The majority of this talk is drawn from Ryan O'Donnell's book *Analysis of Boolean Functions*, primarily chapters 9 and and 10. Some of the later topics, relating this material to the theory of log-Sobolev inequalities, are- drawn from "Hypercontractivity, Logarithmic Sobolev Inequalities, and Applications: A Survey of Surveys" by Leonard Gross, appearing in the volume *Diffusion, Quantum Theory, and Radically Elementary Mathematics*.

# 1 Introduction

If you're like me, you think a lot about the Chernoff bound.

**Theorem 1** (Chernoff bound for $\pm 1$ random variables). *Let $X_1, \ldots, X_n$ be iid uniform $\pm 1$ random variables, and let $a_1, \ldots, a_n \in \mathbb{R}$. Let $S = \sum_{i=1}^n a_i X_i$, and let $\sigma^2 = \sum_{i=1}^n a_i^2$ be the variance of $S$. Then for any $t \geq 0$,*

$$\Pr(|S| \geq t\sigma) \leq 2e^{-2t^2}.$$

Another way of viewing this theorem is as saying that every *linear* function on the hypercube is tightly concentrated about its mean, with exponential tail bounds. In this language, Theorem 1 can be restated as follows.

**Theorem 2** (Chernoff bound restated). *Let $X = (X_1, \ldots, X_n)$ be a uniformly random point in $\{-1, 1\}^n$, and let $f : \{-1, 1\}^n \to \mathbb{R}$ be a linear function with $\mathbb{E}[f(X)] = 0$. Then for any $t \geq 0$,*

$$\Pr(|f(X)| \geq t\|f\|_2) \leq 2e^{-2t^2}. \tag{1}$$

Here, we use the convention that the $L^p$ norm is defined by $\|f\|_p = \mathbb{E}[|f(X)|^p]^{1/p}$, i.e. we take the $L^p$ norm with respect to the uniform measure on $\{-1, 1\}^n$. Also, from here on out, we will always take $X = (X_1, \ldots, X_n)$ to be a uniformly random point in $\{-1, 1\}^n$.

The Chernoff bound is great. However, in many applications, one is interested in *non-linear* functions $f : \{-1, 1\}^n \to \mathbb{R}$. Commonly, we're interested in low-degree polynomials. For instance, if we want to count triangles in a random graph, then we are interested in a degree-3 polynomial of the independent random bits that decide the edges. If we want to count consecutive runs of 10 heads in a sequence of coin flips, then we are interested in a degree-10 polynomial.

Note that for high-degree polynomials, we can't expect strong concentration bounds. A simple example is the function $g$ given by $g(x_1, \ldots, x_n) = (x_1 + 1) \cdots (x_n + 1)$. We have that

$$g(X) = \begin{cases} 0 & \text{with probability } 1 - 2^{-n} \\ 2^n & \text{with probability } 2^{-n} \end{cases}$$

from which we can see that $\|g\|_2 = 2^{n/2}$. Therefore, for any $t < 2^{n/2}$, we have that

$$\Pr(|g(X)| \geq t\|f\|_2) = \Pr(g(X) = 2^n) = 2^{-n},$$

whereas a bound as strong as (1) would allow us to upper-bound this by $\exp(-2t^2) = \exp(-2^n)$ if we take $t = 2^{(n-1)/2}$.

Nonetheless, for low-degree polynomials, there is hope, and there are many techniques for proving strong concentration bounds for low-degree polynomials. The most general result I know is the following.

**Theorem 3** (Tail bounds for low-degree polynomials)**.** *Let $f : \{-1,1\}^n \to \mathbb{R}$ be a polynomial of degree at most $k$. Then for any $t \geq (2e)^{k/2}$, we have that*

$$\Pr(|f(X)| \geq t\|f\|_2) \leq \exp\left(-\frac{k}{2e}t^{2/k}\right).$$

In other words, we are able to get an exponential tail bound, where here "exponential" means exponential in $t^{2/k}$. If we think of the degree $k$ as fixed, then this is quite good: for many applications, it suffices to obtain a bound that is exponentially small in any fixed power of the deviation $t$.

To prove Theorem 3, one needs the theory of hypercontractivity on the Boolean hypercube, which is what most of this talk will be about. At its most basic level, hypercontractivity allows one to prove bounds on norms of random variables.

**Theorem 4.** *Let $f : \{-1,1\}^n \to \mathbb{R}$ be a polynomial of degree at most $k$. Then for any $q \geq 2$,*

$$\|f\|_q \leq \sqrt{q-1}^k \|f\|_2.$$

Recall that on a probability space, the norms are monotonic, i.e. that $\|f\|_p \leq \|f\|_q$ for any $f : \{-1,1\}^n \to \mathbb{R}$ and any $p \leq q$. The point of norm bounds like Theorem 4 is that we can get some inequality in the other direction. This is incredibly powerful; for instance, given Theorem 4, we can easily prove the tail bound in Theorem 3.

*Proof of Theorem 3.* Let $q \geq 2$ be a parameter that we will choose later. By combining Markov's inequality with Theorem 4, we find that

$$\begin{aligned}
\Pr(|f(X)| \geq t\|f\|_2) &= \Pr(|f(X)|^q \geq t^q\|f\|_2^q) \\
&\leq \frac{\mathbb{E}[|f(X)|^q]}{t^q\|f\|_2^q} && \text{[Markov]} \\
&= t^{-q}\left(\frac{\|f\|_q}{\|f\|_2}\right)^q \\
&\leq t^{-q}(q-1)^{kq/2} && \text{[Theorem 4]} \\
&< (q^{k/2}/t)^q.
\end{aligned}$$

It now remains to pick a value of $q$ that yields a good bound. By choosing $q = t^{2/k}/e$, which is at least 2 by assumption, we find that

$$\Pr(|f(X)| \geq t) \leq \exp\left(-\frac{kq}{2}\right) = \exp\left(-\frac{k}{2e}t^{2/k}\right). \qquad \square$$

Theorem 4 allows us to upper-bound $\|f\|_q$ in terms of $\|f\|_2$ for any $q \geq 2$. In many applications, it is also useful to obtain such a bound for $q \leq 2$, and in particular for $q = 1$. Directly applying hypercontractivity will not accomplish this; however, once we have a result like Theorem 4, which gives bounds for *all* $q \geq 2$, a simple interpolation trick using Hölder's inequality (or, if you prefer, the Riesz–Thorin theorem) allows one to obtain upper bounds on $\|f\|_1$ as well.

**Lemma 5.** *Let $f : \{-1,1\}^n \to \mathbb{R}$ be a polynomial of degree at most $k$. Then for any $\|f\|_2 \leq e^k \|f\|_1$.*

*Proof.* The generalized Hölder inequality and Theorem 4 for $p = 2$ imply that for any $\varepsilon > 0$,

$$\|f\|_2 \leq \|f\|_{2+\varepsilon}^{1-\theta} \|f\|_1^{\theta} \leq \sqrt{1+\varepsilon}^{k(1-\theta)} \|f\|_2^{1-\theta} \|f\|_1^{\theta}$$

where $\theta$ is the solution to $\frac{1}{2} = \frac{\theta}{1} + \frac{1-\theta}{2+\varepsilon}$. We now divide by $\|f\|_2^{1-\theta}$ and let $\varepsilon \to 0$ to conclude the desired result. □

Using the same interpolation trick, one can similarly show that $\|f\|_2 \leq (e^{\frac{2}{p}-1})^k \|f\|_p$ for any $1 \leq p \leq 2$. Combining this result with Theorem 4, we find that for any $p \in [1, \infty)$,

$$c_{p,k} \|f\|_p \leq \|f\|_2 \leq C_{p,k} \|f\|_p$$

where the constants $c_{p,k}, C_{p,k}$ depend only on $p$, and the degree $k$. In the case $k = 1$, this is an important probabilistic result known as Khintchine's inequality. Morally, what all of this is saying is that low-degree polynomials are "well-behaved": all their norms are equal up to a constant factor, and therefore they inherit many properties of "nice" random variables.

One further consequence of this control on norms is *anticoncentration*; basically, one can say that with some non-trivial probability, a low-degree polynomial will be far from its mean. Here are two such results that follow easily from these norm bounds; unsurprisingly, one can deduce many other such results using similar techniques.

**Theorem 6.** *Let $f : \{-1,1\}^n \to \mathbb{R}$ be a polynomial of degree at most $k$, and suppose $f$ is non-constant. Then*

$$\Pr\left(|f(X) - \mathbb{E}[X]| \geq \frac{1}{2}\|f\|_2\right) \geq \frac{1}{16} 9^{1-k}$$

*and*

$$\Pr\left(f(X) > \mathbb{E}[X]\right) \geq \frac{1}{4} e^{-2k}.$$

*Proof.* We may assume without loss of generality that $\mathbb{E}[f(X)] = 0$. For the first result, we can apply the Paley–Zygmund inequality to find that

$$\Pr\left(|f(X)| \geq \frac{1}{2}\|f\|_2\right) = \Pr\left(f(X)^2 \geq \frac{1}{4}\|f\|_2^2\right) \geq \frac{9}{16} \frac{\|f\|_2^4}{\|f\|_4^4} \geq \frac{9}{16}(3^{-k/2})^4 = \frac{1}{16} 9^{1-k},$$

where we apply Theorem 4 with $q = 4$ for the last inequality.

For the second result, we use the fact that $\mathbb{E}[f(X)] = 0$ to find that $\mathbb{E}[f(X)\mathbf{1}_{\{f(X)>0\}}] = \frac{1}{2}\|f\|_1$. Applying Cauchy–Schwarz, we find that

$$\frac{1}{4}\|f\|_1^2 \leq \mathbb{E}[f(X)^2]\,\mathbb{E}[\mathbf{1}_{\{f(X)>0\}}^2] = \|f\|_2^2 \Pr(f(X) > 0) \leq e^{2k}\|f\|_1^2 \Pr(f(X) > 0),$$

using Lemma 5 for the final inequality. Rearranging gives the desired result. □

# 2 Hypercontractivity

To state the hypercontractive inequality, we need to set up some terminology and notation. Given $x = (x_1, \ldots, x_n) \in \{-1, 1\}^n$, and given $\rho \in [0, 1]$, we define a probability distribution $N_\rho(x)$ as follows. For each $i \in [n]$, we let

$$Y_i = \begin{cases} x_i & \text{with probability } \rho \\ \pm 1 \text{ uniformly at random} & \text{with probability } 1 - \rho \end{cases}$$

and then let $N_\rho(x)$ be the distribution of the random vector $Y = (Y_1, \ldots, Y_n)$, where the coordinates are independent. This distribution is called the "$\rho$-noisy neighborhood" of $x$. Another way of viewing it is as follows. Consider a continuous-time Markov chain on $\{-1, 1\}^n$, where we have an exponential clock with rate 1 on each $i \in [n]$, and every time the $i$th clock rings we resample uniformly at random the coordinate $i$. If we start this Markov chain at $x$ and run it for time $\log \frac{1}{\rho}$, then its output state will be distributed as $N_\rho(x)$. Note that $N_0(x)$ is uniformly random on $\{-1, 1\}^n$, and that $N_1(x)$ takes on value $x$ with probability 1.

If $X \in \{-1, 1\}^n$ is uniformly random and $Y \sim N_\rho(X)$, then we say that the pair $(X, Y)$ is *$\rho$-correlated*. Note that this is a symmetric condition, i.e. that we would get the same distribution on the pair if we first sampled $Y$ uniformly and then sampled $X \sim N_\rho(Y)$. They are called $\rho$-correlated because $X_i = Y_i$ with probability $\rho$, and $X_i$ and $Y_i$ are independent with probability $1 - \rho$.

With these definitions, we can define our main object of study, the noise operator.

**Definition 7.** Given $\rho \in [0, 1]$, we define the *noise operator* $T_\rho$ by

$$(T_\rho f)(x) = \mathop{\mathbb{E}}_{Y \sim N_\rho(x)}[f(Y)].$$

In other words, we compute $(T_\rho f)(x)$ by averaging the values of $f$ on the $\rho$-noisy neighborhood of $x$.

Up to some simple transformations, $T_\rho$ is just the Laplacian associated to the continuous-time Markov chain discussed above. By virtue of being an averaging operator (or by a simple application of Jensen's inequality), we can see that $\|T_\rho f\|_p \leq \|f\|_p$ for any $\rho$ and $p$. In other words, $T_\rho$ is a *contraction* in $L^p$ for any $p$. The hypercontractive inequality says that in fact, $T_\rho$ is *hypercontractive*, namely is a contraction $L^p \to L^q$ for some $q > p$.

**Theorem 8** (Bonami)**.** *Let $1 \le p \le q \le \infty$, and let $0 \le \rho \le \sqrt{\frac{p-1}{q-1}}$. Then $T_\rho$ is a contraction when viewed as a map $L^p \to L^q$, i.e.*

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

*for any $f : \{-1,1\}^n \to \mathbb{R}$.*

Although the hypercontractive inequality is rightly attributed to Bonami, it's worth noting that the history is complicated: Bonami proved Theorem 8 in her PhD thesis, which was published in French and therefore generally unknown in the English-speaking world for some time. Because of this, variants of Theorem 8, as well as Theorem 8 itself, were discovered independently by Nelson, Gross, and Beckner.

Intuitively, the reason why $T_\rho$ should satisfy some sort of hypercontractivity is the following. The reason why $\|f\|_q$ may be (much) larger than $\|f\|_p$ for $q > p$ is that $f$ may have some sharp peaks, which will receive a higher weight in the computation of the larger moment $\|f\|_q$. However, $T_\rho$ is a "smoothing" operator: since $T_\rho f$ is gotten by averaging $f$ over $\rho$-noisy neighborhoods, any sharp peak of $f$ will be somewhat dampened by $T_\rho$. The smaller $\rho$ is, the greater this effect will be (since the neighborhood $N_\rho(x)$ will be less localized). Hence, we might expect to get some control on $\|T_\rho f\|_q$ in terms of $\|f\|_p$, so long as $\rho$ is sufficiently small relative to $p$ and $q$.

To understand the connections Theorem 8 has with the results discussed in the previous section, we need to take a quick detour to understand Fourier analysis on the Boolean hypercube.

## 3 Fourier analysis

For a set $S \subseteq [n]$, we define the function $\chi_S : \{-1,1\}^n \to \mathbb{R}$ by

$$\chi_S(x_1, \dots, x_n) = \prod_{i \in S} x_i,$$

with the convention that the empty product is taken to be 1. It is easy to check that the functions $\chi_S$ are the characters of the abelian group $\{-1,1\}^n$, which implies that we can use them to define the Fourier transform. Namely, for any $f : \{-1,1\}^n \to \mathbb{R}$, we define

$$\widehat{f}(S) = \mathbb{E}[f(X)\chi_S(X)]$$

and then the Fourier inversion formula says that

$$f(x) = \sum_{S \subseteq [n]} \widehat{f}(S)\chi_S(x) \tag{2}$$

for all $x \in \{-1,1\}^n$. The numbers $\widehat{f}(S)$ are called the *Fourier coefficients* of $f$. The only other property of the Fourier transform we'll need is *Parseval's identity*, which says that

$$\|f\|_2^2 = \sum_{S \subseteq [n]} |\widehat{f}(S)|^2.$$

Note that $\chi_S$ is a multilinear polynomial of degree $|S|$. Therefore, one consequence of (2) is that every function $\{-1,1\}^n \to \mathbb{R}$ has a *unique* representation as a multilinear polynomial of degree at most $n$. This also implies that if $f$ is a polynomial of degree at most $k$, then $\hat{f}(S) = 0$ for all $S$ with $|S| > k$. This is one of the many reasons why the Fourier decomposition is a powerful tool for understanding functions on the hypercube.

To connect this to hypercontractivity, we will need a description of the operator $T_\rho$ in the Fourier basis. That is given by the following lemma.

**Lemma 9.** *For any $S \subseteq [n]$, any $\rho \in [0,1]$, and any $f : \{-1,1\}^n \to \mathbb{R}$, we have that*

$$(\widehat{T_\rho f})(S) = \rho^{|S|}\widehat{f}(S).$$

*Therefore,*

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

*Proof.* We begin by computing $T_\rho \chi_S$. Recalling the definition of $T_\rho$, we find that

$$(T_\rho \chi_S)(x) = \mathop{\mathbb{E}}_{Y \sim N_\rho(x)}[\chi_S(Y)] = \mathop{\mathbb{E}}_{Y \sim N_\rho(x)}\left[\prod_{i \in S} Y_i\right] = \prod_{i \in S} \mathop{\mathbb{E}}_{Y_i \in N_\rho(x_i)}[Y_i],$$

where the final equality uses the fact that if $Y$ is drawn from $N_\rho(x)$, then all the coordinates of $Y$ are independent, with the $i$th one drawn from the distribution $N_\rho(x_i)$, which equals $x_i$ with probability $\rho$ and is otherwise uniformly random on $\pm 1$. From this, we can compute that

$$\mathop{\mathbb{E}}_{Y_i \sim N_\rho(x_i)}[Y_i] = \rho \cdot x_i + (1-\rho) \cdot 0 = \rho x_i.$$

Combining this with our earlier computation, we get that

$$(T_\rho \chi_S)(x) = \prod_{i \in S} \mathop{\mathbb{E}}_{Y_i \in N_\rho(x_i)}[Y_i] = \prod_{i \in S}(\rho x_i) = \rho^{|S|}\chi_S(x).$$

In other words, $\chi_S$ is an eigenfunction of $T_\rho$ with eigenvalue $\rho^{|S|}$. This already implies the desired result by some general theory (since we can conclude from this that $T_\rho$ is self-adjoint), but for completeness, here's the full computation.

We have that

$$\begin{aligned}
(\widehat{T_\rho f})(S) &= \mathop{\mathbb{E}}_X[(T_\rho f)(X)\chi_S(X)] \\
&= \mathop{\mathbb{E}}_X\left[\mathop{\mathbb{E}}_{Y \sim N_\rho(X)}[f(Y)]\chi_S(X)\right] \\
&= \mathop{\mathbb{E}}_{(X,Y)\ \rho\text{-correlated}}[f(Y)\chi_S(X)] \\
&= \mathop{\mathbb{E}}_Y\left[\mathop{\mathbb{E}}_{X \sim N_\rho(Y)}[\chi_S(X)]f(Y)\right] \\
&= \mathop{\mathbb{E}}_Y[\rho^{|S|}\chi_S(Y)f(Y)] \\
&= \rho^{|S|}\widehat{f}(S)
\end{aligned}$$

where we use the property, mentioned above, that $(X, Y)$ being $\rho$-correlated is symmetric. $\square$

From the Fourier expansion of $T_\rho$, we can conclude a number of important properties of $T_\rho$, many of which are non-obvious from the original definition (though some can be proved directly from that definition).

1. $T_\rho$ is self-adjoint.

2. We may extend the definition of $T_\rho$ to $\rho > 1$ via the formula

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

3. With this definition, the operators $\{T_\rho\}_{\rho > 0}$ form a group, with the group operation given by $T_\rho \circ T_\sigma = T_{\rho\sigma}$.

4. Earlier, we gave some intuition for why $T_\rho$ should be a "smoothing operator" for small $\rho$. In the Fourier basis, we see that $T_\rho$ is damping the higher Fourier coefficients of $f$, at least for $\rho < 1$. Since the higher modes should contribute to more local fluctuations, this is consistent with the earlier intuition. Note that for $\rho > 1$, $T_\rho$ is actually accentuating the higher modes, which of course makes some sense: by the group property, we have that $T_{1/\rho} = T_\rho^{-1}$, so $T_{1/\rho}$ must "undo" whatever damping is done by $T_\rho$.

With what we know so far, it is easy to show how the hypercontractive inequality implies the norm bounds Theorem 4, which were what we used to deduce all of our other applications of hypercontractivity.

*Proof of Theorem 4.* Let $\rho = \sqrt{\frac{1}{q-1}} \in (0, 1]$, so that $T_\rho$ is a contraction from $L^2$ to $L^q$ by Theorem 8. By the group property mentioned above, we may write

$$\|f\|_q = \|T_\rho(T_{1/\rho} f)\|_q \leq \|T_{1/\rho} f\|_2.$$

Now, recall that since $f$ has degree at most $k$, we have $\widehat{f}(S) = 0$ whenever $|S| > k$. Therefore, by Parseval's identity,

$$\|T_{1/\rho} f\|_2^2 = \sum_{S \subseteq [n]} |(1/\rho)^{|S|} \widehat{f}(S)|^2 = \sum_{\substack{S \subseteq [n] \\ |S| \leq k}} \rho^{-2|S|} |\widehat{f}(S)|^2 \leq \rho^{-2k} \sum_{S \subseteq [n]} |\widehat{f}(S)|^2 = \rho^{-2k} \|f\|_2^2.$$

Combining this with the earlier computation, we conclude that

$$\|f\|_q \leq \|T_{1/\rho} f\|_2 \leq \rho^{-k} \|f\|_2 = \sqrt{q-1}^k \|f\|_2. \qquad \square$$

# 4 The proof of hypercontractivity

I won't show the proof of Theorem 8. The basic idea is that one wants to induct on the dimension $n$, which is a very appealing thing to do since the statement of the hypercontractive inequality actually has no mention of $n$ whatsoever. The inductive step of the proof is not so bad: basically, one splits the vector $(X_1, \ldots, X_n)$ as the concatenation of $(X_1, \ldots, X_{n-1})$ and $X_n$, and breaks the expectations definining $\|T_\rho f\|_q$ and $\|f\|_p$ into two nested expectations, one depending on $(X_1, \ldots, X_{n-1})$ and one depending on $X_n$. At this point, it basically suffices to apply the inductive hypothesis and the base case of $n = 1$ to conclude the desired result. However, it turns out that really the most convenient way of making this work is to prove a stronger "two-function" version of hypercontractivity: by doing so, one ensures a slightly stronger induction hypothesis that makes it somewhat easier to carry out the proof.

However, it turns out that the base case is really not so trivial to prove. The base case of the hypercontractive inequality is called a "two-point inequality", since it concerns functions $f : \{-1, 1\} \to \mathbb{R}$, namely pairs of points in $\mathbb{R}$. It might be surprising that dealing with two real numbers can be so challenging, but it is! To get an idea of why it's so challenging, consider that the following is the general form of the two-point inequality.

**Lemma 10** (Two-point inequality)**.** *For any $a, b \in \mathbb{R}$, any $1 \le p \le q < \infty$, and any $0 \le \rho \le \sqrt{\frac{p-1}{q-1}}$, we have that $\|a + \rho b X\|_q \le \|a + bX\|_p$, i.e.*

$$\left(\frac{1}{2}(a+\rho b)^q + \frac{1}{2}(a-\rho b)^q\right)^{1/q} \le \left(\frac{1}{2}(a+b)^p + \frac{1}{2}(a-b)^p\right)^{1/p}.$$

Just to get a picture of how one could prove such a thing, let's do it in the case $p = 2, q = 4$. The left-hand side is

$$\left(\frac{1}{2}(a+\rho b)^4 + \frac{1}{2}(a-\rho b)^4\right)^{1/4} = (a^4 + 6\rho^2 a^2 b^2 + \rho^4 b^4)^{1/4},$$

since all the odd terms cancel out when we apply the binomial theorem to the inner terms. Similarly, the right-hand side is

$$\left(\frac{1}{2}(a+b)^2 + \frac{1}{2}(a-b)^2\right)^{1/2} = (a^2 + b^2)^{1/2}.$$

Raising both sides to the fourth power, we wish to prove

$$a^4 + 6\rho^2 a^2 b^2 + \rho^4 b^4 \overset{?}{\le} (a^2 + b^2)^2 = a^4 + 2a^2b^2 + 4b^4.$$

Comparing terms, we see that we'll be all set if $6\rho^2 \le 2$ and $\rho^4 \le 4$. The former condition is equivalent to $\rho \le 1/\sqrt{3}$, while the latter is equivalent to $\rho \le 1/\sqrt{2}$, which is a less stringent requirement. So we find that the two-point inequality holds for $p = 2$ and $q = 4$ if $\rho \le 1/\sqrt{3}$, which is precisely the bound given in the statement of Theorem 8.

The exact same proof will work for $p = 2$ and $q$ any even integer, in basically the same way. For other values of $p$ and $q$, one ends up using the generalized binomial theorem (i.e. writing down a Taylor series), except that things get pretty hairy because of (a) convergence issues and (b) the fact that some coefficients are negative. A fair amount of trickery is required to overcome these issues; everything involved is "elementary", but I would say that a lot of it is far from easy.

## 5 Hypercontractivity and log-Sobolev inequalities

In this section, I'll move away from hypercontractivity and briefly discuss the relations this theory has to various related topics in functional analysis. Since this is material I understand much less well, I'll be somewhat vague at times.

Suppose that $\mu$ is a probability measure on $\mathbb{R}^n$, and suppose that $\mu$ has a smooth, strictly positive density function $h(x)$ with respect to Lebesgue measure. The *Dirichlet form operator* associated to $\mu$ is the operator $A_\mu$ on $L^2$ defined by

$$(A_\mu f)(x) = -\Delta f(x) - \frac{1}{h(x)} \left(\nabla h(x) \cdot \nabla f(x)\right),$$

at least for $f$ for which this expression makes sense (e.g. for $f \in C_c^\infty(\mathbb{R}^n)$). The importance of this operator is its following property, which one can easily prove by integration by parts. For any $f, g \in L^2(\mathbb{R}^n, \mu)$,

$$\langle A_\mu f, g\rangle_{L^2(\mathbb{R}^n,\mu)} = \int_{\mathbb{R}^n} \nabla f(x) \cdot \nabla \overline{g(x)} \, \mathrm{d}\mu(x).$$

Probably the most important example is in case $\mu = \gamma$ is the Gaussian measure on $\mathbb{R}^n$, in which case one has

$$(A_\gamma f)(x) = -\Delta f(x) + x \cdot \nabla f(x).$$

At least in infinite dimensions, this $A_\gamma$ is called the Ornstein–Uhlenbeck operator; I am not sure if it has a name in finite dimensions.

It turns out that Dirichlet form operators are actually a pretty general class of elliptic operators. There's a technique known as Jacobi's trick which allows one to convert many other elliptic differential operators, such as the Schrödinger operator, into

Note that just like $T_\rho$ was "kind of like" the Laplacian on the hypercube, we have that $A_\mu$ is "kind of like" the Laplacian on $\mathbb{R}^n$. Continuing this analogy, we may want to come up with a family $\{D_\rho^\mu\}_{\rho>0}$, built on $A_\mu$, which is forms a group via the multiplication rule $D_\rho^\mu \circ D_\sigma^\mu = D_{\rho\sigma}^\mu$. An obvious way to do this is to define

$$D_\rho^\mu := e^{(\log \rho) A_\mu}.$$

With this definition, we have the following theorem.

**Theorem 11** (Nelson). *Let $\gamma$ be Gaussian measure on $\mathbb{R}^n$, and let $D_\rho^\gamma$ be defined as above. For $1 < p \le q < \infty$, the operator norm of $D_\rho^\gamma$ satisfies*

$$\|D_\rho^\gamma\|_{L^p(\mathbb{R}^n,\gamma)\to L^q(\mathbb{R}^n,\gamma)} = \begin{cases} 1 & \text{if } \rho \le \sqrt{\frac{p-1}{q-1}} \\ \infty & \text{otherwise.} \end{cases}$$

In other words, the operator $D_\rho^\gamma$ also satisfies a hypercontractivity theorem, with the same threshold for hypercontractivity. Though Nelson's proof was analytic in nature, it is also possible to prove Nelson's theorem as a consequence of Bonami's Theorem 8. Indeed, the central limit theorem allows one to convert results about iid uniform $\pm 1$ random variables into results about Gaussian random variables, and this conversion allows one to deduce Theorem 11 from Theorem 8.

As a consequence of Nelson's hypercontractivity theorem, Gross was able to prove the first log-Sobolev inequality, which is as follows. Recall that $\gamma$ denotes Gaussian measure on $\mathbb{R}^n$.

**Theorem 12.** *For any $f : \mathbb{R}^n \to \mathbb{C}$,*

$$\int_{\mathbb{R}^n} |f(x)|^2 \log\left(\frac{|f(x)|}{\|f\|_{L^2(\mathbb{R}^n,\gamma)}}\right) \mathrm{d}\gamma(x) \le \int_{\mathbb{R}^n} |\nabla f(x)|^2 \, \mathrm{d}\gamma(x).$$

Like the Sobolev inequalities, Gross's log-Sobolev inequality involves integrals of both $f$ and its derivatives. However, unlike the Sobolev inequalities, the log-Sobolev inequality is independent of the dimension $n$, which makes it much more useful in certain settings.

In addition to deducing the log-Sobolev inequality from Nelson's hypercontractive inequality, Gross actually proved the following much more general result, showing that log-Sobolev inequalities are *equivalent* to hypercontractive inequalities.

**Theorem 13** (Gross). *Let $\mu$ be a measure on $\mathbb{R}^n$, let $A_\mu$ be the associated Dirichlet form operator, and let $D_\rho^\mu = e^{(\log \rho)A_\mu}$. Then $\mu$ satisfies the log-Sobolev inequality*

$$\int_{\mathbb{R}^n} |f(x)|^2 \log\left(\frac{|f(x)|}{\|f\|_{L^2(\mathbb{R}^n,\mu)}}\right) \mathrm{d}\mu(x) \le \int_{\mathbb{R}^n} |\nabla f(x)|^2 \, \mathrm{d}\mu(x) \qquad \text{for any } f : \mathbb{R}^n \to \mathbb{C}$$

*if and only if it satisfies the hypercontractive inequality*

$$\|D_\rho^\mu\|_{L^p(\mu)\to L^q(\mu)} = 1 \qquad \text{for any } 1 < p \le q < \infty \text{ and any } \rho \le \sqrt{\frac{p-1}{q-1}}.$$

Since there are also direct proofs of the log-Sobolev inequality, this gives another way of proving Nelson's hypercontractivity theorem. Additionally, by directly proving log-Sobolev inequalities for the uniform measure on the hypercube, one can also use this to give alternative proofs of Bonami's Theorem 8 (though strictly speaking, Theorem 13 only applies to measures on $\mathbb{R}^n$, so one needs to prove an analogue for this discrete setting).

The proof of Theorem 13 is actually pretty easy, though I won't show it. Essentially, one takes the $q$-derivative of the hypercontractive inequality (with $p = 2$), and one obtains the log-Sobolev inequality; integrating the log-Sobolev inequality with respect to $q$ gives the hypercontractive inequality.