Analysis of Boolean Functions **(CMU 18-859S, Spring 2007)**

Lecture 16: The Hypercontractivity Theorem

Mar. 8, 2007

Lecturer: Ryan O'Donnell *Scribe: Ryan O'Donnell*

1 Statement and history

In this lecture, we prove the full-blown hypercontractivity theorem for $\{-1,1\}^n$. The idea behind the statement is that the T_ρ operator smooths, or "reasonable-izes", functions.

Theorem 1.1 *(**Hypercontractivity Theorem.**) Let $1 \le p \le q \le \infty$. Provided*

$$\rho \le \sqrt{\frac{p-1}{q-1}},$$

it holds that for all $f : \{-1,1\}^n \to \mathbb{R}$,

$$\|T_\rho f\|_q \le \|f\|_p$$

Some examples, including two that were mentioned in the last lecture:

Example 1.2

- $q = 4$, $p = 2$, $\rho = 1/\sqrt{3}$: $\quad \|T_{1/\sqrt{3}} f\|_4 \le \|f\|_2$
- $q = 2$, $p = 4/3$, $\rho = 1/\sqrt{3}$: $\quad \|T_{1/\sqrt{3}} f\|_2 \le \|f\|_{4/3}$
- $q = q$, $p = 2$, $\rho = 1/\sqrt{q-1}$: $\quad \|T_{1/\sqrt{q-1}} f\|_4 \le \|f\|_2$

One corollary of the last of these is often quite sufficient; it's also a generalization of the original $(2,4,1/\sqrt{3})$-hypercontractivity result we proved easily by induction:

Corollary 1.3 *Let $f : \{-1,1\}^n \to \mathbb{R}$ have degree at most d. Then $\|f\|_q \le \sqrt{q-1}^d \|f\|_2$.*

Proof:

$$\begin{aligned}
\|f\|_q^2 = \|\sum_{k=0}^d f^{=k}\|_q^2 &= \|T_{1/\sqrt{q-1}} \left(\sum_{k=0}^d \sqrt{q-1}^k f^{=k} \right) \|_q^2 \\
&\le \|\sum_{k=0}^d \sqrt{q-1}^k f^{=k}\|_2^2 \\
&= \sum_{k=0}^d (q-1)^k \sum_{|S|=k} \hat{f}(S)^2 \\
&\le (q-1)^d \sum_S \hat{f}(S)^2 = (q-1)^d \|f\|_2^2,
\end{aligned}$$

and the result follows after taking a square-root. □

1.1 History

The history of Theorem 1.1 is quite involved, and interesting. To understand the history, it's important to first note that Theorem 1.1 has a "Gaussian" version. Specifically, imagine that one only considers f's that can be expressed in the form

$$f(x_1, \ldots, x_n, x_{n+1}, \ldots, x_{2n}, \ldots, x_{mn}) = h\left(\frac{x_1 + \cdots + x_n}{\sqrt{n}}, \ldots, \frac{x_{(m-1)n+1} + \cdots + x_{mn}}{\sqrt{n}}\right),$$

where $h : \mathbb{R}^m \to \mathbb{R}$. By the Central Limit Theorem, $(x_{(j-1)n+1} + \cdots + x_{jn})/\sqrt{n}$ has a distribution very close to that of a standard Gaussian random variable, at least for n large. It follows (this was observed by Gross [4]) that Theorem 1.1 must also hold in the "Gaussian setting", where $f : \mathbb{R}^n \to \mathbb{R}$, the domain $\mathbb{R}^n$ is thought of as having the n-dimensional Gaussian distribution, and T_ρ is an appropriately generalized linear operator (specifically, the "Ornstein-Uhlenbeck operator" from mathematical physics, which we'll encounter later).

An early version of Theorem 1.1 in the Gaussian setting was proved by Edward Nelson of Princeton in the mid-60's [6]; I think it was the $q = 4$, $p = 2$ case, possibly with a constant bigger than 1 on the right side. This was in an important paper on quantum field theory. Several works in subsequent years (e.g., Glimm '68, Federbush '69) improved on the result, and the culmination was the complete proof of Theorem 1.1 in the Gaussian setting, due to Nelson again [7]. Nelson's two papers won him the Steele Prize. He is an interesting character, having gone on to work on foundations of mathematics, bounded arithmetic, and automatic proof verification; he is now well-known for having invented Internal Set Theory, a partial axiomatization of Nonstandard Analysis.

In 1973, Leonard Gross proved a limiting version of the theorem called a Logarithmic Sobolev Inequality, and deduced Nelson's Hypercontractive theorem from it [4]. His proof was in the boolean setting, getting the Gaussian setting via the Central Limit Theorem. However, it turned out that the proof of Theorem 1.1 (in the boolean setting) was not new; it was first proved by Aline Bonami [3] in 1979. In fact, Bonami had proved the $q = 4$, $p = 2$ case in the boolean setting even earlier [2].

The history of the theorem can be traced even further back; Bonami's work was informed by that of Walter Rudin, who proved [8] similar inequalities in the setting of $\mathbb{Z}_n$ rather than $\{-1, 1\}^n$ ("one of my favorite papers" — Rudin). Further, a version of the log-sobolev inequality in the Euclidean (rather than Gaussian) setting was proved by A. J. Stam [9] in 1959, in work on Fisher/Shannon information theory — much closer to the world of computer science!

Finally, Theorem 1.1 was introduced to the world of theoretical computer science in the work of Kahn, Kalai, and Linial [5]. Unfortunately, they attributed the theorem to William Beckner [1], which is not really an accurate accreditation. Beckner work was in fact important followup work on the work of Nelson and Gross, making extensions to Euclidean and complex settings.

In the computer science theory literature, Theorem 1.1 is often called "Beckner's Theorem". Lately there has been a move towards "Bonami-Beckner Theorem", although "Bonami Theorem" would respect the original discoverer and "Bonami-Gross Theorem" might more properly respect the independent discovery. To sidestep the issue, we will simply call it the Hypercontractivity Theorem.

1.2 The proof

The proof is in two parts:

Part 1. Prove Theorem 1.1 in the case $n = 1$. This is called the "Two-Point Inequality" because (if p, q, ρ are given) it depends on only two real variables, $f(1)$ and $f(-1)$. The Two-Point Inequality is therefore considered "elementary"; but, it's tricky.

Part 2. Induction on n.

It must be said that both parts are a little annoying to carry out.
We will do them in the opposite order.

2 Part 2: Induction

The induction ultimately only uses two things:

- The triangle inequality for $\|\cdot\|_{q/p}$; i.e., $\|g+h\|_{q/p} \leq \|g\|_{q/p} + \|h\|_{q/p}$.

  Note that $\|\cdot\|_r$ is a norm (satisfies the triangle inequality) for all $r \geq 1$ — and our $r = q/p$ is at least 1 since $q \geq p$.

- A lot of notation.

We will keep p, q, and ρ satisfying the conditions of the theorem fixed throughout.

We will consider a partition of the coordinates $[n]$ into I and $\bar{I}$, and we will write a generic string in $\{-1,1\}^n$ as (x, y), where $x \in \{-1,1\}^I$ and $y \in \{-1,1\}^{\bar{I}}$.

We will prove the Theorem for functions $f : \{-1,1\}^n \to \mathbb{R}$ using the fact that it inductively holds for functions $\{-1,1\}^I \to \mathbb{R}$ and $\{-1,1\}^{\bar{I}} \to \mathbb{R}$. We could have insisted simply that $|I| = 1$ if we wanted, but doing this is actually no simpler or clearer.

We begin with:

$$\begin{aligned}
\|T_\rho f\|_q &= \left(\mathop{\mathbf{E}}_{\boldsymbol{y}} \mathop{\mathbf{E}}_{\boldsymbol{x}} [|(T_\rho f)(\boldsymbol{x}, \boldsymbol{y})|^q] \right)^{1/q} \\
&= \left(\mathop{\mathbf{E}}_{\boldsymbol{y}} \mathop{\mathbf{E}}_{\boldsymbol{x}} [|(T_\rho f)_{\boldsymbol{y}\to\bar{I}}(\boldsymbol{x})|^q] \right)^{1/q} \\
&= \mathop{\mathbf{E}}_{\boldsymbol{y}} \left[\|(T_\rho f)_{\boldsymbol{y}}\|_{q \text{ (fcn of } \boldsymbol{x})}^q \right]^{1/q}. \quad (*)
\end{aligned}$$

We now wish to understand $(T_\rho f)_y$, as a function of $x \in \{-1,1\}^I$.

$$T_\rho f = \sum_{S\subseteq[n]} \rho^{|S|}\hat{f}(S)\chi_S = \sum_{A\subseteq I}\sum_{B\subseteq \bar{I}} \rho^{|A|}\rho^{|B|}\hat{f}(A\cup B)\chi_A\chi_B.$$

Hence, as a function of $x \in \{-1,1\}^I$,

$$\begin{aligned}(T_\rho f)_y &= \sum_{A\subseteq I}\rho^{|A|}\left(\sum_{B\subseteq\bar{I}}\rho^{|B|}\hat{f}(A\cup B)\chi_B(y)\right)\chi_A \\ &= T_\rho g_y,\end{aligned}$$

where $g_y : \{-1,1\}^I \to \mathbb{R}$ is defined by

$$g_y = \sum_{A\subseteq I}\left(\sum_{B\subseteq\bar{I}}\rho^{|B|}\hat{f}(A\cup B)\chi_B(y)\right)\chi_A. \tag{1}$$

Continuing:

$$\begin{aligned}(*) &= \mathop{\mathbf{E}}_{\boldsymbol{y}}\left[\|T_\rho g_{\boldsymbol{y}}\|^q_{q\text{ (fcn of }\boldsymbol{x})}\right]^{1/q} \\ \text{(by induction)}\quad &\leq \mathop{\mathbf{E}}_{\boldsymbol{y}}\left[\|g_{\boldsymbol{y}}\|_p^q\right]^{1/q} \\ &= \mathop{\mathbf{E}}_{\boldsymbol{y}}\left[\mathop{\mathbf{E}}_{\boldsymbol{x}}[|g_{\boldsymbol{y}}(\boldsymbol{x})|^p]^{q/p}\right]^{1/q} \\ &= \mathop{\mathbf{E}}_{\boldsymbol{y}}[(\text{something non-neg.})^{q/p}]^{1/q} \\ &= \|\text{something}\|^{p/q\cdot 1/q}_{q/p\text{ (fcn of }\boldsymbol{y})} \\ &= \left(\|\quad \mathop{\mathbf{E}}_{\boldsymbol{x}}[|g_{\boldsymbol{y}}(\boldsymbol{x})|^p]\quad\|_{q/p\text{ (fcn of }\boldsymbol{y})}\right)^{1/p}. \quad (**)\end{aligned}$$

We now use the triangle inequality for $\|\cdot\|_{q/p}$. Note that inside the $\|\cdot\|_{q/p}$ we have an expectation over $\boldsymbol{x}$, which is just a constant times a sum over $\boldsymbol{x}$. Pulling the constant out of the $\|\cdot\|_{q/p}$ and then using the triangle inequality, we have that $\|\mathbf{E}_{\boldsymbol{x}}[\text{anything}]\|_{q/p} \leq \mathbf{E}_{\boldsymbol{x}}[\|\text{anything}\|_{q/p}]$. Continuing:

$$\begin{aligned}(**) &\leq \left(\mathop{\mathbf{E}}_{\boldsymbol{x}}\left[\|\,|g_{\boldsymbol{y}}(\boldsymbol{x})|^p\,\|_{q/p}\right]\right)^{1/p} \\ &= \left(\mathop{\mathbf{E}}_{\boldsymbol{x}}\left[\mathop{\mathbf{E}}_{\boldsymbol{y}}[|g_{\boldsymbol{y}}(\boldsymbol{x})|^q]^{p/q}\right]\right)^{1/p} \\ &= \left(\mathop{\mathbf{E}}_{\boldsymbol{x}}\left[\|g_{\boldsymbol{y}}(\boldsymbol{x})\|^p_{q\text{ (fcn of }\boldsymbol{y})}\right]\right)^{1/p}. \quad (***)\end{aligned}$$

We now would like to understand $g_y(x)$ as a function of $y \in \{-1,1\}^{\bar{I}}$. From (1) we have

$$\begin{aligned} g_y(x) &= \sum_{A \subseteq I} \left(\sum_{B \subseteq \bar{I}} \rho^{|B|} \hat{f}(A \cup B) \chi_B(y) \right) \chi_A(x) \\ &= \sum_{B \subseteq \bar{I}} \rho^{|B|} \left(\sum_{A \subseteq I} \hat{f}(A \cup B) \chi_A(x) \right) \chi_B(y). \end{aligned}$$

Hence we see that $g_y(x) = T_\rho h$ for some function h of y, and that this function h has as its Fourier expansion

$$h = \sum_{B \subseteq \bar{I}} \left(\sum_{A \subseteq I} \hat{f}(A \cup B) \chi_A(x) \right) \chi_B = \sum_{S \subseteq [n]} \hat{f}(S) \chi_{S \cap I}(x) \chi_{S \cap \bar{I}}.$$

So h is nothing more than the restriction of f given by fixing x for the coordinates I. I.e., as a function of y, $g_y(x) = T_\rho f_{x \to O}$. Continuing once more:

$$\begin{aligned} (***) &= \left(\mathop{\mathbf{E}}_{\boldsymbol{x}} [\| T_\rho f_{\boldsymbol{x} \to I} \|_q^p] \right)^{1/p} \\ \text{(by induction)} \quad &\leq \left(\mathop{\mathbf{E}}_{\boldsymbol{x}} [\| f_{\boldsymbol{x} \to I} \|_p^p] \right)^{1/p} \\ &= \left(\mathop{\mathbf{E}}_{\boldsymbol{x}} [\mathop{\mathbf{E}}_{\boldsymbol{y}} [|f_{\boldsymbol{x} \to I}(\boldsymbol{y})|^p]] \right)^{1/p} \\ &= (\mathbf{E}[|f|^p])^{1/p} \\ &= \|f\|_p. \end{aligned}$$

The induction is complete.

3 Part 1: The Two-Point Inequality

This is the $n = 1$ base case. In this case, any function $f : \{-1,1\} \to \mathbb{R}$ can be represented by two real numbers, $a = f(1)$ and $b = f(-1)$. Applying T_ρ to f gives the function with two values

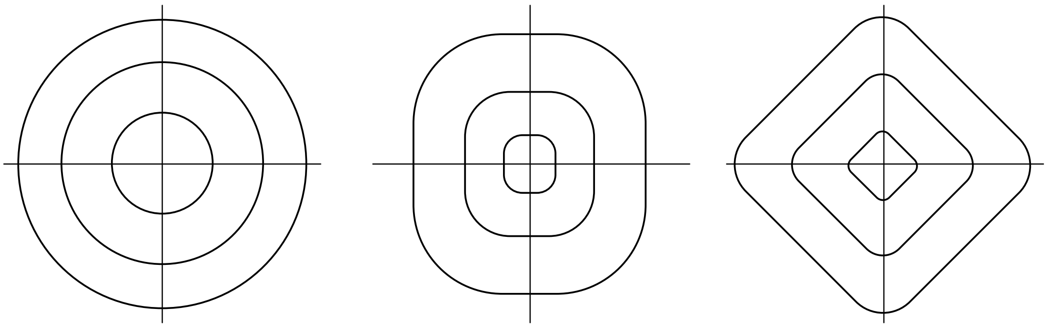
$$T_\rho f(1) = (\tfrac{1+\rho}{2})a + (\tfrac{1-\rho}{2})b, \qquad T_\rho f(-1) = (\tfrac{1-\rho}{2})a + (\tfrac{1+\rho}{2})b.$$

We will think of all functions $f : \{-1,1\} \to \mathbb{R}$ as points in the plane, $(a,b) \in \mathbb{R}^2$, in which case the functions $T_\rho f$ are represented by all points on the line segment joining (a,b) to (b,a). When $\rho = 1$, $T_\rho f$ agrees with f; and as $\rho \to 0$, the resulting function/point moves towards the midpoint of the line segment (a,b)-(b,a).

Given a and b, we wish to find the largest ρ, as a function of p and q, so that $\|T_\rho(a,b)\|_q \leq \|(a,b)\|_p$, where we are identifying functions and points here.

Think of the number $\|(a,b)\|_p$ as being fixed. The set of all points (functions) (a,b) that achieve this number is a kind of "level set". Specifically, it is the "ℓ_p sphere"; the set of points $(a,b) \in \mathbb{R}^2$

such that $(\frac{|a|^p+|b|^p}{2})^{1/p}$ achieves some fixed value. When $p = 2$, for example, these level sets are circles. For larger p, these circles become more like squares with rounded corners; in the limit $p = \infty$, the level sets are indeed squares. On the other hand, for $p < 2$, the level sets become shaped more like rounded diamonds, with the limiting case $p = 1$ indeed giving diamonds.



$|| \cdot ||_2$ level set

$|| \cdot ||_r$ level set, *r* large

$|| \cdot ||_r$ level set, *r* small

To see how these level sets compare with one another, simply observe that all norms have the same value on constant functions; i.e., $\|(a, a)\|_p$ is independent of p. Hence for a given "level", all ℓ_p spheres touch at the points $(\pm a, \pm a)$.

Now again, think of the number $\|(a, b)\|_p$ as being fixed, with a and b "varying". We think of p as being small, so we've drawn a flattish curve in the left diagram below, with various (a, b) and (b, a) pairs on it. We've also drawn in a very squarelike curve in the same diagram, representing the ℓ_q curve at the same level (recall that q is larger).

Wherever (a, b) — equivalently, (b, a) — is on the ℓ_p curve, $T_\rho(a, b)$ is somewhere on the line segment joining these points. We require that $\|T_\rho(a, b)\|_q$ be at most $\|(a, b)\|_p$, so we are effectively asking —

"How far towards the middle of the line segment do we need to go to get inside the ℓ_q curve?"

various points/functions on a fixed $|| \; ||_p$ curve, with the T_ρ values shown dotted, and a $|| \; ||_q$ curve also shown

for nearby points, one has to go almost all the way to the midpoint to get inside the *q* curve

At this point, we will make an unjustified, pictorial, claim: The "worst case" is when a and b are close together. From the diagram on the right, above, one can see that when a and b are far apart, one only has to go a modest distance inward along the line segment to get inside the ℓ_q curve (i.e., ρ need not be that small). But when a and b are close, one has to go almost all the way to the midpoint (i.e., ρ has to get close to 0). The diagram only illustrates $a, b > 0$, but it's the same when $a, b < 0$, and when the two have opposite signs, they are quite far apart and things are only easier.

We conclude that the most constraining case for ρ is when a and b are very near. Since the picture is scale-invariant, we can take $a = 1 + \epsilon$ and $b = 1 - \epsilon$, for $eps \to 0^+$. (Note that $\epsilon = 0$, i.e., $a = b$, is actually not a hard case; here ρ can be 1.) So we are trying to understand how small ρ need be so that

$$\begin{aligned} & \|T_\rho(1+\epsilon, 1-\epsilon)\|_q \leq \|(1+\epsilon, 1-\epsilon)\|_p \\ \Leftrightarrow \quad & \|(1+\rho\epsilon, 1-\rho\epsilon)\|_q \leq \|(1+\epsilon, 1-\epsilon)\|_p \\ \Leftrightarrow \quad & \left(\frac{(1+\rho\epsilon)^q + (1-\rho\epsilon)^q}{2}\right)^{1/q} \leq \left(\frac{(1+\epsilon)^p + (1-\epsilon)^p}{2}\right)^{1/p}. \end{aligned} \tag{2}$$

Now by the (generalized) Binomial Theorem,

$$(1+\rho\epsilon)^q = 1 + q\rho\epsilon + \frac{q(q-1)}{2!}\rho^2\epsilon^2 + \frac{q(q-1)(q-1)}{3!}\rho^3\epsilon^3 + \cdots$$

and hence

$$\frac{(1+\rho\epsilon)^q + (1+\rho\epsilon)^q}{2} = 1 + \frac{q(q-1)}{2}\rho^2\epsilon^2 + O(\epsilon^4).$$

Using this on the left-hand side of (2), and further using the expansion $(1+\delta)^q = 1 + q\delta + O(\delta^2)$, we get that the left-hand side of (2) is

$$\text{LHS} = 1 + \frac{q-1}{2}\rho^2\epsilon^2 + O(\epsilon^4).$$

Doing a similar expansion for the right-hand side of (2) yields

$$\text{RHS} = 1 + \frac{p-1}{2}\epsilon^2 + O(\epsilon^4).$$

Hence as $\epsilon \to 0$, we see that LHS $\leq$ RHS if and only if

$$\frac{q-1}{2}\rho^2 \leq \frac{p-1}{2} \qquad \Leftrightarrow \qquad \rho \leq \sqrt{\frac{p-1}{q-1}},$$

as required by the theorem.

We remark that making this argument rigorous is quite easy; one only needs to use Bernoulli's inequality and compare the series expansions from the generalized Binomial Theorem term-by-term.

References

[1] W. Beckner. Inequalities in Fourier analysis. *Ann. of Math.*, 102(1):159–182 (1975).

[2] A. Bonami. Ensembles $\Lambda(p)$ dans le dual de D^∞. *Ann. Inst. Fourier*, 18(2):193–204 (1968).

[3] A. Bonami. Étude des coefficients de Fourier des fonctions de $L^p(G)$. *Ann. Inst. Fourier*, 20:335–402 (1970).

[4] L. Gross. Logarithmic Sobolev inequalities. *Amer. J. Math.* 97, 1061-1083 (1975).

[5] J. Kahn, G. Kalai, and N. Linial. The influence of variables on boolean functions. In *Proceedings of the 29th Annual Symposium on Foundations of Computer Science*, 68–80 (1988).

[6] E. Nelson. A quartic interaction in two dimensions. *In Mathematical Theory of Elementary Particles*, MIT Press, 69–73 (1966).

[7] E. Nelson. Construction of quantum fields from Markoff fields. J. Functional Analysis 12, 97–112 (1973).

[8] W. Rudin. Trigonometric series with gaps. *J. of Math. and Mech.* 9, 203–227 (1960).

[9] A. J. Stam. Some inequalities satisfied by the quantities of information of Fisher and Shannon. *Information and Control* 2, 101–112 (1959).