectral basis. For a product domain such as $\{0,1\}^n$, such a canonical basis is the Fourier basis, given by the Fourier characters. In the context of the symmetric group, the counterpart is the representation theory of the symmetric group, which partitions the space of all functions into subspaces of dimension larger than 1. Several orthogonal bases exist, but there are no canonical bases in which all coordinates play the same role. In the context of the slice there is similar representation theory (corresponding to the Johnson association scheme), but again no canonical basis.

Other challengs depend on the statement being proved. The proof of some statements uses a property of $\{0,1\}^n$ known as *hypercontractivity*, which fails for the symmetric group. The proof of the invariance principle uses the exchange method in which we replace X_i by G_i one variable at a time; this strongly uses the product nature of the domain.

We have used three different approaches to overcome these difficulties, sometimes in tandem:

1. Reduction to a result on a product domain.
2. Use a canonical spanning set instead of a canonical basis.
3. Use a non-canonical basis.

5 Harmonic polynomials [Option 1]

To finish, let me illustrate the third approach by describing a new orthogonal basis for the slice (or rather, the Johnson scheme), which also doubles as an explicit basis of eigenvectors for the Johnson and Kneser graphs.

While proving the invariance principle on the slice, we were faced with the following problem. Given a function on the slice $\binom{[n]}{k}$, what is the "correct" way to lift it to a function on other slices, on $\{0,1\}^n$ (with a skewed product measure), on Gaussian space?

Every function on $\{0,1\}^n$ can be represented uniquely as a multilinear polynomial, which is closely related to its Fourier expansion. The classical invariance principle shows that the "correct" way to lift the function to Gaussian space is by interpreting the multilinear polynomial as a function on Gaussian space.

Similarly, a classical fact due to Dunkl shows that every function on the slice $\binom{[n]}{k}$ (for $k \leq n/2$) can be represented uniquely as a multilinear polynomial of degree at most k which is annihilated by the operator $\sum_{i=1}^n \frac{\partial}{\partial x_i}$ ("harmonic functions"). This somewhat strange condition comes up naturally from the representation theory of the slice. What kinds of functions are harmonic? $x_1 - x_2$ and $(x_1 - x_2)(x_3 - x_4)$ are harmonic. The function x_1 is not.

Many different bases for the slice appear in the literature, but to the best of my knowledge, none of them is orthogonal. The basis we have constructed is orthogonal at one and the same time with respect to the uniform measure on *all* slices, and beyond!

Let $A = a_1, \ldots, a_\ell$ and $B = b_1, \ldots, \ell$ be two sequences of distinct numbers in $\{1, \ldots, n\}$. We say that $A < B$ if $b_1 < \cdots < b_\ell$ and $a_i < b_i$. Our basis is parametrized by sets B for which $A < B$ for some A. The corresponding basis element is

$$\chi_B = \sum_{A<B} (x_{a_1} - x_{b_1}) \cdots (x_{a_\ell} - x_{b_\ell}).$$

This basis is orthogonal with respect to *all* exchangeable measures. By explcitly computing the norms of basis elements with respect to different measures, we can show for example that if we lift a function from $\binom{[n]}{k}$ to $\binom{[n]}{t}$ (where $k \approx t$), then the functions have almost the same norms. This observation forms the basis of our proof of the invariance principle for the slice.

6 EFF2 [Option 2]

To finish, let me describe one of our results more fully, FKN for balanced functions on the symmetric group. We can think of the symmetric group as the group of permutation matrices (x_{ij}). The "easy" version, proved by Ellis, Friedgut and Pilpel, states that if $f \colon S_n \to \{0,1\}$ has the form $f = \sum_{ij} a_{ij} x_{ij}$ then either $f = \sum_{j \in J} a_{ij}$ for some i, J or $f = \sum_{i \in I} a_{ij}$ for some I, j. In other words, f either depends on the image of some point i or on the inverse image of some point j. We call such a function a *dictatorship*.

The corresponding "hard" version states that if f is "balanced" (e.g. $1/3 \leq \mathbb{E}[f] \leq 2/3$) and close to $\sum_{ij} a_{ij} x_{ij}$ then f is close to a dictatorship. This only works when f is balanced: the function $f = x_{11} + x_{22} - x_{11}x_{22}$ is Boolean and close to the linear function $x_{11} + x_{22}$ but is not close to a dictatorship (in the relative sense); this function is heavily skewed to 0.

Let $f_1 = \sum_{ij} a_{ij} x_{ij}$. The idea of the proof is to construct a matrix (b_{ij}) such that $f_1(\pi) = \sum_{i=1}^n b_{i\pi(i)}$ ("generalized diagonals"). Unfortunately I don't have time to describe the entire proof, so let me just describe one step. Consider an $n \times n$ matrix consisting of 0s and 1s. Easy version: If every generalized diagonal contains exactly a single 1, then the matrix contains a 1 row or a 1 column. Hard version: If most generalized diagonals contain exactly a single 1, then the matrix contains a row or a column consisting mostly of 1s.