Diffuse Decompositions of Polynomials

Daniel Kane

Department of Computer Science and Engineering / Department of Mathematics
University of California, San Diego
dakane@ucsd.edu

November 11th, 2014
Overview

- Introduction
- Invariance Principle and the Replacement Method
- Bottlenecks for Existing Techniques
- Diffuse Decompositions
- Diffuse Invariance Principle
- Application to Noise Sensitivity
Definitions

We briefly recall some basic definitions:

**Definition**

We call a function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ a (degree-$d$) **Polynomial Threshold Function** (or PTF) if it is of the form $f(x) = \text{sgn}(p(x))$ for $p$ a (degree-$d$) polynomial in $n$ variables.

We will be interested in questions involving the evaluation of polynomial threshold functions at random Gaussian or Bernoulli inputs. We will make extensive use of the $L^t$-norms with respect to each of these distributions.

**Definition**

For $p : \mathbb{R}^n \rightarrow \mathbb{R}$ and $t \in \mathbb{R}$ define

$$|p|_t := \left( \mathbb{E}_{X \sim G^n} \left[ |p(X)|^t \right] \right)^{1/t}.$$  

$$|p|_{B,t} := \left( \mathbb{E}_{B \sim u\{-1,1\}^n} \left[ |p(B)|^t \right] \right)^{1/t}.$$
Invariance Principles

- Many questions about PTFs are easier in the Gaussian case
- Many questions about PTFs are more interesting in the Bernoulli case
- Would like to reduce the latter to the former
- An *Invariance Principle* is a theorem that shows that for multilinear $p$

\[
\text{Dist}(p(G)) \approx \text{Dist}(p(B))
\]

usually in cdf distance.
- For degree-1 polynomials, reduces to the Berry-Esseen Theorem.
The Replacement Method

The replacement method is a technique for proving:

\[ \mathbb{E}[f(X_1, \ldots, X_n)] \approx \mathbb{E}[f(Y_1, \ldots, Y_n)] \]

where

- \( f \) is some function
- \( X_i, Y_i \) independent random variables
- Low degree moments of \( X_i \) agree with corresponding moments of \( Y_i \)

The replacement method is used to prove:

- Invariance principles
- Error bounds for pseudorandom generators
- Bounds on noise sensitivity
Replacement Method - Setup

- Approximate $f$ by a smooth function $g$
- Show
  \[ \mathbb{E}[f(X)] \approx \mathbb{E}[g(X)] \approx \mathbb{E}[g(Y)] \approx \mathbb{E}[f(Y)] \]

- Often the last approximation will be hard
- Sandwich $f$ between smooth functions $g_+ \geq f \geq g_-$
- Show
  \[ \mathbb{E}[f(X)] \approx \mathbb{E}[g_+(X)] \approx \mathbb{E}[g_-(Y)] \]
  \[ \mathbb{E}[g_+(Y)] \geq \mathbb{E}[f(Y)] \geq \mathbb{E}[g_-(Y)] \]

- We consider simply $f \approx g$
To show $\mathbb{E}[g(X)] \approx \mathbb{E}[g(Y)]$ replace $X_i$ by $Y_i$ one at a time.

Show:

$$\mathbb{E}[g(Y_1, \ldots, Y_{i-1}, Y_i, X_{i+1}, \ldots, X_n)] 
\approx \mathbb{E}[g(Y_1, \ldots, Y_{i-1}, X_i, X_{i+1}, \ldots, X_n)].$$

For fixed $Y_1, \ldots, Y_{i-1}, X_{i+1}, \ldots, X_n$ Taylor expand

$$g(Y_1, \ldots, Y_{i-1}, Z, X_{i+1}, \ldots, X_n) = a_0 + a_1 Z + \ldots + a_k Z^k + O(|Z^{k+1}|)$$

First few terms have same expectation for $Z = X_i$ and $Z = Y_i$

Error from higher degree terms

If $k$ moments agree, error is roughly

$$\left| \frac{\partial^{k+1} g}{\partial Z^{k+1}} \right|$$
Approximation Error

- To show $\mathbb{E}[f(X)] \approx \mathbb{E}[g(X)]$, show with high probability that $f(X) \approx g(X)$.
- This generally amounts to proving an anticoncentration result, telling us that $X$ (or some function of $X$) has small probability of lying in some small region.
Using the Replacement Method

- \( f(x) = \text{sgn}(p(x)) \), \( \text{Var}(p) = 1 \), \( p \) multilinear
- \( X_i \) Gaussian, \( Y_i \) Bernoulli agree in first \( k = 3 \) moments
- Want \( \mathbb{E}[f(X_1, \ldots, X_n)] \approx \mathbb{E}[f(Y_1, \ldots, Y_n)] \)
- \( g(x) = \rho(p(x)) \) for \( \rho(x) = \text{sgn}(x) \) when \( |x| > \eta \)
- Errors
  - Replacement \( \sum_{i=1}^{n} \mathbb{E} \left[ \left| \frac{\partial^4 g}{\partial X^4_i} \right| \right] \)
  - Approximation \( \mathbb{E}[|g(X) - f(X)|] \)
Replacement Error

Error

\[
\sum_{i=1}^{n} \mathbb{E} \left[ \frac{\partial^4 g}{\partial X_i^4} \right] \approx \rho^{(4)} \sum_{i=1}^{n} \mathbb{E} \left[ \left| \frac{\partial p(X)}{\partial X_i} \right|^4 \right] \approx \eta^{-4} \sum_{i=1}^{n} \mathbb{E} \left[ \left| \frac{\partial p(X)}{\partial X_i} \right|^4 \right]
\]

Definition

If \( p \) is multilinear, define \( \text{Inf}_i(p) = \mathbb{E} \left[ \left| \frac{\partial p(X)}{\partial X_i} \right|^2 \right] \).

Note: \( \sum_i \text{Inf}_i(p) \approx \text{Var}(p) \)

Replacement Error \( \approx \eta^{-4} \sum_{i=1}^{n} \text{Inf}_i(p)^2 \approx \eta^{-4} \max_i \text{Inf}_i(p) \)

Definition

We say that a polynomial \( p \) is \( \tau \)-regular if \( \text{Inf}_i(p) \leq \tau \text{Var}(p) \) for all \( i \).
Approximation Error

- Error $= \mathbb{E}[|f(X) - g(X)|] \approx \Pr(|p(X)| < \eta)$
- *Anticoncentration* result needed

**Lemma (Carbery and Wright)**

*If $p$ is a degree-$d$ polynomial in $n$ variables, $X$ an $n$-dimensional Gaussian, and $\epsilon > 0$ then*

$$Pr(|p(X)| \leq \epsilon |p|_2) = O(d\epsilon^{1/d}).$$

- Error $\approx \eta^{1/d}$
The Invariance Principle

Theorem (Mossel, ODonnell, and Oleszkiewicz)

If $p$ is a $\tau$-regular, degree-$d$ multilinear polynomial, then

$$|Pr(p(B) \leq t) - Pr(p(G) \leq t)| = O(d\tau^{1/(8d)}).$$
Dealing with Irregular Polynomials

Not all $p$ are regular

- $p$ irregular because of a few influential coordinates
- Deal with influential coordinates separately
- Fix a few influential coordinates either:
  - The values of these coordinates essentially determines the polynomial
  - The resulting polynomial is regular
- Use this to reduce to the regular case
- Such a theorem is a *Regularity Lemma*
Regularity Lemma

**Theorem**

Let $p$ be any degree-$d$ polynomial, $\tau > 0$. Then $p$ is equivalent to a decision tree of depth

$$\text{depth}(d, \tau) = \frac{1}{\tau} \cdot O_d(\log(\tau^{-1}))^{O(d)}$$

so that with probability $1 - \tau$ a random leaf corresponds to a polynomial $p_{\rho}$ so that either

- $\text{Var}(p_{\rho}) < \tau^2 |p_{\rho}|_2^2$
- $p_{\rho}$ is $\tau$-regular
Combining the Invariance Principle and Regularity Lemma we find any multilinear polynomial evaluated at a random Bernoulli can be written as a decision tree of depth $\tilde{O}_d(\tau^{-1})$ whose leaves with probability $1 - \tau$ are either:

- Nearly constant
- Within $O(d^{1/(8d)})$ in cdf distance of a polynomial of Gaussians

Each of the Invariance Principle and Regularity Lemma is essentially tight, but this combination is not.
A Bad Case for the Invariance Principle

The $\tau^{1/d}$-dependence in the Invariance Principle is necessary

- Let $p(x) = \tau x_0 + \left( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} x_i \right)^d$ for $d$ even
- $p$ is roughly $\tau^2$-regular
- $p(B) \geq -\tau$ with probability 1
- $p(G) < -\tau$ with probability $\approx \tau^{1/d}$
- Although $x_0$ not very influential, can still have large effect when $\sum_{i=1}^{N} x_i$ is small
Problems arise when the distribution of $p(X)$ is highly clumped
The $\epsilon^{1/d}$ dependence of Carbery-Wright is a major bottleneck
Would like $\Pr(|p(X)| < \epsilon|p|_2) \approx \epsilon$
Not true in general:
- $p(x) = x_1^d$
- $p(x) = x_1^d + x_2^d + x_3^d + \delta$
- $p(x) = q_1(x)^7 + q_2(x)^7 + q_1(x)^2 q_2(x)^2 + \delta q_3(x)$
In all of the above $p(x)$ fails to be well anticoncentrated because of how it can be decomposed into other polynomials
Let $g$ depend on $q_i$
Decompositions of Polynomials

Idea

For any polynomial $p$, we should be able to explain its failure to be anticoncentrated by expressing it in terms of other polynomials with no special distributional properties.

In order to make this rigorous, we will need to introduce some terminology.

Definition

Let $p(x)$ be a degree-$d$ polynomial. We say that a decomposition of $p$ of size $m$ is a sequence of polynomials $(h, q_1, \ldots, q_m)$ so that:

- $p(x) = h(q_1(x), \ldots, q_m(x))$
- For each monomial $\prod x_i^{a_i}$ of $h$, we have $\sum a_i \deg(q_i) \leq d$
Diffuse Decompositions

Definition
We say that a sequence \((q_1, \ldots, q_m)\) of polynomials \(\mathbb{R}^n \to \mathbb{R}\) is an \((\epsilon, N)\)-diffuse set if

- \(|q_i|_2 \leq 1\) for all \(i\)
- For any \((a_1, \ldots, a_m) \in \mathbb{R}^m\),

\[
\Pr \left( |q_i(G) - a_i| < \epsilon \text{ for all } 1 \leq i \leq m \right) \leq N\epsilon^m.
\]

Definition
An \((\epsilon, N)\)-diffuse decomposition of size \(m\) of a polynomial \(p\) is a decomposition \((h, q_1, \ldots, q_m)\) of \(p\) of size \(m\) so that \((q_1, \ldots, q_m)\) is an \((\epsilon, N)\)-diffuse set.
The Diffuse Decomposition Theorem

**Theorem (K.)**

Let \( \epsilon, c, N > 0 \) and \( p(X) \) be a degree-\( d \) polynomial. Then there exists a degree-\( d \) polynomial \( p_0 \) with:

- \( |p - p_0|_2 < O_{c,d,N}(\epsilon^N)|p|_2 \)
- \( p_0 \) has an \((\epsilon, \epsilon^{-c})\)-diffuse decomposition of size at most \( O_{c,d,N}(1) \).

**Remark**

The \( O_{d,c,N}(1) \) terms in the Theorem statement hide an \( A(d + O(1), Nc^{-1}) \). I suspect that \( \text{poly}(d, c, N) \) suffices.
Proof.

- Start with the decomposition $p = \text{Id} \circ p$ and iteratively refine.
- Show that if $(q_1, \ldots, q_m)$ is not $(\epsilon, \epsilon^{-c})$-diffuse then some $q_i$ can be decomposed as a sum of:
  - A linear combination of $q_j$ of the same degree
  - A bounded number of products of lower degree polynomials
  - A small error
- Use an ordinal monovariant (based on the number of $q_i$ of each degree) to show that the above process terminates.
Diffuse Replacement Method

- Have \( f(x) = \text{sgn}(p(x)) \) with \( \text{Var}(p) = 1 \)
- Obtain \( |p_0 - p|_2 < O_{d,c}(\eta^{2d}) \) with \((\eta, \eta^{-c})\)-diffuse decomposition \((h, q_1, \ldots, q_m)\)
- Approximate

\[
\begin{align*}
  f(x) &= \text{sgn}(p(x)) \\
  &\approx \text{sgn}(p_0(x)) \\
  &= \text{sgn}(h(q_1(x), \ldots, q_m(x))) \\
  &\approx \rho(q_1(x), \ldots, q_m(x))
\end{align*}
\]

- Use \( \rho(q_1, \ldots, q_m) = \text{sgn}(h(q_1, \ldots, q_m)) \) as long as \((q_1, \ldots, q_m)\) more than \(\eta\) from the boundary
Approximation Error

- Since probably $|p(X) - p_0(X)| \approx \eta^{2d}$, we have $\mathbb{E}[|\text{sgn}(p(X)) - \text{sgn}(p_0(X))|] = O(\eta)$.
- The error $\mathbb{E}[|\text{sgn}(h(q_1(X), \ldots, q_m(X))) - \rho(q_1(X), \ldots, q_m(X))|]$ is roughly the probability that $(q_1(X), \ldots, q_m(X))$ is within $\eta$ of a zero of $h$
- Standard results imply that $|q_i(X)| \leq \text{polylog}(\eta)$ with high probability
- Cover remaining region with $\eta$-sized boxes and use the diffuse property to bound probability
- Approximation Error $\tilde{O}_{d,m}(\eta^{1-c})$
Have

\[ \frac{\partial^k \rho}{\partial q_{\ell_1} \cdots \partial q_{\ell_k}} = O_{m,k}(\eta^{-k}) \]

Replacement error

\[ \sum_{i=1}^{n} \mathbb{E} \left[ \left| \frac{\partial^k g(X)\partial X_i^k}{\partial X_i^k} \right| \right] \approx O_{m,k}(\eta^{-k}) \left( \sum_{i,j} \mathbb{E} \left[ \left| \frac{\partial q_j(X)}{\partial X_i} \right|^k \right] \right) \]
The Diffuse Invariance Principle

Use the diffuse replacement method to prove invariance principle.
Replacement error depends:

$$\sum_{i,j} \mathbb{E} \left[ \left| \frac{\partial q_j(X)}{\partial X_i} \right|^k \right].$$

We will use the following notion of regularity:

**Definition**

For $p$ a degree-$d$ multilinear polynomial, we say that $p$ has a $(\tau, N, m, \epsilon)$-regular decomposition if there exists degree-$d$ $p_0$ so that:

- $|p - p_0|_{B, 2}^2 \leq \epsilon^2 \text{Var}(p_0(X))$
- $p_0$ has a $(\tau, N)$-diffuse decomposition of size $m$, $(h, q_1, \ldots, q_m)$ so that:
  - $q_i$ is multilinear for each $i$
  - $\inf_j(q_i) \leq \tau$ for each $i, j$
The Diffuse Invariance Principle

**Theorem (K.)**

If \( p \) is a degree-\( d \) multilinear polynomial with a \((\tau, N, m, \epsilon)\)-regular decomposition, then \( |Pr(p(B) \leq t) - Pr(p(G) \leq t)| \) is at most

\[
O_{d,m}(\tau^{1/5} N \log(\tau^{-1})^{dm/2+1} + \epsilon^{1/d} \log(\epsilon^{-1})^{1/2}).
\]

Note that \( p(x) = \tau x_0 + \left( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} x_i \right)^d \) has diffuse decomposition \( p(x) = \tau q_1(x) + q_2(x)^d \) with

\[
q_1(x) = x_0, \quad q_2(x) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} x_i.
\]

The irregularity of \( q_1 \) locates our problem.
The Diffuse Regularity Lemma

**Theorem (K.)**

Let $\tau, c, M > 0$, and let $p$ be a degree-$d$ polynomial with Bernoulli inputs. Then $p$ can be written as a decision tree of depth at most

$$O_{c,d,M} \left( \tau^{-1} \log(\tau^{-1})^{O(d)} \right)$$

so that with probability at least $1 - \tau$, a random leaf corresponds to a polynomial $p_\rho$ which satisfies either

- $\text{Var}(p_\rho) < \tau^M |p_\rho|_2^2$
- $p_\rho$ has an $(\tau, \tau^{-c}, O_{c,d,M}(1), O_{c,d,M}(\tau^M))$-regular decomposition

The proof is essentially the same as the proof of the Decomposition Theorem with a little extra care used to ensure that the $q_i$ are regular.
Combining the diffuse versions of the Invariance Principle and Regularity Lemma we find any multilinear polynomial evaluated at a random Bernoulli can be written as a decision tree of depth $\tilde{O}_d(\tau^{-1})$ whose leaves with probability $1 - \tau$ are either:

- Nearly constant
- Within $O_{c,d}(\tau^{1/5-c})$ in cdf distance of a polynomial of Gaussians
Application: Noise Sensitivity

For a Boolean function $f$, the noise sensitivity of $f$ is a measure of the sensitivity of the output of $f$ to small changes in the input. There are a number of different measures of noise sensitivity.
Bernoulli Noise Sensitivity

Definition

For \( f : \mathbb{R}^n \to \{ -1, 1 \} \) a Boolean function, and \( 1 \geq \delta \geq 0 \) we define the

\[ \text{noise sensitivity of } f \text{ with parameter } \delta \text{ to be} \]

\[ \text{NS}_\delta(f) := \Pr(f(A) \neq f(B)), \]

where \( A \) and \( B \) are Bernoulli random variables with \( B \) obtained from \( A \) by
flipping the sign of each coordinate randomly and independently with
probability \( \delta \).
Gaussian Noise Sensitivity

**Definition**

For $f : \mathbb{R}^n \to \{-1, 1\}$ a Boolean function, and $1 \geq \delta \geq 0$ we define the **Gaussian noise sensitivity of $f$ with parameter $\delta$** to be

$$\text{GNS}_\delta(f) := \Pr(f(X) \neq f(Y)),$$

where $X$ and $Y$ are Gaussian random variables that together form a joint Gaussian with

$$\text{Cov}(X_i, Y_j) = \begin{cases} 
(1 - \delta) & \text{if } i = j \\
0 & \text{otherwise} 
\end{cases}.$$
Basic Problem

For $f$ a degree-$d$ PTF, find upper bounds for $\text{NS}_\delta(f)$ and $\text{GNS}_\delta(f)$ in terms of $d$ and $\delta$. 
The Gotsman-Linial Conjecture

The generally believed answer to these questions follows from a Conjecture of Gotsman and Linial via work of Diakonikolas, Raghavendra, Servedio, and Tan:

**Conjecture**

For $f$ a degree-$d$ PTF and $1 > \delta > 0$

$$NS_{\delta}(f), GNS_{\delta}(f) = O(d\sqrt{\delta}).$$

Furthermore, if this Conjecture holds, the bounds would be tight up to constants for $\delta < d^{-2}$. 
Relationship

There is a close relationship between $\text{NS}_\delta(f)$ and $\text{GNS}_{2\delta}(f)$. The random variables $(A, B)$ and $(X, Y)$ agree in their first three moments. Using an appropriate invariance principle, for regular $f$

$$\text{NS}_\delta(f) \approx \text{GNS}_{2\delta}(f).$$
The first non-trivial bounds were proven independently by Harsha-Klivans-Meka and Diakonikolas-Raghavendra-Servedio-Tan in 2009. For a polynomial of Gaussians, or a regular polynomial of Bernoullis, they show that for appropriate $T$ that

$$|p(X)| > T > |p(X) - p(Y)|$$

with high probability. For general polynomials of Bernoullis, they reduce to the above with a regularity lemma. They prove:

$$\mathbb{NS}_\delta(f) \leq 2^{O(d)} \delta^{1/(4d+2)} \log(1/\delta),$$

$$\mathbb{GNS}_\delta(f) \leq O(d \delta^{1/(2d)} \log(1/\delta)^{1/2}).$$
Past Work II

In 2010, K. found an argument to prove the optimal bound for the Gaussian case

$$GNS_\delta(f) = O(d\sqrt{\delta}).$$

- Depends on symmetries that do not exist in Bernoulli case
- Classical invariance principle, cannot be used to get better than $\delta^{1/O(d)}$ bound for Bernoulli case
- New invariance principle can do better
The Bound for Regular PTFs

Proposition

Let $f = sgn \circ p$ be a PTF for $p$ a degree-$d$ polynomial with a $(\tau, N, m, \epsilon)$-regular decomposition. $\text{NS}_\delta(f)$ is at most

$$O(d \sqrt{\delta}) + O(d \epsilon^{1/2d} \log(\epsilon^{-1})) + O_{d,m}(N \tau^{1/5} \log(\tau^{-1})^{dm/2+1}).$$

The proof is by showing that

$$\text{NS}_\delta(f) \approx \text{GNS}_{2\delta}(f) = O(d \sqrt{\delta}).$$
The Bound for General PTFs

Theorem

For $f$ a degree-$d$ PTF, $\delta, c > 0$,

$$\mathbb{NS}_\delta(f) = O_{c,d}(\delta^{1/6-c}).$$

Proof (sketch).

- Use regularity Lemma to write $f$ as a decision tree of depth $\tilde{O}_{c,d}(\delta^{-5/6})$

- With probability $1 - \tilde{O}_{c,d}(\delta^{1/6})$, $A$ and $B$ agree on coordinates defining decision tree

- With probability $1 - \delta^{5/6}$ have $p_\rho$ with either:
  - $p_\rho$ has constant sign with high probability
  - $p_\rho$ has a $(\delta^{5/6}, \delta^{-c}, O_{c,d}(1), O_{c,d}(\delta^{2d}))$-regular decomposition

- In either case, we have $f(A) = f(B)$ with probability $1 - O_{c,d}(\delta^{1/6-c})$
Conclusion

We have presented some of the basic theory of diffuse decompositions and its applications. These ideas are still new and there are several areas of potential improvement. Most notably, one could try to:

- Improve the $d$-dependence of the size of the decompositions
- Find new applications for the theory
Acknowledgements

This work was done with the support of an NSF postdoctoral research fellowship.


Ilias Diakonikolas, Rocco Servedio, Li-Yang Tan, Andrew Wan *A Regularity Lemma, and Low-Weight Approximators, for Low-Degree Polynomial Threshold Functions*, 25th Conference on Computational Complexity (CCC), 2010


