k-Independent Gaussians Fool Polynomial Threshold Functions

Daniel M. Kane
Outline

• Problem Background
• Fooling Linear Threshold Functions
• Strategy
  – Structure Theorem
  – Smoothing
  – Polynomial Approximation
  – Anticoncentration
• Wrapup
Polynomial Threshold Functions

• F is a (degree-d) polynomial threshold function (PTF) if it is of the form $F(x) = \text{sgn}(p(x))$ for $p$ a (degree-d) polynomial.
  – d=1, Linear Threshold Functions
Fooling Functions

• We say that a probability distribution, $D$, $\epsilon$-fools a class $C$ of functions with respect to some other distribution $B$, if for every $F \in C$ we have that
  \[|E_{X \sim D}[F(X)] - E_{Y \sim B}[F(Y)]| = O(\epsilon).\]

• We wish to show that for any $d$, $\epsilon > 0$, that for sufficiently large $k$ that any $k$-independent family of Gaussians $\epsilon$-fools the family of degree-$d$ PTFs with respect to the Gaussian distribution.

• Notation: $A \approx_{\epsilon} B$ if $|A - B| = O(\epsilon)$
Previous Work

• [Diakonikolas, Gopalan, Jaiswal, Servedio, Viola 2010]
  – $k = O^\sim(\epsilon^{-2})$ for $d=1$ Bernoullis

• [Diakonikolas, K., Nelson, 2010]
  – $k = O(\epsilon^{-2})$ for $d=1$ Gaussians
  – $k = O(\epsilon^{-8})$ for $d=2$ Gaussians
  – $k = O^\sim(\epsilon^{-9})$ for $d=2$ Bernoullis

• [Meka, Zuckerman 2010]
  – $2^{O(d)} \log(n) \epsilon^{-8d-3}$ seed length PRG for Bernoullis

• [K. 2011]
  – $2^{O(d)} \log(n) \epsilon^{-5}$ seed length PRG for Gaussians
Our Result

**Theorem:** Any $k$-independent family of Gaussians $\epsilon$-fools degree-$d$ PTFs for $k = O_d(\epsilon^{-2^O(d)})$
We want to fool \( F = \text{sgn}(p(x)), \ |p|_2 = 1 \)

Fool \( F^\sim = g(p(x)), \ g \) a smooth approximation of \( \text{sgn} \).

1. \( |g(x) - \text{sgn}(x)| \) small if \( |x| \gg \epsilon \)
2. \( |g^{(t)}(x)| = O(1/\epsilon)^t \)

\[ E[F(X)] \approx E[F^\sim(X)] \approx E[F^\sim(Y)] \approx E[F(Y)] \]

First and Last since \( p(X), p(Y) \) anticoncentrated
Taylor Expansion

• Approximate $g$ by its degree-$k$ Taylor polynomial
  \[ g(p(X)) = T(p(X)) + O(\varepsilon^{-k} |p(X)|^k / k!) \]

• $E[T(p(X))] = E[T(p(Y))]$

• Error: $\varepsilon^{-k} E[p(X)^k] / k! = O(\varepsilon^{-1} k^{-1/2})^k$

• Small if $k >> \varepsilon^{-2}$.

• For $d > 1$ this fails since $E[p(X)^k]$ can be as big as $(dk)^{dk/2}$.
Strategy

• Structure theorem for polynomials
  – Write $p(x)$ as $h(p_1(x), p_2(x), ..., p_N(x))$ where $p_i(x)$ has small higher moments

• Smoothing
  – Approximate $f = \text{sgn} \circ h$ by smooth $f^\sim$, define $F^\sim$

• Taylor approximation
  – Show $E[F^\sim(X)] \approx E[F^\sim(Y)]$

• Anticoncentration
  – Show $E[F(X)] \approx E[F^\sim(X)]$, $E[F(Y)] \approx E[F^\sim(Y)]$
Moment Bounds

- **Theorem**: (slight modification of [Latala 2006])
  For $p$ a hom. deg-$d$ polynomial,
  
  $$
  E[|p(Y)|^k] = O_d \left( \sum_{a=1}^{d} M_a(p) k^{a/2} \right)^k 
  $$

- $M_a(p)$ is the biggest correlation of $p$ with a product of polynomials of lower degree.

- Idea: Higher moments are small, unless $p$ has a reasonable component given by a product of smaller degree polynomials.
Structure Theorem

• **Theorem:** For \( p \) a deg-\( d \) poly, \( m_1 < \ldots < m_d \) integers, can write \( p = h(P_{i,j}) \) so that:
  - \( P_{i,j} \) hom polynomial
  - \( E[|P_{i,j}(Y)|^k] = O_d( k^{1/2} )^k \) for \( k \leq m_i \)
  - For each \( i \), \( n_i = O_d(m_1 \cdots m_{i-1}) \) \( P_{i,j} \)'s
  - (Some extra technical stuff about \( h \))
Structure Theorem

Proof (sketch):

• Use hom. parts of $p(x)$ unless moments are too big.

• This only happens when has large component of $q_1(x) \cdots q_a(x)$

• Replace $p(x)$ by $p(x) - c q_1(x) \cdots q_a(x)$ add new term. Decreases $|p|$ by noticeable amount

• Iterate until all moments are small
Mollification

- $F(x) = \text{sgn}(h(P_{i,j}(x))) = f(P_{i,j}(x))$
- Want to approximate $f$ by smooth $f^\sim$
- $f^\sim = f * \rho$
Which \( \rho \) to Use

**Lemma:** For \( n, C > 0 \),

- \( B(\xi) = C^2/4 - |\xi|^2 \) for \( |\xi| \leq C/2 \), 0 else
- \( \rho_C = |B^\wedge|^2/|B|^2 \)

Then

- \( \rho_C \geq 0 \)
- \( \int_{\mathbb{R}^n} \rho_C(x) dx = 1 \).
- For unit vector \( v \), \( \int_{\mathbb{R}^n} |D_v^k \rho_C(x)| dx \cdot C^k \).
- For \( D > 0 \), \( \int_{|x| > D} |\rho_C(x)| dx = O \left( \left( \frac{n}{CD} \right)^2 \right) \).
Smooth Approximation

- $P_i(x) = (P_{i,j}(x))_{1 \leq j \leq n_i}$
- $P(x) = (P_i(x))_{1 \leq i \leq d}$
- $F(x) = \text{sgn}(h(P(x))) = f(P(x))$
- For appropriately chosen $C_i$, define:
  - $f^\sim = f(P) \ast (\rho_{C_1}(P_1) \times \rho_{C_2}(P_2) \times \cdots \times \rho_{C_d}(P_d))$
  - $F^\sim = f^\sim(P)$
Taylor Approximation

- Let $T(P)$ be the Taylor polynomial of $f$

$$
|T(P) - \tilde{f}(P)| \cdot \prod_{i=1}^{d} \left(1 + \frac{C_i^{m_i} |P_i|^{m_i}}{m_i!} \right) - 1
$$

$$
= O_d \left( \sum_{i=1}^{d} \sum_{e=1}^{d} \left( \frac{C_i^{m_i} |P_i|^{m_i}}{m_i!} \right)^e \right)
$$

Expectation $= O_d \left( \sum_{i=1}^{d} \sum_{e=1}^{d} O_d \left( C_i n_i^{1/2} m_i^{-1/2} \right)^{m_i e} \right)$. 
Taylor Error

• Hence if:
  – \( m_i \gg n_i C_i^2 \)
  – \( k \gg d^2 m_i \)

• \( \mathbb{E}[F^\sim(X)] \approx \epsilon \mathbb{E}[T(P(X))] = \mathbb{E}[T(P(Y))] \approx \epsilon \mathbb{E}[F^\sim(Y)] \)
Approximation Error

**Lem:** If \((x_1, \ldots, x_d) \in \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_d}\), \(D > 0\) so that for all \((y_1, \ldots, y_d)\) with \(|x_i - y_i| < Dn_i d^{1/2} / C_i\) for all \(i\), \(h(x)\) and \(h(y)\) have the same sign, then

\[ |f(x) - f^\sim(x)| = O(\min\{1, D^{-2}\}). \]

**pf (sketch):** \(O(1)\), since \(f, f^\sim\) both \(O(1)\). Otherwise, \(f\) is constant near \(x\), follows from concentration of \(\rho\).
Approximation Error

**Lem:** Given constants $B_i$ with

- $B_i >> (\log(n_i/\epsilon))^{1/2}$
- $m_i >> B_i^2$
- $C_i >> n_i^2 B_i^{i-1} \epsilon^{-2d}$

\[
|E[F(Y)] - E[F^\sim(Y)]| = O(\epsilon)
\]

**pf (sketch):** Unless $|P_{i,j}| > B_i$, or $|p(Y)|$ small (unlikely due to anticoncentration results), previous Lemma applies.
For X

- Don’t have anticoncentration for $p(X)$.
- We use a trick to get around this.
Putting it Together:

• Need $k$, $m_i$, $B_i$, $n_i = O_d(m_1 \cdots m_{i-1})$
  
  – $B_i \gg (\log(n_i/\epsilon))^{1/2}$
  
  – $m_i \gg n_i C_i^2$
  
  – $C_i \gg n_i^2 B_i^{i-1} \epsilon^{-2d}$
  
  – $k \gg d m_i$

• Use:
  
  – $B_i \sim \log(1/ \epsilon)$
  
  – $m_i \sim \epsilon^{-3 \cdot 7^i \cdot d}$
  
  – $C_i \sim \epsilon^{-7^i \cdot d}$
Final Result

- \( k = O_d(\epsilon^{-3.7d\cdot d}) \)-independence \( \epsilon \)-fools degree-d PTFs with respect to the Gaussian distribution

- Note: We were sloppy in several places and the 7 is not optimal

- Conj: \( k = O_d(\epsilon^{-\text{poly}(d)}) \) suffices

- Conj: \( k = O(d^2\epsilon^{-2}) \) suffices

- Note: proof generalizes to Bernoulli except for moment bounds
\[ M_a(p) \]

- **Def:** \( p \) hom. deg-d. \( M_a(p) \) is the:
  - max over:
    - \( q_1, \ldots, q_a \) hom polynomials
    - \( \text{deg. } d_1, \ldots, d_a \)
    - \( d_1 + d_2 + \ldots + d_a = d \)
    - \( E[q_i^2(Y)] = 1 \)
    - of: \( E[p(Y)q_1(Y)q_2(Y) \cdots q_a(Y)] \)

- Biggest component of \( p \) given by a product of a different polynomials.
For X

- Don’t have anticoncentration for $p(X)$.
- $F_c(x) := \text{sgn}(p(x) + c)$
- For small $c$:
  $$E[F_{c}(Y)] \approx_{\epsilon} E[F_{c}(Y)] \approx_{\epsilon} E[F_{c}(X)] > E[F(X)] + O(\epsilon)$$
- Similarly, $E[F(X)] > E[F_{-c}(Y)] + O(\epsilon)$
- Anticoncentration for $Y \Rightarrow$
  $$E[F(Y)] \approx_{\epsilon} E[F_{c}(Y)] \approx_{\epsilon} E[F_{-c}(Y)]$$