ON THE STRUCTURE OF BOOLEAN FUNCTIONS WITH SMALL SPECTRAL NORM

Ben Lee Volk (Technion)

Joint work with
Amir Shpilka (Technion)
Avishay Tal (Weizmann)

FOURIER TRANSFORM: BASICS

- Represent Boolean functions as functions: $f: \mathbb{Z}_2^n \to \{-1, +1\} \subseteq \mathbb{R}$
- $\chi_\alpha(x) = (-1)^{\langle \alpha, x \rangle}$ (for all $\alpha \in \mathbb{Z}_2^n$) is an orthonormal basis for functions $f: \mathbb{Z}_2^n \to \mathbb{R}$ with respect to the inner product
$$\langle f, g \rangle = \mathbb{E}_x[f(x)g(x)]$$
- Write $f(x) = \sum_\alpha \hat{f}(\alpha)\chi_\alpha(x)$.

FOURIER TRANSFORM AND COMPUTATION

- Restrictions on the computational model imply special Fourier structure (e.g. **[Linial-Mansour-Nisan93]**: small depth circuit ⇒ low degree concentration)
- This talk: restrictions on the Fourier spectrum imply low complexity in certain computational models

Complexity Measures (I)

For $f: \mathbb{Z}_2^n \to \{-1,1\}$:

- The *sparsity* of f: the number of non-zero Fourier coefficients:
$$\|\hat{f}\|_0 = \#\{\, \alpha \in \mathbb{Z}_2^n \mid \hat{f}(\alpha) \neq 0 \,\}$$
- The *spectral norm* (also L_1 *norm*) of f:
$$\|\hat{f}\|_1 = \sum_{\alpha \in \mathbb{Z}_2^n} |\hat{f}(\alpha)|$$

Complexity Measures (II)

"Typical" (random) Boolean functions have "large" (exponential in n) sparsity and spectral norm.

Questions:

1. When is $\|\hat{f}\|_0$ small?
2. When is $\|\hat{f}\|_1$ small?

Boolean Functions with Small Spectral Norm

Fact: Let $f\colon \mathbb{Z}_2^n \to \{-1,1\}$ be a characteristic function of a coset (affine subspace) $U \subseteq \mathbb{Z}_2^n$ (e.g. AND, OR, XOR).

Then $\|\hat{f}\|_1 \leq 3$.

Boolean functions with small spectral norm

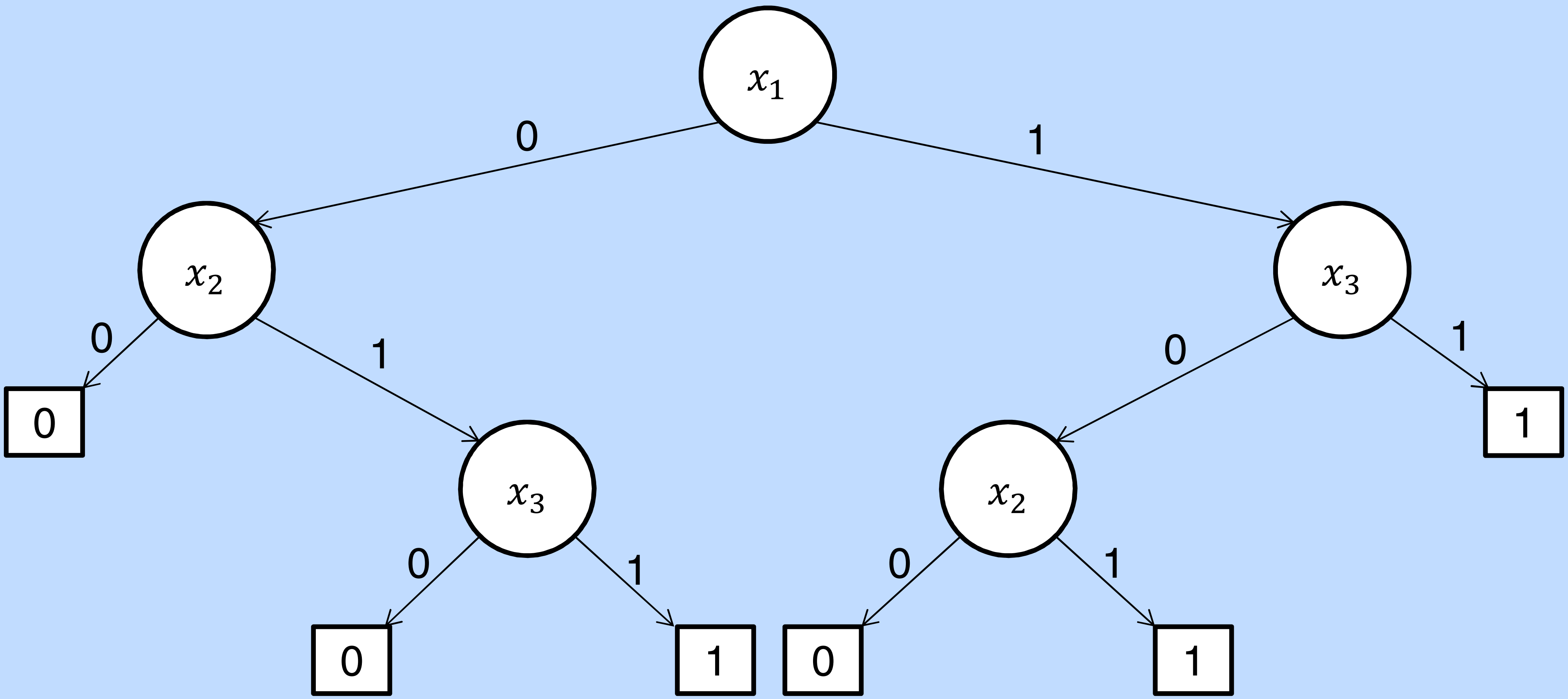
Theorem [Green-Sanders08]: Let $f: \mathbb{Z}_2^n \to \{-1,1\}$ with $\|\hat{f}\|_1 = A$.

Then

$$f = \sum_{i=1}^{L} \pm \mathbf{1}_{U_i}$$

where every $\mathbf{1}_{U_i}$ is a characteristic function of a coset and $L \leq 2^{2^{O(A^4)}}$.

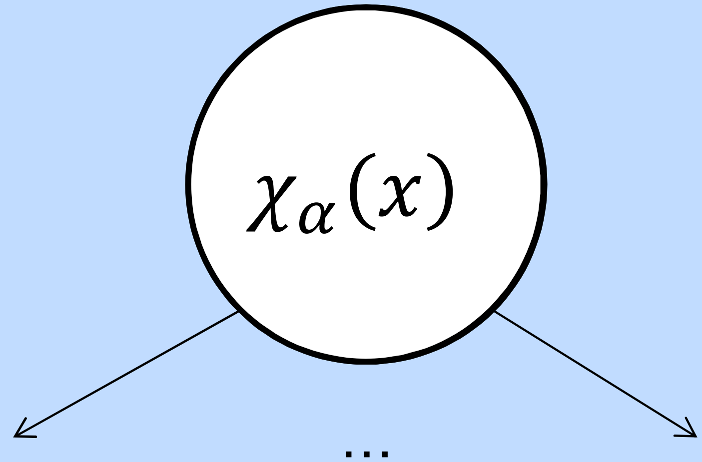
COMPUTATIONAL MODELS: DECISION TREE



Depth of the tree: longest path from the root to a leaf.
$D(f) :=$ minimal-depth decision tree for f.

Computational Models: Parity Decision Tree ($\oplus$-DT)

Same as decision tree, except that every internal node is labeled with a linear function over $\mathbb{Z}_2^n$:



$D^{\oplus}(f) :=$ minimal *depth* of a $\oplus$-DT for f (obviously $D^{\oplus}(f) \leq D(f)$).

$\mathrm{size}_{\oplus}(f) :=$ minimal *size* of a $\oplus$-DT for f (minimal number of leaves).

PARITY DECISION TREES AND NORMS

Fact: Let $f\colon \mathbb{Z}_2^n \to \{-1,+1\}$ be computed by a parity decision tree of depth d and size s. Then:

$$\|\hat{f}\|_1 \leq s \leq 2^d$$
$$\|\hat{f}\|_0 \leq s \cdot 2^d \leq 4^d$$

Question: Is there an inverse theorem? Can any function with small spectral norm or small sparsity be represented by a small parity decision tree?

OUR RESULTS

Theorem(s): Suppose $\|\hat{f}\|_1 = A$. Then:

1. $\exists$ affine subspace $V \subseteq \mathbb{Z}_2^n$, $\mathrm{codim}(V) \le A^2$, such that $f|_V$ is constant.
2. $\mathrm{size}_{\oplus}(f) \le 2n^{A^2}$
3. $D^{\oplus}(f) \le A^2 \log\|\hat{f}\|_0$.
4. $\exists$ $\oplus$-DT of depth $O(A^2 + \log(1/\epsilon))$ that ϵ-approximates f and can be learned efficiently.
5. All of the above: also for $\mathbb{Z}_p^n$.

All Proofs: Corollaries of the following lemma:

MAIN LEMMA

Suppose $\|\hat{f}\|_1 = A > 1$, $\hat{f}(\alpha)$ largest coefficient, $\hat{f}(\beta)$ second largest.

Denote $f|_{\chi_{\alpha+\beta}=z}$ the restriction of f to the subspace $\{\, x \mid \chi_{\alpha+\beta}(x) = z \,\}$.

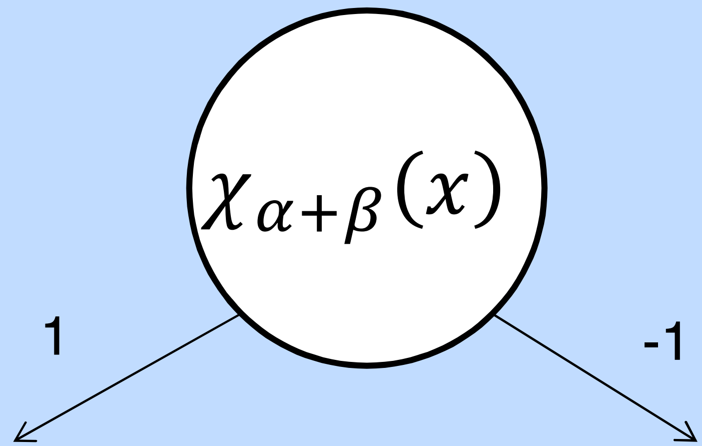
Then $\exists\, b \in \{1, -1\}$ s.t.

$$\left\| \widehat{f|_{\chi_{\alpha+\beta}=b}} \right\|_1 \le A - |\hat{f}(\alpha)| \le A - 1/A$$

$$\left\| \widehat{f|_{\chi_{\alpha+\beta}=-b}} \right\|_1 \le A - |\hat{f}(\beta)|$$

MAIN LEMMA

$$\|\hat{f}\|_1 = A$$



$$\left\|\widehat{f|_{\chi_{\alpha+\beta}=1}}\right\|_1 \leq A - 1/A$$

$$\left\|\widehat{f|_{\chi_{\alpha+\beta}=-1}}\right\|_1 \leq A - |\hat{f}(\beta)|$$

PROOFS
(ASSUMING MAIN LEMMA)

Theorem: Suppose $\|\hat{f}\|_1 = A$. Then there exists an affine subspace $V \subseteq \mathbb{Z}_2^n$, $\mathrm{codim}(V) \leq A^2$, such that $f|_V$ is constant.

Proof:

- Apply Main Lemma on f to obtain a restriction f' such that $\|\widehat{f'}\|_1 \leq A - 1/A$
- Iterate at most A^2 times to obtain a restriction g such that $\|\hat{g}\|_1 = 1$, at which point $g = \chi_\alpha$
- Restrict further on $\chi_\alpha = 1$ if $\alpha \neq 0$

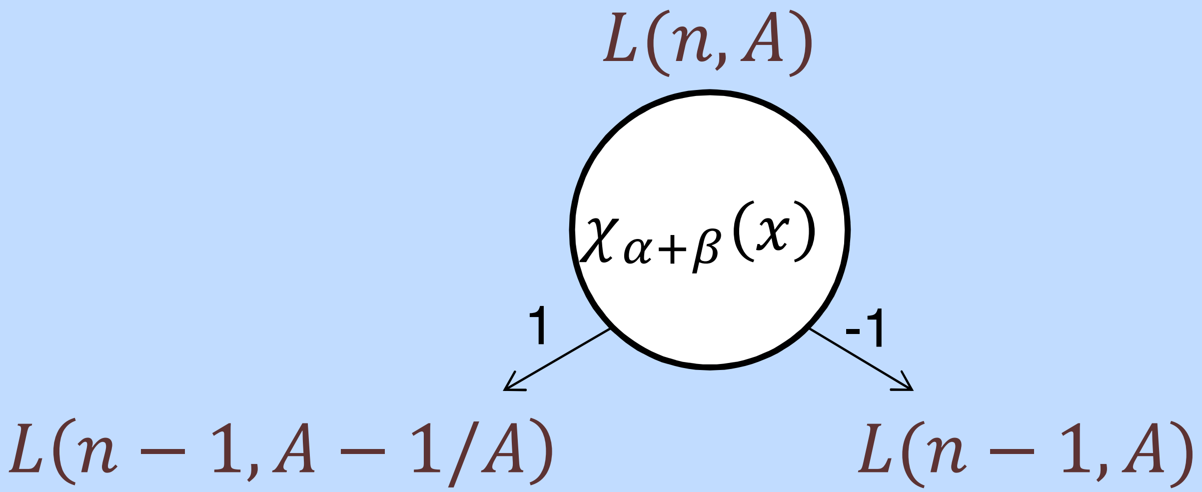
Remark: A slightly more careful analysis of the above algorithm **[Tsang-Wong-Xie-Zhang13]** implies $\mathrm{codim}(V) = O(A)$.

Theorem: If $\|\hat{f}\|_1 = A$ then $\text{size}_{\oplus}(f) \leq 2n^{A^2}$.

Proof: By Induction on n. Let

$$L(n, A) = \max_{f : \|\hat{f}\|_1 \leq A} \text{size}_{\oplus}(f).$$

By Main Lemma:



$$\Rightarrow L(n, A) \leq L(n-1, A-1/A) + L(n-1, A).$$

Remark: Better recursion gives $2^{A^2} n^A$.

Corollary 1: If $\|\hat{f}\|_1 = A$ then

$$f = \sum_{i=1}^{L} \pm \mathbf{1}_{U_i}$$

where every U_i is an affine subspace of $\mathbb{Z}_2^n$ and $L \leq 2^{A^2} n^A$.

[GS08] gives $L = 2^{2^{O(A^4)}}$.

Here L depends on n.

However, we get a $\bigoplus$-DT structure, and a better upper bound if $A = \Omega\left((\log\log n)^{1/4}\right)$.

Corollary 2: f has formula of size $O\left(2^{A^2} n^A \cdot n^2\right)$ and depth $O(A \log n + A^2)$

APPLICATION:
COMPUTATIONAL LEARNING THEORY

LEARNING MODEL

Probably Approximately Correct (PAC) with ***membership queries*** (oracle for f) under the ***uniform distribution***:

Goal: Output, with probability $1-\delta$, an hypothesis g such that $\Pr_x[f(x) \neq g(x)] \leq \epsilon$ (probability taken over the uniform distribution).

Running time = $\mathrm{poly}(n, 1/\epsilon, \log 1/\delta)$

Theorem [Kushilevitz-Mansour93]: If $\|\hat{f}\|_1 = A$,

$$g(x) = \mathrm{sign}\left(\sum_{i=1}^{A^2/\epsilon} \hat{f}(\alpha_i)\chi_{\alpha_i}(x)\right)$$

(α_i's are the largest Fourier coefficients of f) ϵ-approximates f.

Furthermore: there is an algorithm which learns the list of these coefficients (w.p. $1-\delta$) in time $\mathrm{poly}(n, 1/\epsilon, \log 1/\delta)$.

g can be computed by a $\oplus$-DT of depth A^2/ϵ by querying $\chi_{\alpha_i}(x)$ for all i.

OUR RESULT

Theorem: Suppose $\|\hat{f}\|_1 = A$. Then there is an algorithm which outputs (w.p. $1-\delta$) a $\oplus$-DT of depth $O(A^2 + \log(1/\epsilon))$ which computes a function g that ϵ-approximates f.

Running time: $\text{poly}(n, \exp(A^2), 1/\epsilon, \log(1/\delta))$.

	Our Result	**KM**
Hypothesis size	$\text{poly}(\exp(A^2), 1/\epsilon)$	A^2/ϵ
Time complexity of hypothesis on worst input	$O(A^2 + \log(1/\epsilon))$	A^2/ϵ
Structure	$\oplus$-DT (*proper* if input is given as a $\oplus$-DT)	Polynomial Threshold Function (PTF)

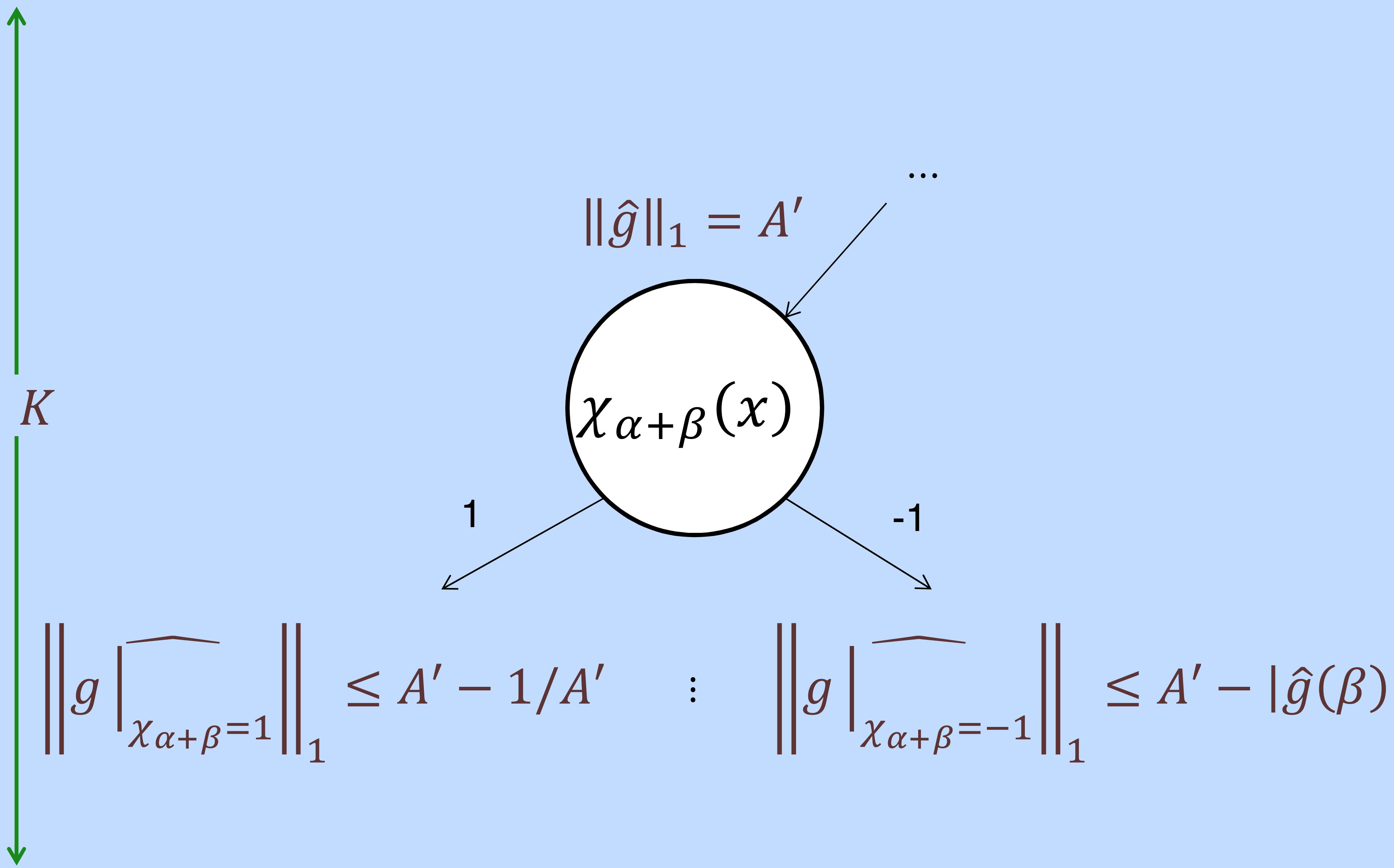
Proof Overview:

Definition: The *bias* of a Boolean function f:

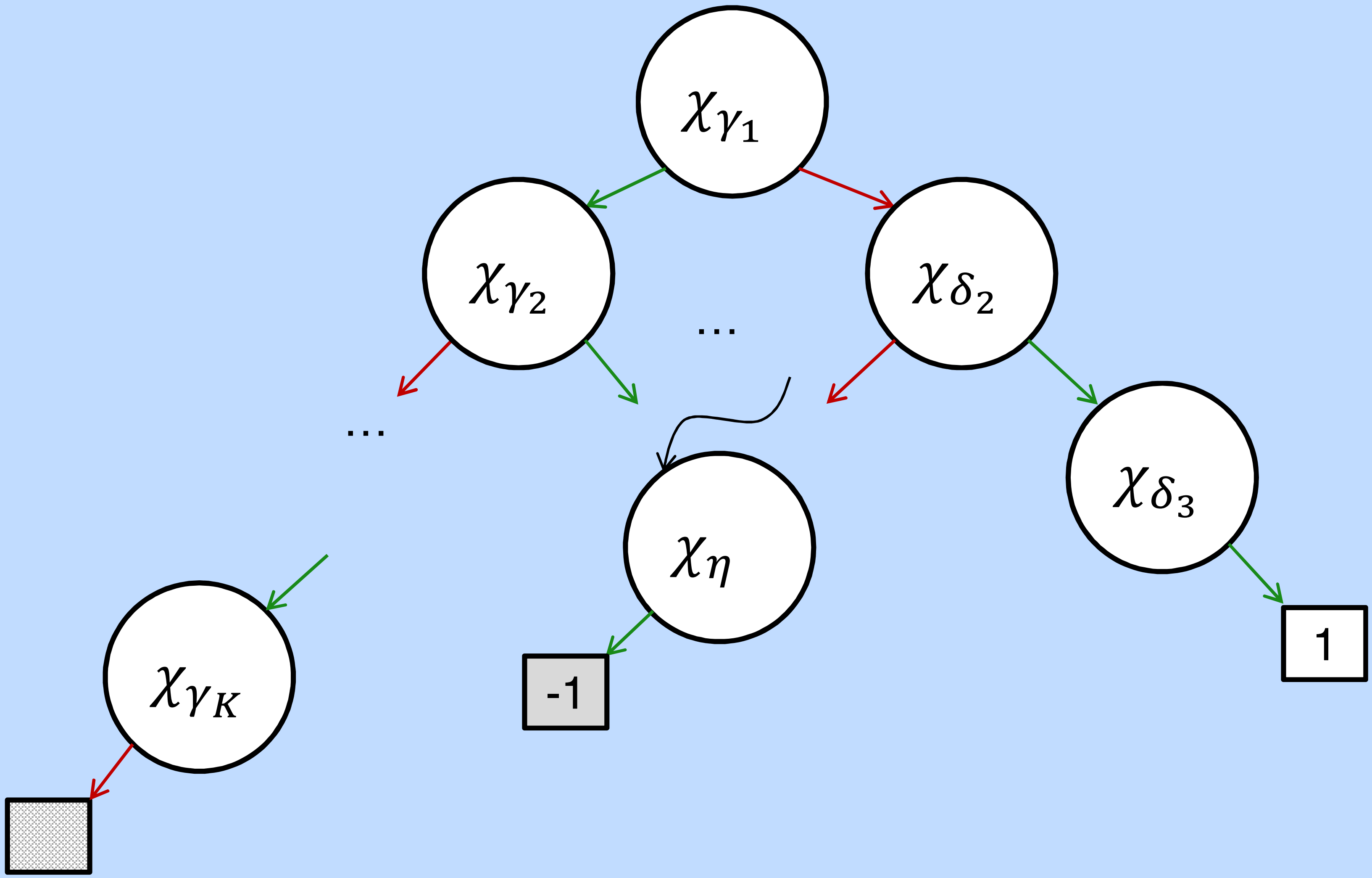
$$\left|\Pr_x[f(x) = 1] - \Pr_x[f(x) = -1]\right|$$

(also equals $\left|\hat{f}(0)\right|$)

Let $K = \max\{10A^2, 2\log(1/\epsilon)\}$. The construction is recursive and stops after at most K levels or when the restricted subfunction has bias $\geq 1 - \epsilon$.



...
$\|\hat{g}\|_1 = A'$
$\chi_{\alpha+\beta}(x)$
K
1
-1
$\left\|\widehat{g|_{\chi_{\alpha+\beta}=1}}\right\|_1 \leq A' - 1/A'$
$\vdots$
$\left\|\widehat{g|_{\chi_{\alpha+\beta}=-1}}\right\|_1 \leq A' - |\hat{g}(\beta)|$



Theorem follows by replacing every subfunction in the leaves with the constant it is biased towards, and then learning largest coefficients by KM algorithm.

Need to show: fraction of inputs that arrive at an unbiased leaf is at most ϵ.

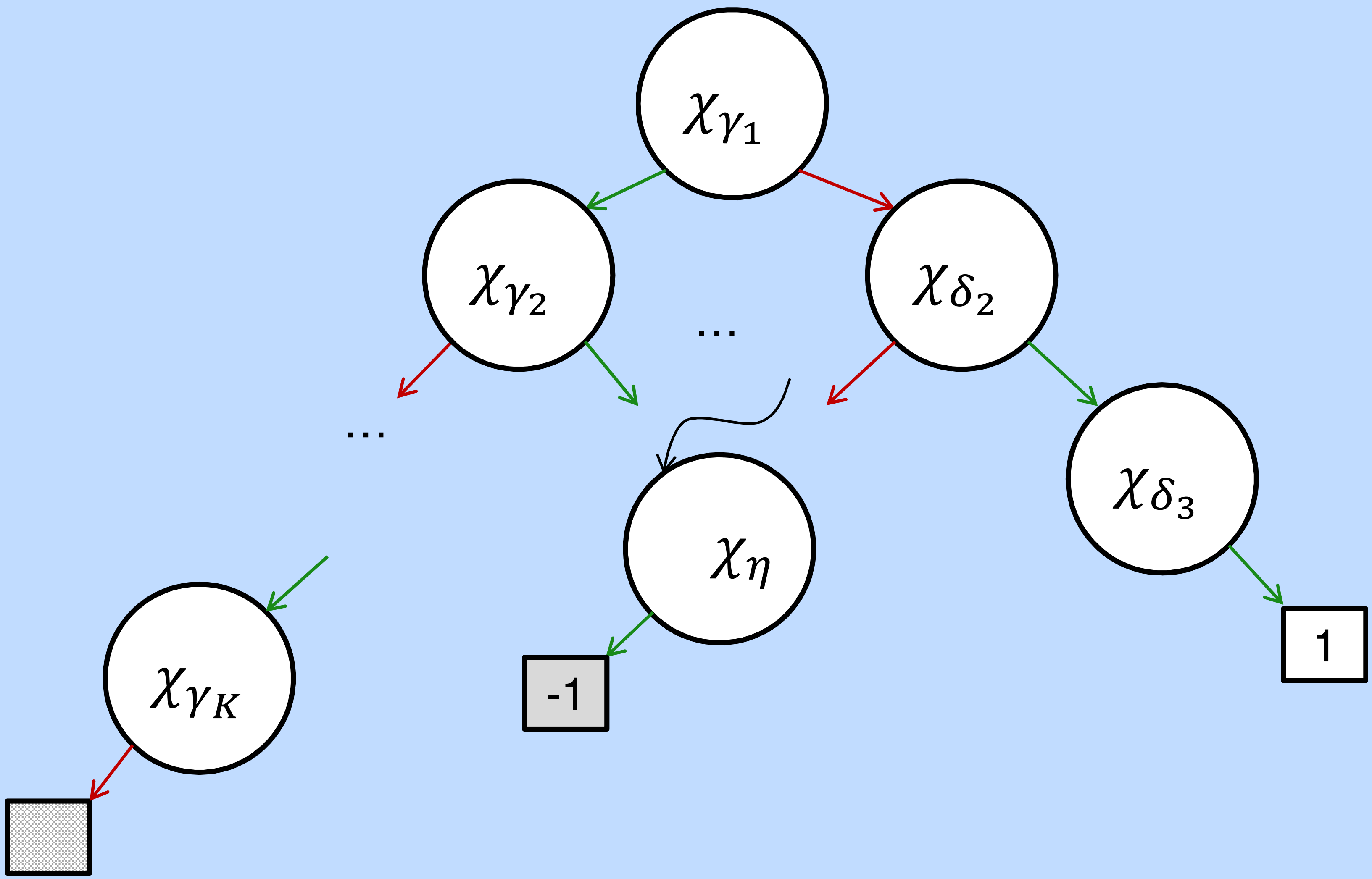
By construction (and Main Lemma), half of the inputs that arrive at every internal node follow a path which reduces the spectral norm by $1/A$.

After A^2 such norm reductions the function is constant (in particular highly biased).

Question: How many inputs make that many "bad moves"?

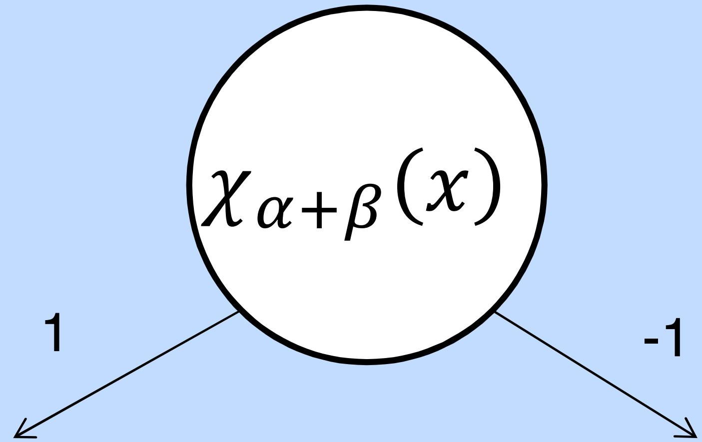
In expectation: A^2 good moves in $2A^2$ moves.

$\Rightarrow$ Probability of following a path of length $K \geq 10A^2$ w/o arriving at a constant leaf is at most $2^{-K/2} \leq \epsilon$.



MAIN LEMMA: PROOF IDEA

$$\|\hat{f}\|_1 = A$$



$$\left\| \widehat{f|_{\chi_{\alpha+\beta}=1}} \right\|_1 \leq A - |\hat{f}(\alpha)| \qquad \left\| \widehat{f|_{\chi_{\alpha+\beta}=-1}} \right\|_1 \leq A - |\hat{f}(\beta)|$$

Suppose for simplicity: $\alpha = 0$, $\hat{f}(0)\hat{f}(\beta) > 0$.

MAIN LEMMA: PROOF IDEA

Restriction on $\chi_\beta = \pm 1$:

For all γ, $\hat{f}(\gamma)$ and $\hat{f}(\beta+\gamma)$ collapse to the same coefficient with absolute value $\left|\hat{f}(\gamma) \pm \hat{f}(\beta+\gamma)\right|$.

BEFORE	$\hat{f}(\gamma_1)$	$\hat{f}(\gamma_1+\beta)$	$\hat{f}(\gamma_2)$	$\hat{f}(\gamma_2+\beta)$	$\cdots$
AFTER	$\hat{f}(\gamma_1) \pm \hat{f}(\gamma_2+\beta)$		$\hat{f}(\gamma_2) \pm \hat{f}(\gamma_2+\beta)$		$\cdots$

MAIN LEMMA: PROOF IDEA

Consider $f' = f|_{\chi_\beta = -1}$:

BEFORE	$\hat{f}(0)$	$\hat{f}(\beta)$	$\cdots$	Contributes $\lvert\hat{f}(0)\rvert + \lvert\hat{f}(\beta)\rvert$
AFTER	$\hat{f}(0) - \hat{f}(\beta)$		$\cdots$	Contributes $\lvert\hat{f}(0) - \hat{f}(\beta)\rvert$

Assumption: $\hat{f}(0)\hat{f}(\beta) > 0$

$$\Rightarrow \|\widehat{f'}\|_1 \leq \|\hat{f}\|_1 - 2|\hat{f}(\beta)|$$

The real action is going on in $f|_{\chi_\beta = 1}$.

FUNCTIONS $f: \mathbb{Z}_p^n \to \{-1,1\}$

All the results also hold for functions $f: \mathbb{Z}_p^n \to \{-1,1\}$ for any prime p.

But more technical work required (characters are no longer Boolean, Fourier coefficients are complex numbers).

Open Problems

- **Conjecture [Zhang-Shi10, Montanaro-Osborne09]:** For all Boolean f, $D^{\oplus}(f) \leq \left(\log\|\hat{f}\|_0\right)^c$
  - Our result: true when $\|\hat{f}\|_1 = \text{poly}\log\|\hat{f}\|_0$
  - Implies the log-rank conjecture for $F(x, y) = f(x \oplus y)$.
  - Known: $c \geq \log_3 6 = 1.63$ ... **[Nisan-Wigderson95, Kushilevitz94]**
  - Equivalent to: ∀ f ∃ subspace V, $\text{codim}(V) \leq \left(\log\|\hat{f}\|_0\right)^{c'}$, $f|_V$ const.
- **Conjecture [TWXZ13]:** ∀ f ∃ subspace V, $\text{codim}(V) \leq \left(\log\|\hat{f}\|_1\right)^c$, $f|_V$ constant.
- $\text{size}_{\oplus}(f) = \text{poly}(n, A)$?

THANK YOU

Braydon Fuller (CC-BY-SA)