CS 880: Advanced Complexity Theory	2/15/2008
Lecture 10: Hypercontractivity	
Instructor: Dieter van Melkebeek	Scribe: Baris Aydinlioglu

This is a technical lecture throughout which we prove the hypercontractivity of the noise operator, a result that will be used in later lectures. The reader may wish to review the notes of lecture 6 for a discussion of the noise operator T_α, the p-norm of a function from the Boolean cube to the reals, and the notion of hypercontractivity.

1 Hypercontractivity Theorem

Recall the definition of the noise operator on functions $f : \{-1, 1\}^n \to \mathbb{R}$:

$$(T_\alpha f)(x) = \mathop{\mathrm{E}}_{y \sim_\epsilon x}[\, f(y)\,], \text{ with } \epsilon = \tfrac{1-\alpha}{2},$$

where $y \sim_\epsilon x$ refers to the string y obtained by flipping each bit of x independently with probability ϵ. Also recall that

$$(T_\alpha f)(x) = \sum_{S \subseteq [n]} \alpha^{|S|} \hat{f}(S) \chi_S(x) \; .$$

Intuitively, the noise operator has a "smoothening effect" on a function, in the sense that the resulting function is a weighted average of the original function around some neighborhood of its argument. In the Fourier spectrum, the effect is that the higher frequencies get dampened out. One consequence of this is hypercontractivity.

Theorem 1. *For all p, q, α such that $1 \leq p \leq q \leq \infty$ and $|\alpha| \leq \sqrt{\frac{p-1}{q-1}}$, and for all functions $f : \{-1, 1\}^n \to \mathbb{R}$,*

$$\left\|T_\alpha f\right\|_q \leq \left\|f\right\|_p \; . \tag{1}$$

Remark: If $\alpha = 0$ then (1) always holds, for then $\left\|T_\alpha f\right\|_q = \left|\mathrm{E}[f]\right| \leq (\, \mathrm{E}[\,|f|^p\,]\,)^{\frac{1}{p}} = \left\|f\right\|_p$, where the inequality follows from Hölder's inequality: $\mathrm{E}\big[fg\big] \leq \left\|f\right\|_p \left\|g\right\|_q$, if $\frac{1}{p} + \frac{1}{q} = 1$ with $p, q \geq 1$.

If $\alpha = \pm 1$ then (1) fails unless $p = q$ or f is constant in absolute value. This follows because $(T_{\pm 1} f)(x) = f(\pm x)$, where $-x$ denotes x with all its bits flipped, and because the only functions f for which $\left\|f\right\|_p = \left\|f\right\|_q$ for $p \neq q$ are those that are constant in absolute value.

In proving the theorem we will arrive at the condition in the statement of the theorem on $|\alpha|$ as the weakest one that guarantees (1) to hold.

2 Proof of the Hypercontractivity Theorem

The proof is by induction on n. In the base case we develop the condition on α and in the inductive step we maintain it.

2.1 Base Case

The base case is for $n = 1$. Note that in this case we can represent any function $f : \{-1, 1\}^n \to \mathbb{R}$ with the point $\big(f(-1), f(1)\big)$ in 2-space. In what follows we say f to mean either the function or the point that it corresponds to in 2-space, which is clear in context.

We want to show

$$\sqrt{\tfrac{p-1}{q-1}} \le \min_f \ \max\left\{\beta :\ \forall|\alpha| \le \beta \ \left\|T_\alpha f\right\|_q \le \left\|f\right\|_p\right\} \ . \tag{2}$$

Let $f(-1) = a$ and $f(1) = b$. Then $(T_\alpha f)(-1) = \frac{1+\alpha}{2}\,a + \frac{1-\alpha}{2}\,b$, and $(T_\alpha f)(1) = \frac{1+\alpha}{2}\,b + \frac{1-\alpha}{2}\,a$. WLOG suppose $a, b > 0$.

Notice that as a point in 2-space $(T_\alpha f)$ is a convex combination of the points (a, b) and (b, a), hence resides somewhere on the line segment joining (a, b) and (b, a). As α gets closer to 1, $(T_\alpha f)$ gets closer to (a, b), which is the point for f. As α gets closer to -1, $(T_\alpha f)$ gets closer to (b, a), which is the point for $T_{-1}f$. For $\alpha = 0$, $(T_\alpha f)$ is the midpoint of the line segment $[(a, b), (b, a)]$.

Recall from Lecture 6 that the p-norm of a vector corresponds to the amount that the unit p-circle should be scaled so that the vector is on the scaled circle. From this it follows that given f, asking for the largest β for which inequality (1) holds for all $|\alpha| \le \beta$ amounts to asking for the largest $|\alpha|$ such that, if we scale the unit p-circle so that f is on it, and if we scale the unit q-circle by the same amount, $(T_\alpha f)$ remains within the (scaled) q-circle.

Note that for fixed (a, b), $(T_\alpha f)$ is linear in α, and therefore the last question is equivalent to asking for the fraction of the line segment $[(a, b), (b, a)]$ that falls within the q-circle.

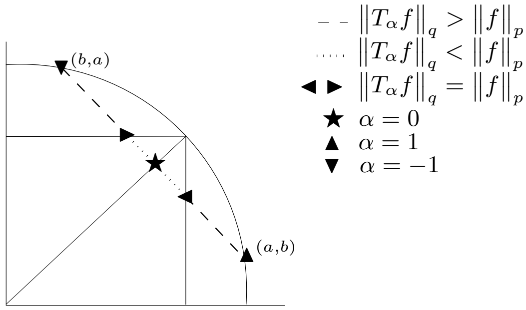

Figure 1: T_α resides on the line segment $[(a, b), (b, a)]$, where $(a, b) = (f(-1), f(1))$. The outer circle represents the unit p-circle scaled by $\left\|f\right\|_p$, and the inner circle represents the unit q-circle scaled by the same amount. Note that the p-circle and q-circle meet on the diagonal. The figure is for $p = 2$ and $q = \infty$.

We state without proof that this ratio decreases as we make a and b get closer while keeping $\left\|f\right\|_p$ fixed. We can see that this is plausible if we take $p = 2$ and $q = \infty$; see figure 2.1.

Letting $\delta = \frac{a-b}{a+b}$, from the foregoing it follows that we want to find, as δ approaches zero, the largest $|\alpha|$ such that $\left\|T_\alpha f\right\|_q \le \left\|f\right\|_p$. We do this in the rest of this section.

First we write $\left\|T_\alpha f\right\|_q$ and $\left\|f\right\|_p$ in terms of δ:

$$\left\|T_\alpha f\right\|_q = \left(\frac{(T_\alpha f(-1))^q + (T_\alpha f(1))^q}{2}\right)^{\frac{1}{q}}$$

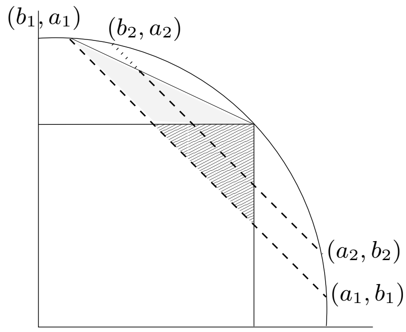

Figure 2: An illustration of the fact, in the case for $p = 2$ and $q = \infty$, that as a and b get closer, $\max\{\beta : \forall|\alpha| \le \beta \ \|T_\alpha f\|_q \le \|f\|_p\}$ decreases. As we move the line segment from $[(a_1, b_1), (b_1, a_1)]$ to $[(a_2, b_2), (b_2, a_2)]$, the portion of the line segment outside the ∞-circle decreases at a slower rate than the portion inside the ∞-circle, due to convexity of the 2-circle.

$$
\begin{aligned}
&= \left[\frac{1}{2}\left(\frac{a+b}{2} + \frac{a-b}{2}\alpha\right)^q + \frac{1}{2}\left(\frac{a+b}{2} - \frac{a-b}{2}\alpha\right)^q\right]^{\frac{1}{q}} \\
&= \frac{a+b}{2}\left[\frac{1}{2}\left(1 + \frac{a-b}{a+b}\alpha\right)^q + \frac{1}{2}\left(1 - \frac{a-b}{a+b}\alpha\right)^q\right]^{\frac{1}{q}} \\
&= \frac{a+b}{2}\left(\frac{1}{2}(1+\delta\,\alpha)^q + \frac{1}{2}(1-\delta\,\alpha)^q\right)^{\frac{1}{q}},
\end{aligned}
$$

and by setting $\alpha = 1$ and replacing q by p,

$$
\|f\|_p = \|T_1 f\|_p = \frac{a+b}{2}\left(\frac{1}{2}(1+\delta)^p + \frac{1}{2}(1-\delta)^p\right)^{\frac{1}{p}}.
$$

Therefore $\|T_\alpha f\|_q \le \|f\|_p$ iff

$$
\left(\frac{1}{2}(1+\delta\,\alpha)^q + \frac{1}{2}(1-\delta\,\alpha)^q\right)^{\frac{1}{q}} \le \left(\frac{1}{2}(1+\delta)^p + \frac{1}{2}(1-\delta)^p\right)^{\frac{1}{p}}. \tag{3}
$$

Now, using Taylor's expansion $(1 \pm \delta\alpha)^q = 1 \pm q\delta\alpha + \frac{q(q-1)}{2}(\delta\alpha)^2 \pm \cdots$, we can rewrite the LHS of (3) as

$$
\begin{aligned}
\text{LHS} \quad &\underset{\delta\to 0}{\sim} \quad \left(1 + \frac{q(q-1)}{2}(\delta\alpha)^2\right)^{\frac{1}{q}} \\
&\underset{\delta\to 0}{\sim} \quad 1 + \frac{q-1}{2}(\delta\alpha)^2,
\end{aligned}
$$

where the second line follows again from Taylor's expansion.

Similarly,

$$\text{RHS} \underset{\delta\to 0}{\sim} 1 + \frac{p-1}{2}\,\delta^2 \,.$$

Putting together,

$$\begin{aligned}\text{LHS} < \text{RHS} &\underset{\delta\to 0}{\iff} \frac{q-1}{2}\,\alpha^2 < \frac{p-1}{2} \\ &\iff |\alpha| < \sqrt{\frac{p-1}{q-1}}\,,\end{aligned}$$

and by continuity

$$\text{LHS} \le \text{RHS} \underset{\delta\to 0}{\iff} |\alpha| \le \sqrt{\frac{p-1}{q-1}}\,,$$

proving the base case.

2.2 Inductive Step

Consider $x \in \{-1,1\}^n$ and a nontrivial partition $x = x_1x_2$ with $|x_1| = k$, $0 < k < n$. Our first step is to obtain an expression for $(T_\alpha f)(x)$ in terms of T_α applied to functions on fewer variables, so we can apply our induction hypothesis.

$$(T_\alpha f)(x_1x_2) = \sum_{S\subseteq[n]} \alpha^{|S|}\hat{f}(S)\chi_S(x_1x_2)$$

partitioning S consistent with the way x is partitioned,

$$\begin{aligned}&= \sum_{\substack{S_1\subseteq[k]\\ S_2\subseteq\{k+1,\ldots,n\}}} \alpha^{|S_1|}\alpha^{|S_2|}\hat{f}(S_1\cup S_2)\chi_{S_1}(x_1)\chi_{S_2}(x_2) \\ &= \sum_{S_1}\alpha^{|S_1|}\underbrace{\Big(\sum_{S_2}\alpha^{|S_2|}\hat{f}(S_1\cup S_2)\chi_{S_2}(x_2)\Big)}_{(+)}\chi_{S_1}(x_1)\end{aligned}$$

defining a function $g_{x_2} : \{-1,1\}^k \to \mathbb{R}$ such that $\widehat{g_{x_2}}(S_1)$ is $(+)$,

$$\begin{aligned}&= \sum_{S_1}\alpha^{|S_1|}\widehat{g_{x_2}}(S_1)\chi_{S_1}(x_1) \\ &= \big(T_\alpha g_{x_2}\big)(x_1)\,, \qquad (4)\end{aligned}$$

where T_α in the last line is an operator for functions on the Boolean k-cube. For future reference we point out that for any fixed x_1, $g_{x_2}(x_1)$ as a function of x_2 can be expressed as the result of applying the noise operator T_α on the restriction of f to the Boolean $(n-k)$-cube defined by x_1. We formalize this observation in the following lemma.

Lemma 1. *Let $g_{x_2} : \{-1,1\}^k \to \mathbb{R}$ be the function such that its Fourier coefficient corresponding to $S_1 \subseteq [k]$ is given by $\widehat{g_{x_2}}(S_1) = \sum_{S_2} \alpha^{|S_2|} \hat{f}(S_1 \cup S_2)\chi_{S_2}(x_2)$. Then*

$$g_{x_2}(x_1) = \big(T_\alpha(f|_R)\big)(x_2),$$

where $R = ([k], x_1)$ restricts f to give $f|_R : \{-1,1\}^{n-k} \to \mathbb{R}$ with $(f|_R)(x_2) = f(x_1x_2)$.

Proof.

$$\begin{aligned}
g_{x_2}(x_1) &= \sum_{S_1 \subseteq [k]} \widehat{g_{x_2}}(S_1)\chi_{S1}(x_1) \\
&= \sum_{S_1} \Big(\sum_{S_2 \subseteq \{k+1,\dots,n\}} \alpha^{|S_2|} \hat{f}(S_1 \cup S_2)\chi_{S_2}(x_2) \Big) \chi_{S_1}(x_1) \\
&= \sum_{S_2} \alpha^{|S_2|} \underbrace{\Big(\sum_{S_1} \hat{f}(S_1 \cup S_2)\chi_{S_1}(x_1) \Big)}_{(\#)} \chi_{S_2}(x_2).
\end{aligned}$$

Recall that as part of the discussion in Lecture 9 on random restrictions, we obtained an expression for $\widehat{f|_R}(S_2)$, where $R = (I, v)$ and $S_2 \subseteq [n] \backslash I$, as $\widehat{f|_R}(S_2) = \sum_{S_1 \subseteq I} \hat{f}(S_1 \cup S_2)\chi_{S_1}(v)$. Setting $I = [k]$ and $v = x_1$ in this expression gives precisely (#), and so

$$\begin{aligned}
g_{x_2}(x_1) &= \sum_{S_2} \alpha^{|S_2|}\, \widehat{f|_R}(S_2)\, \chi_{S_2}(x_2) \\
&= \big(T_\alpha(f|_R)\big)(x_2),
\end{aligned}$$

as claimed. ☐

Now we write $\|T_\alpha f\|_q$ in terms of $\|T_\alpha g_{x_2}\|_q$ so that we can apply the induction hypothesis to the latter:

$$\begin{aligned}
\|T_\alpha f\|_q &= \Big(\mathrm{E}_{x_2}\big[\ \mathrm{E}_{x_1}\big[\, |(T_\alpha f)(x_1x_2)|^q \,\big]\ \big] \Big)^{\frac{1}{q}} && \text{(by the definition of } q\text{-norm)} \\
&= \Big(\mathrm{E}_{x_2}\big[\ \mathrm{E}_{x_1}\big[\, |(T_\alpha g_{x_2})(x_1)|^q \,\big]\ \big] \Big)^{\frac{1}{q}} && \text{(by (4))} \\
&= \Big(\mathrm{E}_{x_2}\big[\ (\|T_\alpha g_{x_2}\|_q)^q\ \big] \Big)^{\frac{1}{q}} && \begin{pmatrix}\text{the inner expectation is just} \\ \text{the } q\text{-th power of a } q\text{-norm}\end{pmatrix} \\
&\le \Big(\mathrm{E}_{x_2}\big[\ (\|g_{x_2}\|_p)^q\ \big] \Big)^{\frac{1}{q}}. && \text{(by the induction hypothesis)}
\end{aligned}$$

Finally, we massage this last expression such that we get a q-norm instead of a p-norm so that by Lemma 1 we can apply the induction hypothesis again:

$$\begin{aligned}
\|T_\alpha f\|_q &\le \Big(\mathrm{E}_{x_2}\big[\ \big(\mathrm{E}_{x_1}\big[\, |g_{x_2}(x_1)|^p \,\big]\big)^{\frac{q}{p}}\ \big] \Big)^{\frac{1}{q}} && \text{(by the definition of } p\text{-norm)} \\
&= \Bigg(\Big(\mathrm{E}_{x_2}\big[\ \big(\underbrace{\mathrm{E}_{x_1}\big[\, |g_{x_2}(x_1)|^p \,\big]}_{(*)}\big)^{\frac{q}{p}}\ \big] \Big)^{\frac{p}{q}} \Bigg)^{\frac{1}{p}}
\end{aligned}$$

viewing $(*)$ as a function of x_2, this entire expression is the p-th root of the $\frac{q}{p}$-norm of $(*)$:

$$= \left(\Big\| \overbrace{\mathrm{E}_{x_1}\left[\, |g_{x_2}(x_1)|^p \,\right]}^{(*)} \Big\|_{\frac{q}{p}} \right)^{\frac{1}{p}}$$

$(*)$ can be viewed as a convex combination of 2^k functions of x_2. By linearity and the triangle inequality, the $\frac{q}{p}$-norm—which is a valid norm, since $\frac{q}{p} \geq 1$—of the convex combination is at most the convex combination of the $\frac{q}{p}$-norms of each term, thus yielding

$$
\begin{aligned}
&\leq \left(\mathrm{E}_{x_1}\left[\Big\| |g_{x_2}(x_1)|^p \Big\|_{\frac{q}{p}} \right] \right)^{\frac{1}{p}} \\
&= \left(\mathrm{E}_{x_1}\left[\left(\mathrm{E}_{x_2}\left[\, |g_{x_2}(x_1)|^{p\cdot\frac{q}{p}} \right] \right)^{\frac{p}{q}} \right] \right)^{\frac{1}{p}} && \text{(by the definition of } \tfrac{q}{p}\text{-norm)} \\
&= \left(\mathrm{E}_{x_1}\left[\|g_{x_2}(x_1)\|_q^p \right] \right)^{\frac{1}{p}} \\
&= \left(\mathrm{E}_{x_1}\left[\|T_\alpha(f|_R)\|_q^p \right] \right)^{\frac{1}{p}} && \text{(by Lemma 1)} \\
&\leq \left(\mathrm{E}_{x_1}\left[\|f|_R\|_p^p \right] \right)^{\frac{1}{p}} && \text{(by the induction hypothesis)} \\
&= \left(\mathrm{E}_{x_1}\left[\mathrm{E}_{x_2}\left[|f|_R(x_2)|^p \right] \right] \right)^{\frac{1}{p}} && \text{(by the definition of } p\text{-norm)} \\
&= \left(\mathrm{E}_{x_1}\left[\mathrm{E}_{x_2}\left[|f(x_1x_2)|^p \right] \right] \right)^{\frac{1}{p}} \\
&= \|f\|_p .
\end{aligned}
$$

□

3 Next Lecture

We will use the instantiation of this hypercontractivity result with $q = 2$ to prove two theorems stated in lecture 1, namely that for every balanced function there is a variable with influence $\Omega\left(\frac{\log n}{n}\right)$, and that for every ϵ any function f is ϵ-close to a $2^{O(I(f)/\epsilon)}$-junta.