Hypercontractivity and its Applications

Punyashloka Biswal
Advisors: James Lee and Paul Beame

University of Washington

February 10, 2010

Theory

- Problem: smoothing a function
- Log-Sobolev inequality
- Hypercontractivity

Applications

- Dictatorship testing with perfect completeness
- Integrality gap for Unique Games

Problem: smoothing a function

Given a function $f\colon \{-1,1\}^n \to [-1,1]$,
we want a smoothed version $g\colon \{-1,1\}^n \to [-1,1]$

Problem: smoothing a function

Given a function $f\colon \{-1,1\}^n \to [-1,1]$,
we want a smoothed version $g\colon \{-1,1\}^n \to [-1,1]$

1. Values of g should depend linearly on values of f

Problem: smoothing a function

Given a function $f\colon \{-1,1\}^n \to [-1,1]$,
we want a smoothed version $g\colon \{-1,1\}^n \to [-1,1]$

1. Values of g should depend linearly on values of f
2. $\mathbf{E}[g] = \mathbf{E}[f]$

Problem: smoothing a function

Given a function $f\colon \{-1,1\}^n \to [-1,1]$,
we want a smoothed version $g\colon \{-1,1\}^n \to [-1,1]$

1. Values of g should depend linearly on values of f
2. $\mathbf{E}[g] = \mathbf{E}[f]$
3. g should vary less than f

Problem: smoothing a function

Given a function $f\colon \{-1,1\}^n \to [-1,1]$,
we want a smoothed version $g\colon \{-1,1\}^n \to [-1,1]$

1. Values of g should depend linearly on values of f
2. $\mathbf{E}[g] = \mathbf{E}[f]$
3. g should vary less than f
4. $g(x)$ should depend on values of f near x

Global properties from local ones

$$g\colon \{-1,1\}^n \to [-1,1]$$

Global properties from local ones

$$g\colon \{-1,1\}^n \to [-1,1]$$

Suppose we are able to control local variation

$$\textbf{Energy}(g) = \frac{1}{2}\,\mathbf{E}_{x\sim y}\big[(g(x)-g(y))^2\big]$$

Global properties from local ones

$$g\colon \{-1,1\}^n \to [-1,1]$$

Suppose we are able to control local variation

$$\mathbf{Energy}(g) = \tfrac{1}{2}\,\mathbf{E}_{x\sim y}\big[(g(x)-g(y))^2\big]$$

What can we say about the variance?

$$\mathbf{Var}(g) = \mathbf{E}\big[\big(g - \mathbf{E}[g]\big)^2\big]$$

Global properties from local ones

$$g\colon \{-1,1\}^n \to [-1,1]$$

Suppose we are able to control local variation

$$\mathbf{Energy}(g) = \tfrac{1}{2}\,\mathbf{E}_{x\sim y}\big[(g(x)-g(y))^2\big]$$

What can we say about the variance?

$$\mathbf{Var}(g) = \mathbf{E}\Big[\big(g - \mathbf{E}[g]\big)^2\Big]$$

Poincaré

$$\mathbf{Var}(g) \le \tfrac{n}{2}\,\mathbf{Energy}(g)$$

Global properties from local ones

$$g\colon \{-1,1\}^n \to [-1,1]$$

Suppose we are able to control local variation

$$\mathbf{Energy}(g) = \tfrac{1}{2}\,\mathbf{E}_{x\sim y}\big[(g(x)-g(y))^2\big]$$

What can we say about the variance?

$$\mathbf{Var}(g) = \mathbf{E}\big[\big(g - \mathbf{E}[g]\big)^2\big]$$

Poincaré

$$\mathbf{Var}(g) \le \tfrac{n}{2}\,\mathbf{Energy}(g)$$

What can we say about the entropy?

$$\mathbf{Ent}(g) = \mathbf{E}\big[g \ln g - \mathbf{E}[g] \ln \mathbf{E}[g]\big]$$

Global properties from local ones

$$g\colon \{-1,1\}^n \to [-1,1]$$

Suppose we are able to control local variation

$$\mathbf{Energy}(g) = \tfrac{1}{2}\,\mathbf{E}_{x\sim y}\big[(g(x)-g(y))^2\big]$$

What can we say about the variance?

$$\mathbf{Var}(g) = \mathbf{E}\Big[\big(g - \mathbf{E}[g]\big)^2\Big]$$

Poincaré

$$\mathbf{Var}(g) \le \tfrac{n}{2}\,\mathbf{Energy}(g)$$

What can we say about the entropy?

$$\mathbf{Ent}(g) = \mathbf{E}\big[g\ln g - \mathbf{E}[g]\ln\mathbf{E}[g]\big]$$

Log-Sobolev [Gross]

$$\mathbf{Ent}(g^2) \le n\,\mathbf{Energy}(g)$$

Some candidates for g

- $g(x) = f(x)^2$?
  - Not linear

Some candidates for g

- $g(x) = f(x)^2$?
  - Not linear
- $g(x) = \frac{1}{2} f(x)$?
  - Shrinks everything
  - Changes the expectation

Some candidates for g

- $g(x) = f(x)^2$?
  - Not linear
- $g(x) = \mathbf{E}[f]$?
  - Very lossy
- $g(x) = \frac{1}{2} f(x)$?
  - Shrinks everything
  - Changes the expectation

Some candidates for g

- $g(x) = f(x)^2$?
  - Not linear
- $g(x) = \mathbf{E}[f]$?
  - Very lossy
- $g(x) = \frac{1}{2} f(x)$?
  - Shrinks everything
  - Changes the expectation
- $g(x) = \mathbf{E}_{y \text{ near } x}[f(y)]$?
  - Like a blur kernel in graphics
  - Hmm…

The Bonami-Gross-Beckner operator

For any $\rho \in [0, 1]$,

$$\mathbf{T}_\rho[f](x_1, \ldots, x_n) = \mathbf{E}[f(y_1, \ldots, y_n)]$$

where

$$y_i = \begin{cases} x_i & \text{with probability } \frac{1+\rho}{2} \\ -x_i & \text{with probability } \frac{1-\rho}{2} \end{cases}$$

The Bonami-Gross-Beckner operator

For any $\rho \in [0, 1]$,

$$\mathbf{T}_\rho[f](x_1, \ldots, x_n) = \mathbf{E}[f(y_1, \ldots, y_n)]$$

where

$$y_i = \begin{cases} x_i & \text{with probability } \frac{1+\rho}{2} \\ -x_i & \text{with probability } \frac{1-\rho}{2} \end{cases}$$

e.g., $\mathbf{T}_0[f](x) = \mathbf{E}[f]$

The Bonami-Gross-Beckner operator

For any $\rho \in [0,1]$,

$$\mathbf{T}_\rho[f](x_1,\ldots,x_n) = \mathbf{E}[f(y_1,\ldots,y_n)]$$

where

$$y_i = \begin{cases} x_i & \text{with probability } \frac{1+\rho}{2} \\ -x_i & \text{with probability } \frac{1-\rho}{2} \end{cases}$$

e.g., $\mathbf{T}_0[f](x) = \mathbf{E}[f]$ $\quad$ $\mathbf{T}_1[f](x) = f(x)$

p-norms

- For any $f\colon \{-1,1\}^n \to [-1,1]$

$$\begin{aligned} \|f\|_p &= \mathbf{E}\left[|f|^p\right]^{1/p} && 1 \le p < \infty \\ \|f\|_\infty &= \lim_{p\to\infty} \|f\|_p = \max f \end{aligned}$$

p-norms

- For any $f\colon \{-1,1\}^n \to [-1,1]$

$$\|f\|_p = \mathbf{E}\big[|f|^p\big]^{1/p} \qquad 1 \le p < \infty$$
$$\|f\|_\infty = \lim_{p\to\infty} \|f\|_p = \max f$$

- $\|f\|_1 \le \|f\|_2 \le \cdots \le \|f\|_\infty$

- Lower norms pay more attention to the average
  Higher norms pay more attention to spikes

p-norms

- For any $f\colon \{-1,1\}^n \to [-1,1]$

$$\begin{aligned} \|f\|_p &= \mathbf{E}\big[|f|^p\big]^{1/p} & \qquad 1 \le p < \infty \\ \|f\|_\infty &= \lim_{p\to\infty} \|f\|_p = \max f \end{aligned}$$

- $\|f\|_1 \le \|f\|_2 \le \cdots \le \|f\|_\infty$

- Lower norms pay more attention to the average
  Higher norms pay more attention to spikes

Intuition

Noise spreads out the mass of f from its spikes,
so we should be able to bound the higher norms of $\mathbf{T}_\rho[f]$

Hypercontractivity for $\{-1,1\}^n$ [Gross]

For any function $f\colon \{-1,1\}^n \to [-1,1]$ and $1 \le p \le q$, $0 \le \rho \le 1$,

$$\rho \le \sqrt{\frac{p-1}{q-1}} \text{ implies } \| \mathbf{T}_\rho[f] \|_q \le \|f\|_p$$

Hypercontractivity for $\{-1,1\}^n$ [Gross]

For any function $f\colon \{-1,1\}^n \to [-1,1]$ and $1 \le p \le q$, $0 \le \rho \le 1$,

$$\rho \le \sqrt{\frac{p-1}{q-1}} \text{ implies } \| \mathbf{T}_\rho[f] \|_q \le \|f\|_p$$

Application

- For any unbiased boolean function $f(x_1, \ldots, x_n)$ there is an index x_i such that $f(\ldots, x_i, \ldots) \neq f(\ldots, -x_i, \ldots)$ at least $\Omega(\frac{\log n}{n}) \cdot \mathbf{Var}(f)$ of the time. [Kahn, Kalai, Linial]

Hypercontractivity in other spaces

- More generally, if X has notions of probability and distance

Hypercontractivity in other spaces

- More generally, if X has notions of probability and distance
- We can define $\mathbf{Energy}(f) = \frac{1}{2}\,\mathbf{E}_{x \text{ near } y}[(f(x) - f(y))^2]$

Hypercontractivity in other spaces

- More generally, if X has notions of probability and distance
- We can define $\mathbf{Energy}(f) = \frac{1}{2}\,\mathbf{E}_{x \text{ near } y}[(f(x) - f(y))^2]$
- If we can prove a Log-Sobolev inequality $\mathbf{Ent}(f^2) \leq C \cdot \mathbf{Energy}(f)$

Hypercontractivity in other spaces

- More generally, if X has notions of probability and distance
- We can define $\mathbf{Energy}(f) = \frac{1}{2}\,\mathbf{E}_{x \text{ near } y}[(f(x) - f(y))^2]$
- If we can prove a Log-Sobolev inequality $\mathbf{Ent}(f^2) \leq C \cdot \mathbf{Energy}(f)$
- Then we can define a smoothing operator $\mathbf{T}_\rho$ for $0 \leq \rho \leq 1$

Hypercontractivity in other spaces

- More generally, if X has notions of probability and distance
- We can define $\mathbf{Energy}(f) = \frac{1}{2}\,\mathbf{E}_{x \text{ near } y}[(f(x) - f(y))^2]$
- If we can prove a Log-Sobolev inequality $\mathbf{Ent}(f^2) \leq C \cdot \mathbf{Energy}(f)$
- Then we can define a smoothing operator $\mathbf{T}_\rho$ for $0 \leq \rho \leq 1$
- Such that for any $f\colon X \to [-1, 1]$ and $1 \leq p \leq q$,
  we can give an explicit $\rho = \rho(p, q, C)$ such that $\|\,\mathbf{T}_\rho\, f\|_q \leq \|f\|_q$

Hypercontractivity in other spaces

- More generally, if X has notions of probability and distance
- We can define $\mathbf{Energy}(f) = \frac{1}{2}\,\mathbf{E}_{x \text{ near } y}[(f(x) - f(y))^2]$
- If we can prove a Log-Sobolev inequality $\mathbf{Ent}(f^2) \leq C \cdot \mathbf{Energy}(f)$
- Then we can define a smoothing operator $\mathbf{T}_\rho$ for $0 \leq \rho \leq 1$
- Such that for any $f \colon X \to [-1,1]$ and $1 \leq p \leq q$,
  we can give an explicit $\rho = \rho(p, q, C)$ such that $\| \mathbf{T}_\rho f\|_q \leq \|f\|_q$

- e.g., $\mathbb{R}$ with Gaussian measure:

$$\mathbf{Energy}(f) = \frac{1}{2}\,\mathbf{E}[(f')^2]$$

$$\mathbf{Ent}(f^2) \leq \mathbf{Energy}(f)$$

Hypercontractivity in other spaces

- More generally, if X has notions of probability and distance
- We can define $\mathbf{Energy}(f) = \frac{1}{2}\,\mathbf{E}_{x \text{ near } y}[(f(x) - f(y))^2]$
- If we can prove a Log-Sobolev inequality $\mathbf{Ent}(f^2) \le C \cdot \mathbf{Energy}(f)$
- Then we can define a smoothing operator $\mathbf{T}_\rho$ for $0 \le \rho \le 1$
- Such that for any $f \colon X \to [-1, 1]$ and $1 \le p \le q$, we can give an explicit $\rho = \rho(p, q, C)$ such that $\| \mathbf{T}_\rho f \|_q \le \| f \|_q$
- e.g., $\mathbb{R}$ with Gaussian measure:

$$\mathbf{Energy}(f) = \frac{1}{2}\,\mathbf{E}[(f')^2]$$
$$\mathbf{Ent}(f^2) \le \mathbf{Energy}(f)$$

$$\mathbf{T}_\rho[f](x) = \mathbf{E}_{y \sim \mathcal{N}(0,1)}\left[f\left(\rho x + (1-\rho^2)^{1/2} y\right)\right]$$
$$\| \mathbf{T}_\rho[f] \|_q \le \| f \|_p \text{ when } \rho < \sqrt{(q-1)/(p-1)}$$

Applications of HC in other spaces

- **Gaussian**
  Strong isoperimetric inequality for Gaussian space, leading to fast algorithms for graph partitioning [Sherman]

Applications of HC in other spaces

- **Gaussian**
  Strong isoperimetric inequality for Gaussian space, leading to fast algorithms for graph partitioning [Sherman]

- **Schreier graphs**
  Every monotone function from $\{-1,1\}^n$ is $(\frac{1}{2} - \Omega(\frac{\log n}{n}))$-close to one of $\{0, 1, x_1, \ldots, x_n, \mathbf{Maj}(x)\}$. [O'Donnell-Wimmer]

Dictatorship testing

Given a function $f\colon \{-1,1\}^n \to \{-1,1\}$,

Dictatorship testing

Given a function $f\colon \{-1,1\}^n \to \{-1,1\}$,
query it at 3 points $x, y, z \in \{-1,1\}^n$

Dictatorship testing

Given a function $f\colon \{-1,1\}^n \to \{-1,1\}$,
query it at 3 points $x, y, z \in \{-1,1\}^n$ chosen non-adaptively and then

Dictatorship testing

Given a function $f\colon \{-1,1\}^n \to \{-1,1\}$,
query it at 3 points $x, y, z \in \{-1,1\}^n$ chosen non-adaptively and then

1. if f is a dictator
   - i.e., it only depends on only one of its input coordinates

Dictatorship testing

Given a function $f\colon \{-1,1\}^n \to \{-1,1\}$,
query it at 3 points $x, y, z \in \{-1,1\}^n$ chosen non-adaptively and then

1. if f is a dictator
   - i.e., it only depends on only one of its input coordinates
   - you must accept with probability $\geq c$

Dictatorship testing

Given a function $f\colon \{-1,1\}^n \to \{-1,1\}$,
query it at 3 points $x, y, z \in \{-1,1\}^n$ chosen non-adaptively and then

1. if f is a dictator
   - i.e., it only depends on only one of its input coordinates
   - you must accept with probability $\geq c$
2. f is quasirandom
   - i.e., it is far from every small junta

Dictatorship testing

Given a function $f\colon \{-1,1\}^n \to \{-1,1\}$,
query it at 3 points $x, y, z \in \{-1,1\}^n$ chosen non-adaptively and then

1. if f is a dictator
   - i.e., it only depends on only one of its input coordinates
   - you must accept with probability $\geq c$
2. f is quasirandom
   - i.e., it is far from every small junta
   - you must accept with probability $\leq s$

Dictatorship testing

Given a function $f\colon \{-1,1\}^n \to \{-1,1\}$,
query it at 3 points $x, y, z \in \{-1,1\}^n$ chosen non-adaptively and then

1. if f is a dictator
   - i.e., it only depends on only one of its input coordinates
   - you must accept with probability $\geq c$
2. f is quasirandom
   - i.e., it is far from every small junta
   - you must accept with probability $\leq s$

Quasirandomness and Fourier analysis

- Every function $f\colon \{-1,1\}^n \to \mathbb{R}$ can be written as a multilinear polynomial, e.g. $f(x) = \frac{3}{4}x_1 - \frac{1}{2}x_3x_4$

Quasirandomness and Fourier analysis

- Every function $f\colon \{-1,1\}^n \to \mathbb{R}$ can be written as a multilinear polynomial, e.g. $f(x) = \frac{3}{4}x_1 - \frac{1}{2}x_3x_4$
- For any set $S \subseteq [n]$, the coefficient of $\prod_{i \in S} x_i$ is denoted $\hat{f}(S)$

Quasirandomness and Fourier analysis

- Every function $f\colon \{-1,1\}^n \to \mathbb{R}$ can be written as a multilinear polynomial, e.g. $f(x) = \frac{3}{4}x_1 - \frac{1}{2}x_3x_4$
- For any set $S \subseteq [n]$, the coefficient of $\prod_{i \in S} x_i$ is denoted $\hat{f}(S)$
- If $g = \mathbf{T}_\rho[f]$, then $\hat{g}(S) = \rho^{|S|}\hat{f}(S)$

Quasirandomness and Fourier analysis

- Every function $f\colon \{-1,1\}^n \to \mathbb{R}$ can be written as a multilinear polynomial, e.g. $f(x) = \frac{3}{4}x_1 - \frac{1}{2}x_3x_4$
- For any set $S \subseteq [n]$, the coefficient of $\prod_{i\in S} x_i$ is denoted $\hat{f}(S)$
- If $g = \mathbf{T}_\rho[f]$, then $\hat{g}(S) = \rho^{|S|}\hat{f}(S)$
- f is said to be (ϵ,δ)-quasirandom if $|\hat{f}(S)| \leq \epsilon$ whenever $|S| \geq 1/\delta$,

Quasirandomness and Fourier analysis

- Every function $f\colon \{-1,1\}^n \to \mathbb{R}$ can be written as a multilinear polynomial, e.g. $f(x) = \frac{3}{4}x_1 - \frac{1}{2}x_3x_4$
- For any set $S \subseteq [n]$, the coefficient of $\prod_{i\in S} x_i$ is denoted $\hat{f}(S)$
- If $g = \mathbf{T}_\rho[f]$, then $\hat{g}(S) = \rho^{|S|}\hat{f}(S)$
- f is said to be (ϵ, δ)-quasirandom if $|\hat{f}(S)| \leq \epsilon$ whenever $|S| \geq 1/\delta$,

Testing with perfect completeness [O'Donnell-Wu]

For every $0 < \delta < 1/8$, there is a 3-query nonadaptive test that accepts any dictator with probability 1 but accepts any $(\delta, \frac{\delta}{\log(1/\delta)})$-quasirandom function with probability $\leq \frac{5}{8} + O(\sqrt{\delta})$.

The test

- For each $1 \leq i \leq n$, sample (x_i, y_i, y_i) as follows:

x_1	$\dots$	x_i	$\dots$	x_n
y_1	$\dots$	y_i	$\dots$	y_n
z_1	$\dots$	z_i	$\dots$	z_n

The test

- For each $1 \leq i \leq n$, sample (x_i, y_i, y_i) as follows:

x_1	$\ldots$	x_i	$\ldots$	x_n
y_1	$\ldots$	y_i	$\ldots$	y_n
z_1	$\ldots$	z_i	$\ldots$	z_n

  - with probability $1 - \delta$,
    pick x_i, y_i, z_i uniformly from the subset that satisfies $x_i y_i z_i = -1$

The test

- For each $1 \leq i \leq n$, sample (x_i, y_i, y_i) as follows:

x_1	$\dots$	x_i	$\dots$	x_n
y_1	$\dots$	y_i	$\dots$	y_n
z_1	$\dots$	z_i	$\dots$	z_n

  - with probability $1 - \delta$,
    pick x_i, y_i, z_i uniformly from the subset that satisfies $x_i y_i z_i = -1$
  - with probability δ,
    pick $x_i = y_i = z_i$ uniformly between $\{-1, 1\}$

The test

- For each $1 \leq i \leq n$, sample (x_i, y_i, y_i) as follows:

x_1	$\dots$	x_i	$\dots$	x_n
y_1	$\dots$	y_i	$\dots$	y_n
z_1	$\dots$	z_i	$\dots$	z_n

  - with probability $1 - \delta$,
    pick x_i, y_i, z_i uniformly from the subset that satisfies $x_i y_i z_i = -1$
  - with probability δ,
    pick $x_i = y_i = z_i$ uniformly between $\{-1, 1\}$
- Query $f(x), f(y), f(z)$.
- If exactly two of the values are -1, then reject. Otherwise accept.

Analysis: completeness

- $(x_i, y_i, z_i) \in \underbrace{\{(-1,1,1),(1,-1,1),(1,1,-1)\}}_{x_i y_i z_i = -1} \cup \underbrace{\{(-1,-1,-1),(1,1,1)\}}_{x_i = y_i = z_i}$
- Zero, one, or three occurences of -1!
- So if $f(x) = x_i$, our test would pass it. ($c = 1$)

Analysis: soundness

a	-1	-1	-1	-1	1	1	1	1
b	-1	-1	1	1	-1	-1	1	1
c	-1	1	-1	1	-1	1	-1	1
$\texttt{NTW}(a,b,c)$	1	0	0	1	0	1	1	1

Analysis: soundness

a	-1	-1	-1	-1	1	1	1	1
b	-1	-1	1	1	-1	-1	1	1
c	-1	1	-1	1	-1	1	-1	1
$\mathtt{NTW}(a,b,c)$	1	0	0	1	0	1	1	1

- $\mathtt{NTW}(a,b,c) = \frac{5}{8} + \frac{1}{8}(a+b+c) + \frac{1}{8}(ab+bc+ca) - \frac{3}{8}abc$

Analysis: soundness

a	-1	-1	-1	-1	1	1	1	1
b	-1	-1	1	1	-1	-1	1	1
c	-1	1	-1	1	-1	1	-1	1
$\texttt{NTW}(a,b,c)$	1	0	0	1	0	1	1	1

- $\texttt{NTW}(a,b,c) = \frac{5}{8} + \frac{1}{8}(a+b+c) + \frac{1}{8}(ab+bc+ca) - \frac{3}{8}abc$

$$\begin{aligned}\Pr[\text{accept } f] &= \mathbf{E}[\texttt{NTW}(f(x), f(y), f(z))] \\ &= \frac{5}{8} + \frac{3}{8}\,\mathbf{E}[f(x)] + \frac{3}{8}\,\mathbf{E}[f(x)f(y)] - \frac{3}{8}\,\mathbf{E}[f(x)f(y)f(z)]\end{aligned}$$

Analysis: soundness

a	-1	-1	-1	-1	1	1	1	1
b	-1	-1	1	1	-1	-1	1	1
c	-1	1	-1	1	-1	1	-1	1
$\texttt{NTW}(a,b,c)$	1	0	0	1	0	1	1	1

- $\texttt{NTW}(a,b,c) = \frac{5}{8} + \frac{1}{8}(a+b+c) + \frac{1}{8}(ab+bc+ca) - \frac{3}{8}abc$

$$\begin{aligned}\Pr[\text{accept } f] &= \mathbf{E}[\texttt{NTW}(f(x), f(y), f(z))] \\ &= \frac{5}{8} + \frac{3}{8}\,\mathbf{E}[f(x)] + \frac{3}{8}\,\mathbf{E}[f(x)f(y)] - \frac{3}{8}\,\mathbf{E}[f(x)f(y)f(z)]\end{aligned}$$

- Proceed using
  - linearity of expectation
  - Plancherel's theorem: $\mathbf{E}[f^2] = \sum_S \hat{f}(S)^2$
  - elementary algebra

Analysis: soundness

a	-1	-1	-1	-1	1	1	1	1
b	-1	-1	1	1	-1	-1	1	1
c	-1	1	-1	1	-1	1	-1	1
$\texttt{NTW}(a,b,c)$	1	0	0	1	0	1	1	1

- $\texttt{NTW}(a,b,c) = \frac{5}{8} + \frac{1}{8}(a+b+c) + \frac{1}{8}(ab+bc+ca) - \frac{3}{8}abc$

$$\begin{aligned}\Pr[\text{accept } f] &= \mathbf{E}[\texttt{NTW}(f(x), f(y), f(z))] \\ &= \frac{5}{8} + \frac{3}{8}\,\mathbf{E}[f(x)] + \frac{3}{8}\,\mathbf{E}[f(x)f(y)] - \frac{3}{8}\,\mathbf{E}[f(x)f(y)f(z)]\end{aligned}$$

- Proceed using
  - linearity of expectation
  - Plancherel's theorem: $\mathbf{E}[f^2] = \sum_S \hat{f}(S)^2$
  - elementary algebra
- Need to bound $-\frac{3}{8}\,\mathbf{E}[f(x)f(y)f(z)]$

The cubic term

- The contribution due to each $A \subseteq [n]$ can be bounded by

$$4(1-\delta)^{|A|}\left(|\hat{f}(A)|^3 + \| \mathbf{T}_{\sqrt{\delta}}\, g_A\|_3^3\right)$$

where $g_A \colon \{-1,1\}^{[n]\setminus A} \to \mathbb{R}$ is given by

$$\hat{g}_A(X) = \begin{cases} 0 & X = \emptyset \\ \hat{f}(A \cup X) & \text{otherwise} \end{cases}$$

The cubic term

- The contribution due to each $A \subseteq [n]$ can be bounded by

$$4(1-\delta)^{|A|}\left(|\hat{f}(A)|^3 + \|\mathbf{T}_{\sqrt{\delta}}\, g_A\|_3^3\right)$$

where $g_A \colon \{-1,1\}^{[n]\setminus A} \to \mathbb{R}$ is given by

$$\hat{g}_A(X) = \begin{cases} 0 & X = \emptyset \\ \hat{f}(A \cup X) & \text{otherwise} \end{cases}$$

- $\displaystyle\sum_{|A| \le \frac{1}{\delta}\log\frac{1}{\delta}} (1-\delta)^{|A|}|\hat{f}(A)|^3 \le \sqrt{\delta}$

The cubic term

- The contribution due to each $A \subseteq [n]$ can be bounded by

$$4(1-\delta)^{|A|}\left(|\hat{f}(A)|^3 + \| \mathbf{T}_{\sqrt{\delta}}\, g_A\|_3^3\right)$$

where $g_A\colon \{-1,1\}^{[n]\setminus A} \to \mathbb{R}$ is given by

$$\hat{g}_A(X) = \begin{cases} 0 & X = \emptyset \\ \hat{f}(A \cup X) & \text{otherwise} \end{cases}$$

- $\displaystyle\sum_{|A| \le \frac{1}{\delta}\log\frac{1}{\delta}} (1-\delta)^{|A|}|\hat{f}(A)|^3 \le \sqrt{\delta}$

- $\displaystyle\sum_{|A| > \frac{1}{\delta}\log\frac{1}{\delta}} (1-\delta)^{|A|}|\hat{f}(A)|^3 \le (1-\delta)^{1/\delta} \le O(\delta)$

The cubic term

- Goal: bound $\sum_A (1-\delta)^{|A|} \| \mathbf{T}_{\sqrt{\delta}}\, g_A\|_3^3$
- Using a slight variation of the hypercontractive inequality, we have for $\lambda = 1/\log_2(1/\delta) < 1/3$ that

$$\| \mathbf{T}_{\sqrt{\delta}}\, g_A\|_3^3 \leq \| \mathbf{T}_{\sqrt{\delta}}\, g_A\|_2^{3-3\lambda} \|g_A\|_2^{3\lambda}$$

The cubic term

- Goal: bound $\sum_A (1-\delta)^{|A|} \| \mathbf{T}_{\sqrt{\delta}} g_A \|_3^3$
- Using a slight variation of the hypercontractive inequality, we have for $\lambda = 1/\log_2(1/\delta) < 1/3$ that

$$\| \mathbf{T}_{\sqrt{\delta}} g_A \|_3^3 \leq \| \mathbf{T}_{\sqrt{\delta}} g_A \|_2^{3-3\lambda} \|g_A\|_2^{3\lambda}$$

- Plancherel: $\|g_A\|_2^{3\lambda} \leq 1$

The cubic term

- Goal: bound $\sum_A (1-\delta)^{|A|} \| \mathbf{T}_{\sqrt{\delta}}\, g_A \|_3^3$
- Using a slight variation of the hypercontractive inequality, we have for $\lambda = 1/\log_2(1/\delta) < 1/3$ that

$$\| \mathbf{T}_{\sqrt{\delta}}\, g_A \|_3^3 \leq \| \mathbf{T}_{\sqrt{\delta}}\, g_A \|_2^{3-3\lambda} \|g_A\|_2^{3\lambda}$$

- Plancherel: $\|g_A\|_2^{3\lambda} \leq 1$
- Algebraic manipulation: $\| \mathbf{T}_{\sqrt{\delta}}\, g_A \|_2^{3-3\lambda} \leq O(\sqrt{\delta}) \sum_{\emptyset \neq B \subseteq \overline{A}} \delta^{|B|} \hat{f}(A \cup B)^2$

The cubic term

$$\sum_A (1-\delta)^{|A|} \|T_{\sqrt{\delta}} g_A\|_3^3 \quad \leq \quad O(\sqrt{\delta}) \sum_A (1-\delta)^{|A|} \delta^{|B|} \hat{f}(A \cup B)^2$$

The cubic term

$$\sum_A (1-\delta)^{|A|} \|T_{\sqrt{\delta}} g_A\|_3^3 \quad \leq \quad O(\sqrt{\delta}) \sum_A (1-\delta)^{|A|} \delta^{|B|} \hat{f}(A \cup B)^2$$

- Contribution due to each $A \cup B$ is $\sum (1-\delta)^{|A|} \delta^{|B|} = 1$ (Binomial sum)

The cubic term

$$\sum_A (1-\delta)^{|A|} \|T_{\sqrt{\delta}} g_A\|_3^3 \quad \leq \quad O(\sqrt{\delta}) \sum_A (1-\delta)^{|A|} \delta^{|B|} \hat{f}(A \cup B)^2$$

- Contribution due to each $A \cup B$ is $\sum (1-\delta)^{|A|} \delta^{|B|} = 1$ (Binomial sum)
- Total of all $\hat{f}(A \cup B)^2$ contributions is ≤ 1 (Plancherel)

Thank You!

Unique Games

- **Label Cover**: Given a set V of variables over a domain L and weighted constraints on each pair, assign values to maximize the fraction of satisfied constraints.

Unique Games

- **Label Cover**: Given a set V of variables over a domain L and weighted constraints on each pair, assign values to maximize the fraction of satisfied constraints.
- **Unique Label Cover**: As above, but every constraint is a bijection: a constraint on the pair $u, v \in V$ takes the form of a permutation $\pi\colon L \to L$.

Unique Games

- **Label Cover**: Given a set V of variables over a domain L and weighted constraints on each pair, assign values to maximize the fraction of satisfied constraints.

- **Unique Label Cover**: As above, but every constraint is a bijection: a constraint on the pair $u, v \in V$ takes the form of a permutation $\pi \colon L \to L$.
- **Example**: Linear equations of the form $x_u - x_v = r \pmod{p}$

Unique Games

- **Label Cover**: Given a set V of variables over a domain L and weighted constraints on each pair, assign values to maximize the fraction of satisfied constraints.

- **Unique Label Cover**: As above, but every constraint is a bijection: a constraint on the pair $u, v \in V$ takes the form of a permutation $\pi\colon L \to L$.
- **Example**: Linear equations of the form $x_u - x_v = r \pmod{p}$
- **Hardness**
  - Easy when there exists a perfect solution
  - But if there exists a solution satisfying 99% of the constraints, we don't even know how to find a 1% satisfying solution

Unique Games

- **Label Cover**: Given a set V of variables over a domain L and weighted constraints on each pair, assign values to maximize the fraction of satisfied constraints.
- **Unique Label Cover**: As above, but every constraint is a bijection: a constraint on the pair $u, v \in V$ takes the form of a permutation $\pi \colon L \to L$.
- **Example**: Linear equations of the form $x_u - x_v = r \pmod{p}$
- **Hardness**
  - Easy when there exists a perfect solution
  - But if there exists a solution satisfying 99% of the constraints, we don't even know how to find a 1% satisfying solution
- **Unique Games Conjecture**

An SDP for ULC

$$
\begin{aligned}
\text{maximize} \quad & \mathbf{E}_{e\{u,v\}} \textstyle\sum_{i \in L} \langle u_i, v_{\pi_e(i)} \rangle \\
\text{subject to} \quad & \langle u_i, v_j \rangle \geq 0 && \forall u, v \in V, \forall i, j \in L \\
& \textstyle\sum_{i \in L} \langle v_i, v_i \rangle = 1 && \forall v \in V \\
& \langle \textstyle\sum_{i \in L} u_i, \textstyle\sum_{j \in L} v_j \rangle = 1 && \forall u, v \in L \\
& \langle v_i, v_j \rangle = 0 && \forall v \in V, \forall i \neq j \in L
\end{aligned}
$$

An SDP for ULC

$$
\begin{array}{lll}
\text{maximize} & \mathbf{E}_{e\{u,v\}} \sum_{i\in L} \langle u_i, v_{\pi_e(i)} \rangle & \\
\text{subject to} & \langle u_i, v_j \rangle \geq 0 & \forall u, v \in V, \forall i, j \in L \\
& \sum_{i \in L} \langle v_i, v_i \rangle = 1 & \forall v \in V \\
& \langle \sum_{i\in L} u_i, \sum_{j \in L} v_j \rangle = 1 & \forall u, v \in L \\
& \langle v_i, v_j \rangle = 0 & \forall v \in V, \forall i \neq j \in L
\end{array}
$$

Integrality gap [Khot-Vishnoi]

For domain size 2^k and any value $0 < \eta < \frac{1}{2}$, there is a ULC instance whose integer optimum is $\leq 2^{-k\eta}$ but whose SDP admits solutions of value $\geq 1 - \eta$.

Gap instance

- Take $V =$ all functions $f\colon \{-1,1\}^n \to \{-1,1\}$
- Take $L =$ all monomials $\prod_{i\in S} x_i$
- Hard constraints:
  - If $f = g\chi$ for some monomial χ,
    then $\mathbf{Label}(f) = \mathbf{Label}(g)\chi$ must hold
- Fix one f from each group tied by hard constraints

Gap instance

- Take V = all functions $f\colon \{-1,1\}^n \to \{-1,1\}$
- Take L = all monomials $\prod_{i \in S} x_i$
- Hard constraints:
  - If $f = g\chi$ for some monomial χ, then $\mathbf{Label}(f) = \mathbf{Label}(g)\chi$ must hold
- Fix one f from each group tied by hard constraints
- Soft constraints
  - Weight $= \Pr_{h,h'}[\{f,g\} = \{h,h'\}]$ where h, h' are $(1-2\eta)$-correlated
  - Permutation: $\frac{\mathbf{Label}(f\chi)}{\chi} = \frac{\mathbf{Label}(g\psi)}{\psi}$

Soundness

- Objective value is precisely $\Pr[\mathbf{Label}(h) = \mathbf{Label}(h')]$
- Let $\phi\colon V \to \{0,1\}$ indicate the set that received some label χ

Soundness

- Objective value is precisely $\Pr[\mathbf{Label}(h) = \mathbf{Label}(h')]$
- Let $\phi\colon V \to \{0,1\}$ indicate the set that received some label χ

$$\begin{aligned}\Pr[\mathbf{Label}(h) = \mathbf{Label}(h') = \chi] \\ = \mathbf{E}[\phi(h)\phi(h')] = \mathbf{E}[h, \mathbf{T}_{1-2\eta}\, h] = \|T_{\sqrt{1-2\eta}} h\|_2^2\end{aligned}$$

Soundness

- Objective value is precisely $\Pr[\mathbf{Label}(h) = \mathbf{Label}(h')]$
- Let $\phi \colon V \to \{0,1\}$ indicate the set that received some label χ

$$\Pr[\mathbf{Label}(h) = \mathbf{Label}(h') = \chi] \\ = \mathbf{E}[\phi(h)\phi(h')] = \mathbf{E}[h, \mathbf{T}_{1-2\eta}\, h] = \|T_{\sqrt{1-2\eta}} h\|_2^2$$

- By hypercontractivity, $\leq \|h\|^2_{2(1-\eta)} = 1/2^{\frac{k}{1+\eta}}$