Boolean function analysis beyond the Boolean cube

Yuval Filmus

Technion

5 March 2018
Institute for Advanced Study

Plan for the talks

- This talk: Introduction and applications.
- Tomorrow: In-depth survey.

Classical Boolean function analysis

- Central object of study: real-valued functions on the Boolean cube $\{\pm 1\}^n$.
- Often, but not always, the functions themselves are also Boolean.
- The Boolean cube can be regarded as a Cayley graph of $\mathbb{Z}_2^n$, a distance-regular graph, a differential poset, ...
- Distance-regularity: $\#\{z : d(x,z) = a, d(y,z) = b\}$ depends only on $d(x,y)$.

Fundamental theorem of Boolean function analysis

Fundamental theorem of Boolean function analysis

Every function $f\colon \{\pm 1\}^n \to \mathbb{R}$ has a unique expansion as a multilinear polynomial in the n inputs $x_1, \ldots, x_n \in \{\pm 1\}$.

- The coefficients of the monomials are the *Fourier coefficients*.
- $\hat{f}(S)$ is coefficient of $x_S := \prod_{i \in S} x_i$.
- The monomials form an orthonormal basis with respect to the inner product $\langle f, g \rangle = \mathbb{E}[fg]$.
- The monomials are also all characters of $\mathbb{Z}_2^n$.
  - Important for linearity testing.

Degree

Spectral degree:

- The degree of the unique representation is the *degree* of f.
- Aside: $\deg f = O((\widetilde{\deg} f)^6)$ for Boolean f.

Spatial degree:

- A d-junta is a function depending on d coordinates.
- $\deg f \leq d$ iff f is a linear combination of d-juntas.

Influence

Spatial definition:

- Laplacian: $Lf(x) = \frac{1}{2}\sum_{y\sim x}[f(x) - f(y)]$.
- Total influence: $\mathrm{Inf}[f] = \langle f, Lf\rangle$.

Spectral formula:

$$Lf = \sum_S |S|\hat{f}(S)x_S = \sum_d df^{=d}.$$

- At least $\mathbb{V}[f]$ (Poincaré inequality).
- At most $(\deg f)\mathbb{V}[f]$
- At most $\deg f$ if f is Boolean.

Also have influences in given direction (= generator).

Noise

Spatial definition:

- Markov process on the cube: flip each coordinate with rate 1.
- $N_\rho(x)$ is state after $\frac{1}{2}\ln\frac{1}{\rho}$ steps, starting at x.
- $T_\rho f(x) = \mathbb{E}[f(N_\rho(x))]$.
- Noise stability: $\mathrm{Stab}_\rho[f] = \langle f, T_\rho f\rangle$.

Spectral formula:

$$T_\rho f = \sum_S \rho^{|S|}\hat{f}(S)x_S = \sum_d \rho^d f^{=d}.$$

Coarse decomposition

Spectral definition:

- $f^{=d}$ is homogeneous degree d part of Fourier expansion of f.
- $f = \sum_{d=0}^{n} f^{=d}$. Orthogonal decomposition of $\mathbb{R}[\{\pm 1\}^n]$.

Spatial definition:

- Function is homogeneous degree d if it has degree d and is orthogonal to all $(d-1)$-juntas.
- Theory of differential posets: there exists unique decomposition $f = \sum_d f^{=d}$.

Theory of association schemes: If $A(x, y)$ depends only on $d(x, y)$ then

$$Af = \sum_d \lambda_d f^{=d}.$$

Examples: Lf, $T_\rho f$.

Structure theorems

Notions of simplicity for Boolean functions:

- d-junta.
- Degree d.
- Total influence d.

Fundamental theorems:

- FKN: Almost degree 1 $\longrightarrow$ almost 1-junta.
- Kindler–Safra: Almost degree d $\longrightarrow$ almost $O(2^d)$-junta.
- Friedgut: Total influence d $\longrightarrow$ almost $2^{O(d)}$-junta.

Other highlights

- Invariance principle: extending functions from Boolean cube to Gaussian space.
- Small-set expansion: $\Pr_{x \in A}[N_\rho(x) \notin A] \approx 1$ for small A.
- Hypercontractivity: $\|T_\rho f\|_q \leq \|f\|_p$ for $q > p$.
- KKL: Every balanced function has an influential coordinate.

Other domains

- p-biased cube: important in random graph theory (via sharp threshold theorems).
- Johnson scheme (slice): all k-subsets of $[n]$.
    - Setting of Erdős–Ko–Rado theorem.
    - Used by O'Donnell–Wimmer in statistical learning theory.
        - KKL on the slice implies robust Kruskal–Katona.
- Grassmann scheme: all k-dimensional subspaces of $\mathbb{F}_q^n$. (In the application, $q = 2$.)
    - Used recently to prove 2-to-1 conjecture.
- Other groups ($\mathbb{Z}_k^n$, S_n), other association schemes, Gaussian space, Cayley graphs of codes, high-dimensional expanders, ...

2-to-1 conjecture

Label cover

Given: edge-weighted bipartite graph (A, B, E) and constraints $\pi_e \subseteq \Sigma_A \times \Sigma_B$.
Goal: find assignment to vertices which satisfies maximum weight of constraints.

- *a*-to-*b* constraints: $\pi_e = \bigcup_i A_i \times B_i$, where $|A_i| = a$, $|B_i| = b$ are partitions of Σ_A, Σ_B.
- *a*-to-*b* conjectures: for every $\epsilon > 0$, if $|\Sigma_A|, |\Sigma_B|$ are large enough, NP-hard to distinguish between $val \geq 1 - \epsilon$ and $val \leq \epsilon$ when all constraints are *a*-to-*b*.
  - Unique games conjecture: $a = b = 1$.
  - Variant: perfect completeness (not for UGC!).
  - Stronger version when $a = b$: $\Sigma_A = \Sigma_B = \mathbb{Z}_2^n$ and all constraints are linear, i.e. $\ell_1(x) + \ell_2(y) \in S$ for $|S| = a$.

2-to-1 theorem

Recently proved by Dinur, Khot, Kindler, Minzer, Safra.
Corollaries:

- $\sqrt{2}$-hardness for vertex cover (improving over Dinur–Safra's 1.36-hardness).
- Max-cut-gain: distinguishing $1/2+\epsilon$ and $1/2+\epsilon/\log(1/\epsilon)$.
- Distinguishing almost 4-colorable to not almost $1/\epsilon$-colorable.
- Hard to color more than $1-1/k+O(\frac{\ln k}{k^2})$ vertices of almost k-colorable graphs.
- Lasserre integrality gaps.

Grassmann encoding

Traditional PCPs use the Long Code:

- $x \in \{0,1\}^k$ encoded by a table $T_x \colon \{0,1\}^{\{0,1\}^k} \to \{0,1\}$.
- Encoding is $T_x[f] = f(x)$.

Proof of 2-to-1 conjecture using Grassmann Code:

- $x \in \{0,1\}^k$ identified with linear function Λ_x on $\mathbb{Z}_2^k$.
- Λ_x encoded by a table F_x with input an ℓ-dim subspace L and output a linear function on L.
- Encoding is $F_x[L] = \Lambda_x|_L$.

Similar in spirit to the Short Code of Barak, Gopalan, Håstad, Meka, Raghavendra, and Steurer.

Grassmann agreement test

How do we test that F is a valid encoding?

Grassmann agreement test

- Input: For every ℓ-dim subspace L, a linear function $F[L]$ on L.
- Choose L_1, L_2 of dim ℓ with $\dim(L_1 \cap L_2) = \ell - 1$.
- Verify that $F[L_1]|_{L_1 \cap L_2} = F[L_2]|_{L_1 \cap L_2}$.

(Cf. Long Code test, which uses 3 queries.)
Some properties:

- Completeness: test always passes if $F[L] = \Lambda|_L$.
- Test is 2-to-2: $F[L_i]|_{L_1 \cap L_2}$ can be extended to $F[L_i]$ in 2 ways.
- Can be converted to 2-to-1 using two tables.

Soundness of Grassmann agreement test

Grassmann agreement test

- Input: For every ℓ-dim subspace L, a linear function $F[L]$ on L.
- Choose L_1, L_2 of dim ℓ with $\dim(L_1 \cap L_2) = \ell - 1$.
- Verify that $F[L_1]|_{L_1 \cap L_2} = F[L_2]|_{L_1 \cap L_2}$.

Soundness:

- If test passes w.p. $1 - \delta$ then $F[L] = \Lambda|_L$ w.p. $1 - \epsilon(\delta)$.
- What happens if test passes with constant probability δ?
  - Guess: can "list decode" into $C(\delta)$ many Λ's.
  - Counterexample 1: $F[L] = \Lambda_{\min L}|_L$.
  - Counterexample 2: $F[L] = \Lambda_{\min L^\perp}|_L$.

Reduction to small-set expansion

An idea of Barak, Kothari, and Steurer.

Grassmann agreement test

- Input: For every ℓ-dim subspace L, a linear function $F[L]$ on L.
- Choose L_1, L_2 of dim ℓ with $\dim(L_1 \cap L_2) = \ell - 1$.
- Verify that $F[L_1]|_{L_1 \cap L_2} = F[L_2]|_{L_1 \cap L_2}$.

Suppose F passes the test w.p. δ. For random Λ, let

$$S = \{L : F[L] = \Lambda|_L\}.$$

- If $L_1 \in S$ and $\dim(L_1 \cap L_2) = \ell - 1$ then $\Pr[L_2 \in S] = \delta/2$.
- Hence S has expected expansion $1 - \delta/2$.
  $\implies$ can find non-empty S with "small" expansion.

On the Boolean cube, small sets have expansion ≈ 1.
What about the Grassmann scheme?

Small-set expansion on Grassmann scheme

Do all small sets have expansion ≈ 1?

- Counterexample 1: $\{L : x \in L\}$ has expansion $1/2$.
- Counterexample 2: $\{L : y \perp L\}$ has expansion $1/2$.

Grassmann expansion hypothesis

If S has expansion $1-\delta$ then S has density $\epsilon(\delta)$ inside

$$\{L : x_1, \ldots, x_{C(\delta)} \in L, y_1, \ldots, y_{C(\delta)} \perp L\}.$$

Implies soundness of Grassmann agreement test:

- If F passes test w.p. δ then F agrees with Λ on $\epsilon(\delta)$ points of $\{L : x_1, \ldots, x_{C(\delta)} \in L, y_1, \ldots, y_{C(\delta)} \perp L\}$.
- Can cover more of the domain by repeated randomization.

Proof of Grassmann expansion hypothesis

Recently proved by Khot, Minzer, and Safra.

Grassmann expansion hypothesis

If S has expansion $1-\delta$ then S has density $\epsilon(\delta)$ inside

$$\{L : x_1, \ldots, x_{C(\delta)} \in L, y_1, \ldots, y_{C(\delta)} \perp L\}.$$

Proof idea:

- By assumption, $\frac{\langle 1_S, Lap 1_S \rangle}{\langle 1_S, 1_S \rangle} = \delta$.
- Can only happen if $\|1_S^{=d}\| / \|1_S\| \geq \gamma$ for some small d.
- This implies some lower bound on $\mathbb{E}[(1_S^{=d})^4]$.
- Hypothesis follows by expanding $\mathbb{E}[(1_S^{=d})^4]$. (Hard!)

Actual proof uses the bilinear scheme graph with self-loops.

Robust version of Kruskal–Katona

Kruskal–Katona theorem

If $0 \ll k/n \ll 1$ and $A \subseteq \binom{[n]}{k}$ satisfies $0 \ll |A|/\binom{n}{k} \ll 1$ then

$$\frac{|\partial A|}{\binom{n}{k+1}} \geq \frac{|A|}{\binom{n}{k}} + \Omega\left(\frac{1}{n}\right).$$

Extremal example: x_i.

Robust Kruskal–Katona (O'Donnell–Wimmer)

Either A has correlation $\Omega(1/n^\epsilon)$ with some x_i, or

$$\frac{|\partial A|}{\binom{n}{k+1}} \geq \frac{|A|}{\binom{n}{k}} + \Omega\left(\frac{\log n}{n}\right).$$

Follows from KKL on the slice.

Monotone nets (O'Donnell–Wimmer)

Implication: Every monotone function on $\{0,1\}^n$ has correlation $1/2 + \Omega\left(\frac{\log n}{n}\right)$ with one of:

$$0, 1, x_1, \ldots, x_n, \mathrm{Maj}.$$

In fact, for every monotone function f on $\{0,1\}^n$, either

- f has $1-\epsilon$ correlation with 0 or 1; or
- f has $1/2 + 1/n^\epsilon$ correlation with one of $x_1, \ldots, x_n$; or
- f has $1/2 + \Omega\left(\frac{\log n}{n}\right)$ correlation with majority.

Correlation $1/2 + \Omega\left(\frac{\log n}{n}\right)$ is optimal for polynomial size nets (Blum–Burch–Langford).
Can improve size of net to $O(n/\log n)$ using local majorities.