CAS CS 599 B: Mathematical Methods for TCS

Lecturer: Mark Bun | Spring 2022

Lecture Notes 7:

Hypercontractivity Continued

Reading.

- O'Donnell, Analysis of Boolean Functions §9.2, 9.5, 9.6

Last time we discussed Bonami's Lemma, which says that low degree functions induce reasonable random variables on the hypercube.

Lemma 1 (Bonami's Lemma)**.** *If $f : \{-1,1\}^n \to \mathbb{R}$ has degree k, then*

$$\mathbb{E}[f(x)]^4 \leq 9^k \mathbb{E}[f(x)^2]^2.$$

1 Hypercontractivity

Let $f : \{-1,1\}^n \to \mathbb{R}$, and let $f^{=k}$ be its degree-k part. The noise operator has a very clean effect on the homogeneous polynomial $f^{=k}$:

$$T_\rho f^{=k}(x) = \sum_{|S|=k} \rho^{|S|} \hat{f}(S) \chi_S(x) = \rho^k f^{=k}(x).$$

Setting $\rho = 1/\sqrt{3}$ and applying Bonami's Lemma,

$$\begin{aligned}\|T_{1/\sqrt{3}} f^{=k}\|_4 &= \frac{1}{\sqrt{3}^k} \|f^{=k}\|_4 \\ &\leq \|f^{=k}\|_2.\end{aligned}$$

It turns out that we can relax the condition that f is a homogeneous polynomial:

Theorem 2 ($(2,4)$-Hypercontractivity Theorem)**.** *For every $f : \{-1,1\}^n \to \mathbb{R}$,*

$$\|T_{1/\sqrt{3}} f\|_4 \leq \|f\|_2.$$

One can conclude this from the calculation above, plus a few nice analytic tricks. Another way to prove it is by just repeating the induction underlying Bonami's Lemma.

One can also conclude Bonami's Lemma from the hypercontractivity theorem. Extending the definition of the noise operator to $\rho > 1$ via $T_\rho f = \sum \rho^{|S|} \hat{f}(S) \chi_S$, we have for all f of degree k that

$$\|f\|_4 = \|T_{1/\sqrt{3}} T_{\sqrt{3}} f\|_4 \leq \|T_{\sqrt{3}} f\|_2 \leq \sqrt{3}^k \|f\|_2.$$

Raising both sides to the 4th power gives Bonami's Lemma.

So the hypercontractivity theorem is basically just a reformulation of Bonami's Lemma, but it says something different. It quantifies the extent to which T_ρ is a "smoothing" operator, i.e., one which mollifies peaks in the distribution of $f(x)$. You saw in the exercises that T_ρ is a contractive map: $\|T_\rho f\|_2 \le \|f\|_2$, for instance. The hypercontractivity theorem says that T_ρ is actually a *hyper*contractive map, in that it smooths out f even when one measures $T_\rho f$ in a higher norm.

With some more machinery (see Section 10.1 of O'Donnell) one can prove the general hypercontractivity theorem.

Theorem 3 (Hypercontractivity Theorem)**.** *Let $f : \{-1,1\}^n \to \mathbb{R}$ and $1 \le p \le q \le \infty$. Then $\|T_\rho f\|_q \le \|f\|_p$ for $0 \le \rho \le \sqrt{\frac{p-1}{q-1}}$.*

We won't prove this, but there is a nice trick to getting another special case "for free" as a consequence of the $(2,4)$-hypercontractivity theorem. It relies on the observation that the noise operator is self-adjoint:

Claim 4. *For $\rho \in \mathbb{R}$ and $f, g : \{-1,1\}^n \to \mathbb{R}$, we have $\langle T_\rho f, g\rangle = \sum \rho^{|S|}\hat{f}(S)\hat{g}(S) = \langle f, T_\rho g\rangle$.*

Corollary 5. *For all $f : \{-1,1\}^n \to \mathbb{R}$,*

$$\|T_{1/\sqrt{3}} f\|_2 \le \|f\|_{4/3}.$$

Proof. Writing $T = T_{1/\sqrt{3}}$, we have

$$\begin{aligned}
\|Tf\|_2^2 &= \langle Tf, Tf\rangle \\
&= \langle f, TTf\rangle \\
&\le \|f\|_{4/3}\|TTf\|_4 && \text{by Hölder's inequality} \\
&\le \|f\|_{4/3}\|Tf\|_2 && \text{by (2, 4)-hypercontractivity.}
\end{aligned}$$

Dividing both sides by $\|Tf\|_2$ gives the claim. □

2 Small Set Expansion

Let's take a closer look at what Corollary 5 is saying. The quantity on the left has a nice combinatorial interpretation. Again letting $T = T_{1/\sqrt{3}}$,

$$\|Tf\|_2^2 = \langle Tf, Tf\rangle = \langle f, TTf\rangle = \langle f, T_{1/3}f\rangle = \mathbf{Stab}_{1/3}[f].$$

In the second-to-last equality, we've used the fact that $T_\rho T_\sigma f = \sum \rho^{|S|}\sigma^{|S|}\hat{f}(S) = T_{\rho\sigma} f$.

Thus, Corollary 5 implies

Corollary 6. *For all $f : \{-1,1\}^n \to \mathbb{R}$, $\mathbf{Stab}_{1/3}[f] \le \|f\|_{4/3}^2$.*

So when $\|f\|_{4/3}$ is small, f is very noise sensitive. This happens, for example, when f is the 0-1 indicator for a small subset A of the Boolean hypercube with $|A| = \alpha 2^n$, whence we have

$$\mathbf{Stab}_{1/3}[\mathbf{1}_A] \le \|\mathbf{1}_A\|_{4/3}^2 = (\mathbb{E}[|\mathbf{1}_A|^{4/3}]^{3/4})^2 = \alpha^{3/2}.$$

Meanwhile, unpacking the LHS:

$$\begin{aligned}
\mathbf{Stab}_{1/3}[\mathbf{1}_A] &= \Pr_{\substack{x\sim\{-1,1\}^n\\ y\sim N_{1/3}(x)}}[x \in A, y \in A] \\
&= \Pr_{x\in\{-1,1\}^n}[x \in A] \cdot \Pr_{\substack{x\sim\{-1,1\}^n\\ y\sim N_{1/3}(x)}}[y \in A \mid x \in A] \\
&= \alpha \cdot \Pr_{\substack{x\sim A\\ y\sim N_{1/3}(x)}}[y \in A].
\end{aligned}$$

Thus we conclude that

$$\Pr_{\substack{x\sim A\\ y\sim N_{1/3}(x)}}[y \in A] \le \sqrt{\alpha}.$$

This can be interpreted as follows. Choose a random point $x \in A$, and then flip each bit independently with probability $1/3$, then we end up at a point y that is outside of A with probability at least $1 - \sqrt{\alpha}$. The bound is essentially tight for Hamming balls. Using the general $(p, 2)$-hypercontractivity theorem, one can prove a stability upper bound of $\alpha^{2/(1+\rho)}$ for any noise value $\rho \in [0, 1]$.

3 The Kahn-Kalai-Linial Theorem

In 1985, Ben-Or and Linial asked the following question in social choice. Let $f : \{-1,1\}^n \to \{-1,1\}$ be an unbiased monotone voting rule. Suppose candidate $b \in \{-1,1\}$ gets to bribe k of the voters, fixing their votes to b. To what extent does this vote-fixing affect the outcome of the election?

This is closely related to the influences of the function f. If just one voter i is bribed to vote for candidate b, then the bias of f changes to

$$\mathbb{E}[f|_{x_i=b}] = \hat{f}(\emptyset) + b\hat{f}(i) = b\mathbf{Inf}_i[f]$$

for unbiased, monotone f. Thus, the power afforded by bribing one voter is given by

$$\mathbf{MaxInf}[f] = \max\{\mathbf{Inf}_i[f] \mid i \in [n]\}.$$

Ben-Or and Linial constructed the "Tribes" function $\mathrm{Tribes}_n : \{-1,1\}^n \to \{-1,1\}$ given by $\mathrm{Tribes}_n = \mathrm{OR}_s(\mathrm{AND}_w, \dots, \mathrm{AND}_w)$ where $s \approx n/\log n$ and $w \approx \log(n/\log n)$. The (awkward) setting of parameters ensures that Tribes_n is nearly unbiased. Moreover, thinking of Tribes_n as a voting rule, voter i is pivotal if and only if a) all other voters in i's tribe vote -1 and b) all other tribes produce the outcome $+1$. For uniformly random votes, this happens with probability $O(2^{-w}) = O(\log n/n)$. Thus, $\mathbf{MaxInf}[\mathrm{Tribes_n}] = O(\log n/n)$.

Ben-Or and Linial made the conjecture, later proved by Kahn, Kalai, and Linial, that this is optimal. Up to constant factors, the Tribes function has the smallest max influence amongst all unbiased functions.

Theorem 7 (KKL Theorem)**.** *For any $f : \{-1,1\}^n \to \{-1,1\}$, we have $\mathbf{MaxInf}[f] \ge \mathrm{Var}[f] \cdot \Omega(\log n/n)$.*

Note that this if a Boolean function is unbiased, $\mathrm{Var}[f] = 1$. Thus, this statement generalizes the above claim to functions f that are not too biased.

Going back to Ben-Or and Linial's original bribery question, one can use the KKL Theorem to show the following. If f is a monotone voting rule that is not too biased, then either candidate b can bribe $O(n/\log n)$ voters to bias the election to, say, 99% in their favor.

For simplicity, we will prove the KKL Theorem when f is unbiased, so $\mathrm{Var}\,[f] = 1$. We'll prove it using the following lemma.

Lemma 8 (Edge-Isoperimetric KKL). *Let $f : \{-1,1\}^n \to \{-1,1\}$ be unbiased. Then*

$$\mathbf{MaxInf}[f] \geq \frac{9}{\mathbf{I}[f]^2} \cdot 9^{-\mathbf{I}[f]}.$$

The reason for the name is as follows. Recall that if f is the $\{\pm 1\}$-indicator for a set A, we can think of $\mathbf{I}[f]$ as n-times the fraction of the hypercube's edges which are on the boundary of A. Poincaré's inequality tells us that $\mathbf{I}[f] \geq \mathrm{Var}\,[f] = 1$ for all balanced functions, and is attained when f is a dictator. KKL tells us that if $\mathbf{I}[f] \leq K$, so A has a near-minimal size boundary, then it must still be similar to a dictator, i.e., there is a coordinate with influence at least $2^{-O(K)}$.

Proof of KKL from Lemma 8. We consider two cases.

Case 1: If $\mathbf{I}[f] \geq \log n/8$, then by averaging, there exists an i such that $\mathbf{Inf}_i[f] \geq \log n/8n$, so we're done.

Case 2: If $\mathbf{I}[f] \leq \log n/8$, then by Lemma 8,

$$\mathbf{MaxInf}[f] \geq \Omega\left(\frac{1}{\log^2 n}\right) \cdot 9^{-\log n/8} = \Omega\left(\frac{1}{\sqrt{n}\log^2 n}\right) = \Omega\left(\frac{\log n}{n}\right).$$

□

Proof of Lemma 8. To prove the lemma, we estimate the noise stability of f in terms of its influences in two ways. One one hand:

$$\mathbf{Stab}_{1/3}[f] = \sum_{S\subseteq[n]} 3^{-|S|}\hat{f}(S)^2 = \mathop{\mathbb{E}}_{S\sim\mathcal{S}_f}\left[3^{-|S|}\right] \geq 3^{-\mathbb{E}[|S|]} = 3^{-\mathbf{I}[f]}. \tag{1}$$

Here, we've used the fact that the function $\varphi(s) = 3^{-s}$ is convex, together with Jensen's inequality.

One the other hand, since f is unbiased,

$$\begin{aligned}\mathbf{Stab}_{1/3}[f] &= \sum_{S\neq\emptyset} 3^{-|S|}\hat{f}(S)^2 \\ &\leq \frac{1}{3}\sum_{S\neq\emptyset} |S|3^{-|S|+1}\hat{f}(S)^2 \\ &= \frac{1}{3}\sum_{i=1}^{n}\sum_{S\ni i} 3^{-|S|+1}\hat{f}(S)^2 \\ &= \frac{1}{3}\sum_{i=1}^{n}\mathbf{Stab}_{1/3}[D_i f].\end{aligned}$$

Now applying Corollary 6 to each term:

$$\mathbf{Stab}_{1/3}[D_i f] \leq \|D_i f\|_{4/3}^2 = (\mathbb{E}[|D_i f|^{4/3}])^{3/2} = \mathbf{Inf}_i[f]^{3/2}.$$

Hence

$$\mathbf{Stab}_{1/3}[f] \leq \frac{1}{3}\sum_{i=1}^{n} \mathbf{Inf}_i[f]^{3/2} \leq \frac{1}{3}\sqrt{\mathbf{MaxInf}[f]} \cdot \sum_{i=1}^{n} \mathbf{Inf}_i[f] = \frac{1}{3}\sqrt{\mathbf{MaxInf}[f]} \cdot \mathbf{Inf}[f]. \quad (2)$$

Combining (1) and (2) gives the lemma. □

Related ideas can be used to prove Friedgut's junta theorem:

Theorem 9. *For every $0 < \varepsilon \leq 1$ and every $f : \{-1,1\}^n \to \{-1,1\}$, the function f is ε-close to an $\exp(O(\mathbf{I}[f]/\varepsilon))$-junta.*

4 Other Applications

We now state a few other applications of the general $(2,q)$-hypercontractivity theorem.

Theorem 10 ($(2,q)$-Hypercontractivity)**.** *Let $f : \{-1,1\}^n \to \mathbb{R}$ and $2 \leq q \leq \infty$. Then $\|T_\rho f\|_q \leq \|f\|_2$ for $0 \leq \rho \leq \sqrt{\frac{1}{q-1}}$.*

The same argument we used to deduce Bonami's Lemma from $(4,2)$-hypercontractivity yields

Theorem 11. *Let $f : \{-1,1\}^n \to \mathbb{R}$ have degree at most k. Then $\|f\|_q \leq (q-1)^{k/2}\|f\|_2$ for every $q \geq 2$.*

This lets us derive exponential concentration bounds for low-degree polynomials. Recall that the Chernoff bound tells us that the linear function $\sum_{i=1}^n a_i x_i$ exceeds t standard deviations with probability $\exp(-\Omega(t^2))$. Degree-k polynomials exceed t standard deviations with probability $\exp(-\Omega(t^{2/k}))$.

Theorem 12. *Let $f : \{-1,1\}^n \to \mathbb{R}$ have degree at most k. Then for any $t \geq (2e)^{k/2}$,*

$$\Pr_{x\sim\{-1,1\}^n}[|f(x)| \geq t\|f\|_2] \leq \exp\left(-\frac{k}{2e}t^{2/k}\right).$$

Proof. By normalizing f, we may assume $\|f\|_2 = 1$. Let $q \geq 2$ be a parameter to be chosen later. Then

$$\begin{aligned}
\Pr[|f(x)| \geq t] &= \Pr[|f(x)|^q \geq t^q] \\
&\leq t^{-q}\mathbb{E}[|f(x)|]^q && \text{by Markov} \\
&\leq t^{-q}(q-1)^{kq/2}\|f\|_2^q && \text{by hypercontractivity} \\
&\leq \left(\frac{q^{k/2}}{t}\right)^q.
\end{aligned}$$

Now set $q = t^{2/k}/e$. □