The FT-Mollification Method

Daniel M. Kane
Outline

• Introduction
• Linear Threshold Functions
• Multiple Linear Threshold Function
• Degree 2 PTFs
• Conclusion
K-Independence

Def: A collection of random variables $x_1, \ldots, x_n$ are $k$-independent if any collection of at most $k$ of the $x_i$ are independent.

- $k$-independent families easy to produce
- Often behave like fully independent families
Fooling Functions

• Want expectations of functions to be correct
• \( Y=(y_1, y_2, \ldots, y_n) \) Collection of independent random variables
• \( X=(x_1, x_2, \ldots, x_n) \) \( k \)-independent, \( x_i \sim y_i \)
• \( f: \mathbb{R}^n \rightarrow \mathbb{R} \)
• Want to show that \( E[f(X)] \approx E[f(Y)] \)
• Note that if \( f \) is a polynomial of total degree at most \( k \), then \( E[f(X)] = E[f(Y)] \)
**Notation**

**Def:** We say that $k$-independence $\epsilon$-fools $f$ if for all $k$-independent $X$, $|E[f(X)] - E[f(Y)]| < \epsilon$.

We will use $A \approx_\epsilon B$ to denote that $|A - B| = O(\epsilon)$.
Linear Programming

• Let \( y_i \) be \( \pm 1 \) with equal probability
• Fix \( f \)
• What is \( \sup_{X \text{k-indep}} E[f(X)] \)?
• \( p_{a_i} = \Pr(x_i = a_i \text{ for all } i) \geq 0 \)
• \( \sum_{a: \text{k entries fixed}} p_a = 2^{-k} \)
• Maximize \( \sum p_a f(a) \)
• Linear Program
Dual Program

• g a polynomial of degree at most k (sum of functions depending on at most k coordinates)
• \( g(a) \geq f(a) \) for all a
• Minimize \( E[g(Y)] \)
• \( E[g(Y)] = E[g(X)] \geq E[f(X)] \)

• Fooling f with k-independence \( \iff \)
  Approximating f by a degree k polynomial
Gaussians

• From here on we will use \( Y = \text{Gaussian} \).
• Most of these results still hold for Bernoulli variables because of:
  – Same moment bounds
  – Low influence & Invariance Principle give anti-concentration
  – If high influence fixing the values of high influence variables reduces to low influence
One Dimensional Version

- For simplicity assume that $F(Z) = f(<W,Z>)$ (i.e. $f$ depends only on some linear function of $Z$)
- Approximate $f$ by polynomial
- Taylor series

$$f(z) = f(0) + zf'(0) + \ldots + \frac{z^{k-1}}{(k-1)!} f^{(k-1)}(0) \pm \frac{|z|^k}{k!} \sup_w |f^{(k)}(w)|$$

- For $k$ even have error $E[<W,X>^k] \ |f^{(k)}| \infty \ / \ k!$
- Hypercontractive inequality: $|W|_2^k k^{-k/2} \ |f^{(k)}| \infty$
Non-Smooth $f$

- Example: $F$ is an LTF, $F(Z) = I_{[a,\infty)}(\langle W, Z \rangle)$
- Idea: Replace $f$ by smooth approximation $f^{\sim}$
- $E[F(Y)] \approx_{\epsilon} E[F^{\sim}(Y)] \approx_{\epsilon} E[F^{\sim}(X)] \approx_{\epsilon} E[F(X)]$
  - Anti-concentration of $\langle W, Y \rangle$
  - Polynomial Approximation, moment bounds
  - Anti-concentration of $\langle W, X \rangle$ (generally derived from anti-concentration of $\langle W, Y \rangle$)
Smoothing

• Mollification: \( f^\sim = f \ast \rho \)

• Smoothness:
  – \( f^{\sim(k)} = f \ast \rho^{(k)} \)
  – Need bounds on derivatives of \( \rho \)
  – Want strong bounds

• Approximation:
  – \( \int \rho(x) \, dx = 1 \)
  – \( |\rho(x)| \) small when \( x \) large
Fourier Transform

- Differentiation $\leftrightarrow$ Multiplication by $x$
- Smoothness $\leftrightarrow$ Decay at $\infty$
- We want $\rho$ to be very smooth (it’s higher derivatives should be very small) and decay at $\infty$
- Make $\rho$ the Fourier transform of a smooth compactly supported function.
Example $\rho$

- $\rho = \text{FT}(\exp(-x^2/(1-x^2)))$ on $[-1,1]$
  
  - $|\rho^{(k)}(x)| = O(1)$
  
  - $|\rho(x)| = O_n((1+|x|)^{-n})$
  
  - $\int \rho(x) \, dx = 1$

- $\rho_c(x) = c\rho(cx)$
  
  - $|\rho_c^{(k)}(x)| = O(c^k)$
  
  - $|\rho_c(x)| = O_n(c(1+|xc|)^{-n})$
  
  - $\int \rho_c(x) \, dx = 1$
\[ f^\sim \]

- \( f_c^\sim = l_{[a, \infty)} \ast \rho_c \)
  - \( f_c^{(k)} = O(c^k) \)
  - \( |f(x) - f_c^\sim(x)| = O_n \left( \min\{1, |c(x-a)|^{-n}\} \right) \)
Polynomial Approximation

\[ \left| E[\tilde{F}(Y)] - E[\tilde{F}(X)] \right| \cdot O(c^k)O\left(\frac{|W|^k}{2}\right)k^{-k/2} \]

\[ = O \left( \frac{c|W|_2}{\sqrt{k}} \right)^k \]

Need \( k > (2c|W|_2)^2 \), \( \log(1/\epsilon) \).
Anti-Concentration

- $|E[F(Y)] - E[F\sim(Y)]| \leq E[|F(Y) - F\sim(Y)|]$ 
- $|f - f\sim|$ is small except near a 
- $\langle W,Y \rangle$ anti-concentrated
Approximation Error

• $| f - f^\sim | = O(\min\{1, |c(x-a)|^{-2}\})$

\[
\text{Error} = O \left( \sum 4^{-n} Pr(| < W, Y > - a | < 2^n c^{-1}) \right)
\]

\[
= O \left( \sum \frac{2^{-n}}{c|W|_2} \right)
\]

\[
= O \left( \frac{1}{c|W|_2} \right).
\]
Anti-Concentration for $<W,X>$

- To show $E[F^\sim(X)] \approx_\epsilon E[F(X)]$ we need anti-concentration of $<W,X>$

$$\text{Error} = O \left( \sum 4^{-n} Pr(| < W, X > -a | < 2^n c^{-1}) \right)$$

- Have bounds for $Y$

- Need $k$-indep. forces $g(Z) = 1_{[a-b,a+b]}(<W,Z>)$ to not have too large expectation.

- Same idea only make $g < g^\sim$

- Obtain $g^\sim$ by mollifying $21_{[a-2b,a+2b]}$
Fooling LTFs

• Errors from approximating $f$ by $f^\sim$ of size $O(1/(c|W|))$.
  – Need $c > 1/(\epsilon \cdot |W|)$

• To have small error of $f^\sim$ need
  – $k > 2(c|W|)^2 = O(\epsilon^{-2})$
Positive Convolving Function

• Often convenient to have $\rho \geq 0$
  
  – Then $f^\sim \in [\inf(f), \sup(f)]$

• Use $\rho = |\text{FT}(b)|^2$ for $b$ compactly supported
  
  – Normalize by letting $\int \rho = |b|_2^2 = 1$

$$|\rho^{(k)}|_1 = \left| \sum \binom{k}{m} < \text{FT}(b)^{(m)}, \text{FT}(b)^{(k-m)} > \cdot \sum \binom{k}{m} |x^m b|_2 |x^{k-m} b|_2 \right.$$  

$$\cdot \sum \binom{k}{m} |\text{supp}(b)|^k |b|_2^2 = (2|\text{supp}(b)|)^k$$
Multidimensional Mollification

- m –dimensional space
- Use $b(r) = C(1 - |r|^2)$ for $|r| < 1$.

\[
\int |x|^2 \rho(x) \, dx = \sum_i |x_i \hat{b}|^2
\]

\[
= \sum_i \left| \frac{\partial}{\partial r_i} b \right|^2 = O(m^2)
\]

So

\[
\int_{|x| > R} \rho(x) \, dx = O \left( \frac{m^2}{R^2} \right).
\]

- $k^{\text{th}}$ derivative (in any direction) is $< 2^k$. 
Fooling Intersections of Halfspaces

• $F$ a product of $m$ LTFs
• Depends on $\langle W_i, X \rangle$
• $b_i$ o.n. basis for the $W_i$
• $L(X)_i := \langle b_i, X \rangle$
• $F(X) = f(L(X))$
• Multivariate FT-Mollification
Basic Plan

- \( E[F(Y)] \approx_\epsilon E[F^\sim(Y)] \approx_\epsilon E[F^\sim(X)] \approx_\epsilon E[F(X)] \)

- Get \( f^\sim \) by convolving with \( c^m \rho(cx) \) for \( \rho \) the multivariate mollification function
Anti-Concentration

\[ |f - \tilde{f}| = O \left( \min \left\{ 1, \left( \frac{m}{cd} \right)^2 \right\} \right). \]

Where \( d \) is distance to the nearest of the \( m \) hyperplanes.

\[ E[|f(Y) - \tilde{f}(Y)|] = O \left( \frac{m^2}{c} \right). \]

\[ c = \left( \frac{m^2}{\epsilon} \right). \]
Polynomial Approximation

- Approximate $f \sim$ by degree $k-1$ Taylor poly at 0
- Taylor error along line 0 to $L(X)$
- $\text{Error} \leq |f^{(k)}|_{\infty} E[|L(X)|^k] / k!$
- $|f^{(k)}|_{\infty} \leq |f|_{\infty} |\rho^{(k)}|_1 \leq (2c)^k$
- $E[|L(X)|^k] \leq m^{k/2} k^{k/2}$
- $\text{Error} \leq (4m c^2 / k)^{k/2}$
- $k >> m c^2 >> m^5 \epsilon^{-2}$
Quadratic Moment Bound

- **Thrm**: $Q$ a quadratic form, $X$ Gaussian,
  
  $$E[|Q(X)|^k] = O \left( E[Q(X)] + |Q|_2 \sqrt{k} + |Q|_\infty k \right)^k.$$

  - $|Q|_2$ Frobenius norm
  - $|Q|_\infty$ largest eigenvalue

- $|L(X)|^2$ a quadratic form

- $E[|L(X)|^k] = O(m + m k^{1/2} + k)^{k/2}$

- $k = O( m^4 \epsilon^{-2} )$
Positive Value Approximations

• If \( f^\sim \) is bounded, it’s derivatives cannot decay faster than exponentially (look at the FT)
• If \( f^\sim \) only bounded on positive numbers can do better (e.g. \( \cos(x^{1/2}) \))
• In general:
  - \( g(x) = f(x^2) \)
  - \( f^\sim = (g \ast \rho_c)(x^{1/2}) \)
  - \( f^\sim^{(k)} = O(c \ k^{-1})^k \)
  - \( f^\sim \) approximates \( f \) on \( \mathbb{R}^+ \)
Fooling Degree-2 PTFs

• $F(X) = \mathbb{I}_{[0,\infty)}(p(X))$, $p$ degree 2 poly

• **Cannot** use FT-Mollification Naively, $f\sim(p(X))$

• Error from polynomial approximation:

  $\text{Error} \cdot \left| \tilde{f}^{(k)} \right| E[|p(X)|^k]/k!$

  $\cdot c^k k^k |p|^k_2/k!$

  $= O \left( c|p|_2^k \right)$.

• Cannot make $c$ big enough.
Decomposition

• Suppose $|p|_2 = 1$

• $p(X) = Q(X) + L(X) + C$
  – Quadratic, Linear, Constant

• $Q(X) = Q_+(X) - Q_-(X) + Q_0(X)$
  – Eigenvalues $> \delta$
  – Eigenvalues $< -\delta$
  – $|\text{Eigenvalues}| \leq \delta$

• $Q_+, Q_-$, positive

• $|Q_0|_\infty < \delta$
Decomposition

- $M(X) = (m_1(X), m_2(X), m_3(X), m_4(X))$
  - $m_1(X) = Q_+(X)^{1/2}$
  - $m_2(X) = Q_-(X)^{1/2}$
  - $m_3(X) = Q_0(X) - \text{tr}(Q_0)$
  - $m_4(X) = L(X)$

- $f(w, x, y, z) = I_{[0, \infty)}(w^2 - x^2 + y + z + \text{tr}(Q_0) + C)$

- $F(X) = f(M(X))$

- $f^{\sim}$ is $f$ convolved with $\rho_c$. 
Anti-Concentration

• $|f - f^\sim| \text{ small unless } M(X) \text{ within } \sim c^{-1} \text{ of boundary}$
• $|m_1(X)|, |m_2(X)| = O(\delta^{-1/2}) \text{ w.h.p.}$
• $|f - f^\sim| \text{ small unless } p(X) \text{ within } \sim \delta^{-1/2} c^{-1} \text{ of } 0$
• Gaussian anti-concentration
• Error = $O(\delta^{-1/4} c^{-1/2})$
Polynomial Approximation

- Error = |f^{(k)}| \ E[|M(X)|^k] / k!
  - |f^{(k)}| = O(c)^k
  - E[|m_1(X)|^k], E[|m_2(X)|^k] = E[k/2^{th} moment of quadratic] = O(k^{1/2+\delta^{-1/2}})^k
  - E[|m_3(X)|^k] = E[|Q_0-\text{tr}(Q_0)|^k] = O(k^{1/2} + k\delta)^k
  - E[|m_4(X)|^k] = E[L(X)^k] = O(k^{1/2})^k
- Error = O(c \ k^{-1/2} + c\delta + c\delta^{-1/2} k^{-1})^k
Putting it Together

• Need:
  - $ck^{-1/2} \ll 1$
  - $c\delta \ll 1$
  - $c\delta^{-1/2} k^{-1} \ll 1$
  - $k \gg \log(1/\epsilon)$
  - $\delta^{-1/4} c^{-1/2} \ll \epsilon$

• $c \sim \epsilon^{-4}$

• $\delta \sim \epsilon^4$

• $k \sim \epsilon^{-8}$
Fooling Degree d PTFs

• New Result:
• $k = O(e^{-2^{O(d)}})$-independence $\epsilon$ fools degree $d$ PTFs of Gaussians
• Moment bounds for polynomials
• Structure theorem for polynomials
• FT-Mollification
Further Work

• Better multi-dimensional version
• Find new applications
References


When Moments Don’t Exist

• $y_i \sim \text{Cauchy} \sim \frac{dy}{1+y^2}$

• Fool $f(\sum a_i y_i)$
  – $f$ some function
  – $a_i$ fixed constants ($\sum |a_i| = 1$)

• Replace $f$ by smooth $f^\sim$

• Approximate by polynomial

• Problem: $E[p(\sum a_i x_i)] \neq E[p(\sum a_i y_i)]$ since neither exists
Cutoff Variables

• Idea: cut off $x_i$ at some point so moments exist
• Pick parameter $\lambda$
• $U_i = 1$ if $|a_i x_i| > \lambda$, 0 else
• $x_i' = (1-U_i) x_i$
• $E[U_i] = O(|a_i| \lambda^{-1})$
• $E[|a_i x_i'|^m] = O(|a_i| \lambda^{m-1})$
First Attempt

• Consider $f(\sum a_i x_i')$
• **Problem:** Incorrect when some $x_i$ overflows
• $E[\sum U_i] = O(\sum |a_i| \lambda^{-1}) = O(\lambda^{-1})$
• Happens often unless $\lambda$ is large.
Second Attempt

• Condition based upon which $x_i$ overflow

$$
\sum_{S \subseteq [n], |S| < k} \prod_{i \in S} U_i \cdot f \left( \sum_{i \in S} a_i x_i + \sum_{i \notin S} a_i x'_i \right).
$$

• **Problem**: Overcounts when more indices than those in $S$ overflow.
Solution

- **Idea:** Use approximate Inclusion-Exclusion

\[
\sum_{S \subseteq [n], |S| \leq k} \left( \prod_{i \in S} U_i \right) \left( \sum_{T \subseteq [n] \setminus S, |T| \leq k} (-1)^{|T|} \prod_{i \in T} U_i \right) f \left( \sum_{i \in S} a_i x_i + \sum_{i \notin S} a_i x'_i \right)
\]

- Gives right answer when at most k overflow

- \( U := \sum U_i \)

\[
\text{Error} = O \left( 2^k \sum_{s=0}^{k} \binom{k}{s+k+1} \right)
\]

\[
E[\text{Error}] = O \left( 2^k \sum_{s=0}^{k} \frac{\lambda^{-(s+k+1)}}{(s+k+1)!} \right) = O \left( \frac{\lambda^{-1}}{k} \right)^{\Theta(k)}
\]
Polynomial Approximation

- Fix $S$, $T$
- Condition on $x_i$ for $i \in S \cup T$

\[ \pm \left( \prod_{i \in S \cup T} U_i \right) f \left( \sum_{i \in S} a_i x_i + \sum_{i \in T} a_i x'_i + \sum_{i \notin S \cup T} a_i x'_i \right) \]

\[ = \pm \left( \prod_{i \in S \cup T} U_i \right) f \left( X_0 + \sum_{i \notin S \cup T} a_i x'_i \right) \]

- Taylor approximate $f$ about $X_0$
- Error \( \leq E[\prod U_i] \ |f^{(k)}| \ E[|\sum a_i x'_i|^k] \) / $k!$
Error Bounds

- Expanding and using some combinatorics
  
  \[ E \left[ \left| \sum a_i x'_i \right|^k \right] \cdot \sum_{m=1}^{k/2} \lambda^{k-m} m^k / m! = \sum_{m=1}^{k/2} O(1)^m (\lambda m)^{k-m}. \]

- If \( \lambda = d/k \), then this is \( O(d)^k \).

- If \( |f^{(k)}| = O(c^k) \)

  \[ \text{Error} \cdot E \left[ \prod_{i \in SUT} U_i \right] O \left( \frac{cd}{k} \right)^k. \]
Error Bounds

- Summing over $S, T$ we get at most
  \[ \prod_i (1 + 2E[U_i]) = \exp(O(E[U])) = O(1)^{\lambda^{-1}} = O(1)^k. \]

- Total Error from polynomial approximation
  \[ O\left(\frac{cd}{k}\right)^k. \]

- Total Error from $I-E$
  \[ O(d^{-1})^{\Theta(k)}. \]
Putting it together

• Need:
  – \( k \gg \log(1/\epsilon) \)
  – \( d \gg 1 \)
  – \( k \gg cd \)

• Hence if \( |f^{(k)}| = O(c^k) \) can \( \epsilon \)-fool \( f \) with \( O(c + \log(1/\epsilon)) \)-independence.
Smoothing

• If f not smooth, FT-Mollify
• Example: f threshold function
• $|f^{\sim(k)}| = O(c^k)$
• $|f - f^{\sim}|$ small except within $c^{-1}$ of boundary
• Introduce error of $O(c^{-1})$
• Use $k = O(\epsilon^{-1})$