1 Hypercontractive inequality

Let $f: \{-1, 1\}^n \to \mathbb{R}$. Its $\ell_p$ norm is $\|f\|_p = \mathbb{E}[|f(x)|^p]^{1/p}$. By convexity $\|f\|_p \leq \|f\|_q$ if $p \leq q$. The hypercontractivity theorem gives the reverse inequality when we add some noise. Recall that $T_\rho f: \{-1, 1\}^n \to \mathbb{R}$ is defined as

$$T_\rho f(x) = \mathbb{E}[f(y)]$$

where $(x, y)$ are $\rho$-correlated, namely for each $i$ independently $\mathbb{E}[x_i] = \mathbb{E}[y_i] = 0$ and $\mathbb{E}[x_i y_i] = \rho$. Again by convexity, if $\rho < \rho'$ then $\|T_\rho f\|_p \leq \|T_{\rho'} f\|_p$. Recall that the Fourier expansion of $T_\rho f$ is given by

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

So: increasing $p$ increases the norm; decreasing $\rho$ decreases the norm. The hypercontractive inequality controls the relation between the two.

**Theorem 1.1** (Hypercontractive theorem). Let $f: \{-1, 1\}^n \to \mathbb{R}$. Then for any $1 \leq p \leq q \leq \infty$ and any $0 \leq \rho \leq \sqrt{\frac{p-1}{q-1}}$ it holds that

$$\|T_\rho f\|_q \leq \|f\|_p.$$

We start by giving a number of applications for the hypercontractive inequality, before proving a (special case) of it.

2 Low degree polynomials are “nice”

Let $f: \{-1, 1\}^n \to \mathbb{R}$ be a low degree polynomial. We can use the hypercontractive theorem to show that its norms are bounded. For concreteness, we focus on the $\ell_4$ vs $\ell_2$ norm. Assume that $f: \{-1, 1\}^n \to \mathbb{R}$ is a homogeneous degree $k$ polynomial (that is, all the monomials have
degree exactly $k$). Note that for such functions, $T_{\rho}f = \rho^k f$ as can be seen by considering the Fourier expansion of $f$. Thus, by the hypercontractive theorem we have

$$3^{-k/2}\|f\|_4 = \|T_{1/\sqrt{d}}f\|_4 \leq \|f\|_2.$$  

Bonami’s lemma shows that this holds for any polynomial of degree $\leq k$, not just homogeneous ones.

**Lemma 2.1** (Bonami’s lemma). Let $f : \{-1, 1\}^n \to \mathbb{R}$ be a degree $d$ polynomial. Then

$$\mathbb{E}[f^4] \leq 9^d \mathbb{E}[f^2]^2.$$

**Proof.** The proof is by induction on $n$. Let $f(x) = x_n g(x) + h(x)$ where $g, h$ are polynomials in $n-1$ variables, deg($g$) $\leq d-1$ and deg($h$) $\leq d$. We have $\mathbb{E}[f^2] = \mathbb{E}[g^2] + \mathbb{E}[h^2]$. Furthermore we compute

$$\mathbb{E}[f^4] = \mathbb{E}[(x_n g + h)^4] = \mathbb{E}[g^4] + 6\mathbb{E}[g^2 h^2] + \mathbb{E}[h^4].$$

We used the fact that $g, h$ do not depend on $x_n$ and that $\mathbb{E}[x_n] = \mathbb{E}[x_n^3] = 0$ and $\mathbb{E}[x_n^2] = \mathbb{E}[x_n^4] = 1$. We use the induction hypothesis to bound $\mathbb{E}[g^4] \leq 9^{d-1}\mathbb{E}[g^2]^2$ and $\mathbb{E}[h^4] \leq 9^d \mathbb{E}[h^2]^2$. By Cauchy-Schwartz we have

$$\mathbb{E}[g^2 h^2] \leq \sqrt{\mathbb{E}[g^4] \mathbb{E}[h^4]} \leq 9^{d-1/2}\mathbb{E}[g^2] \mathbb{E}[h^2].$$

Thus

$$\mathbb{E}[f^4] \leq 9^{d-1}\mathbb{E}[g^2]^2 + 6 \cdot 9^{d-1/2}\mathbb{E}[g^2] \mathbb{E}[h^2] + 9^d \mathbb{E}[h^2]^2 \leq 9^d (\mathbb{E}[g^2] + \mathbb{E}[h^2])^2 = 9^d \mathbb{E}[f^2].$$

$\square$

**Lemma 2.2** (General polynomial tails). Let $f : \{-1, 1\}^n \to \mathbb{R}$ be a degree $d$ polynomial. Then for any $q \geq 2$,

$$\|f\|_q \leq (q - 1)^{d/2}\|f\|_2.$$

**Proof.** Let $\rho = \sqrt{q - 1}$ and define $g = T^{-1}_\rho f$. Its Fourier expansion is

$$g(x) = \sum_S (1/\rho)^{|S|} \hat{f}(S) x^S.$$  

We have

$$\|f\|_q^2 = \|T_{\rho}g\|_q^2 \leq \|g\|_2^2 = \sum_{k=0}^d (1/\rho)^k W_k(f) \leq (1/\rho)^d \|f\|_2^2.$$  

$\square$

**Corollary 2.3.** Let $f : \{-1, 1\}^n \to \mathbb{R}$ be a degree $d$ polynomial. Assume that $\mathbb{E}[f] = 0$ and $\mathbb{E}[f^2] = 1$. Then for any $\lambda \geq 2$ we have

$$\text{Pr}[|f(x)| \geq \lambda] \leq 2^{-(\lambda/2)^{2/d}}.$$  

2
Proof. Pick $q \geq 2$ to be determined later. We have

\[
\|f\|_q^q = \mathbb{E}[|f(x)|^q] \geq \Pr[|f(x)| \geq \lambda] \lambda^q.
\]

On the other hand, as we just saw that $\|f\|_q \leq (q - 1)^{d/2}$. So

\[
\Pr[|f(x)| \geq \lambda] \leq \left(\frac{(q - 1)^{d/2}}{\lambda}\right)^q.
\]

The lemma follows by setting $q = (\lambda/2)^{2/d} + 1$. \hfill \qed

For another application, we show that low degree polynomials cannot be too concentrated around their mean. We will need the Paley-Zygmund inequality.

**Lemma 2.4** (Paley-Zygmund). Let $X \geq 0$ be a random variable with finite second moment. Then

\[
\Pr[X > \theta \mathbb{E}[X]] \geq (1 - \theta)^2 \frac{\mathbb{E}[X]^2}{\mathbb{E}[X^2]}.
\]

**Proof.** We have

\[
\mathbb{E}[X] = \mathbb{E}[X 1_{X \leq \theta \mathbb{E}[X]}] + \mathbb{E}[X 1_{X > \theta \mathbb{E}[X]}] \leq \theta \mathbb{E}[X] + \sqrt{\mathbb{E}[X^2]} \Pr[X > \theta \mathbb{E}[X]],
\]

where the inequality follows by Cauchy-Schwartz. The lemma follows by rearranging terms. \hfill \qed

**Lemma 2.5** (Polynomial anti-concentration). Let $f : \{-1, 1\}^n \to \mathbb{R}$ be a degree $d$ polynomial with $\mu = \mathbb{E}[f]$ and $\sigma^2 = \text{Var}(f)$. Then

\[
\Pr \left[ |f(x)| \geq \frac{\sigma}{2} \right] \geq \frac{1}{16 \cdot 9^{d-1}}.
\]

**Proof.** We may assume without loss of generality that $\mu = 0$ and $\sigma = 1$. Thus $\mathbb{E}[f^2] = 1$ and by Bonami’s lemma, $\mathbb{E}[f^4] \leq 9^d$. By the Paley-Zygmund inequality applied to the random variable $X = f(x)^2$ we have

\[
\Pr[|f(x)| \geq 1/2] = \Pr[X \geq 1/4] \geq (3/4)^2 \frac{\mathbb{E}[X]^2}{\mathbb{E}[X^2]} \geq \frac{(3/4)^2}{9^d}.
\]

\hfill \qed

### 3 (2,4) hypercontractive inequality

Let $f : \{-1, 1\}^n \to \mathbb{R}$. We prove the special case of the hypercontractive inequality with $p = 2, q = 4, \rho = 1/\sqrt{3}$. Note that if $f$ is a homogeneous polynomial of degree $d$ then by Bonami’s lemma,

\[
\|T_{1/\sqrt{3}}f\|_4 = 3^{-d/2}\|f\|_4 \leq 3^{-d/2} \cdot 3^{d/2}\|f\|_2 = \|f\|_2.
\]

We will prove this for any function.
\textbf{Theorem 3.1} \((2,4)\) hypercontractive inequality. Let \(f: \{-1, 1\}^n \to \mathbb{R}\). Then
\[ \|T_{1/\sqrt{3}} f\|_4 \leq \|f\|_2. \]

\textit{Proof.} Let us shorthand \(T = T_{1/\sqrt{3}}\). We will prove \(\mathbb{E}[(Tf)^4] \leq \mathbb{E}[f^2]^2\) in a similar way to the proof of Bonami’s lemma. Let \(f(x) = x_n g(x) + h(x)\) so that \(\mathbb{E}[f^2] = \mathbb{E}[g^2] + \mathbb{E}[h^2]\). Moreover, \(Tf = \frac{1}{\sqrt{3}} x_n Tg + Th\). Thus
\[ \mathbb{E}[(Tf)^4] = \frac{1}{9} \mathbb{E}[(Tg)^4] + \frac{1}{3} \mathbb{E}(Tg)^2(Th)^2 + \mathbb{E}[(Th)^4]. \]

Again by Cauchy-Schwartz we have \(\mathbb{E}(Tg)^2(Th)^2 \leq \sqrt{\mathbb{E}[(Tg)^4] \mathbb{E}[(Th)^4]}\) and apply induction. We get
\[ \mathbb{E}[(Tf)^4] = \frac{1}{9} \mathbb{E}[g^2]^2 + 2 \mathbb{E}[g^2] \mathbb{E}[h^2] + \mathbb{E}[h^2]^2 \leq (\mathbb{E}[g^2] + \mathbb{E}[h^2])^2 = \mathbb{E}[f^2]^2. \]
\[ \square \]

We get as a corollary a \((4/3, 2)\) hypercontractive inequality.

\textbf{Theorem 3.2} \((4/3,2)\) hypercontractive inequality. Let \(f: \{-1, 1\}^n \to \mathbb{R}\). Then
\[ \|T_{1/\sqrt{3}} f\|_2 \leq \|f\|_{4/3}. \]

\textit{Proof.} We use Hölder’s inequality. Writing \(T = T_{1/\sqrt{3}}\) for brevity gives
\[ \|Tf\|_2^2 = (Tf, Tf) = (f, TTf) \leq \|f\|_{4/3} \|TTf\|_4 \leq \|f\|_{4/3} \|Tf\|_2. \]
\[ \square \]

The nice fact about this is that \(\|T_{1/\sqrt{3}} f\|_2\) has a nice form in terms of the noise stability of \(f\):
\[ \|T_{1/\sqrt{3}} f\|_2^2 = \sum_{k \geq 2} 3^{-k} W_k(f) = \text{Stab}_{1/3}(f). \]

As a corollary, we get that small subsets of \(\{0, 1\}^n\) cannot be too noise stable.

\textbf{Lemma 3.3.} Let \(A \subset \{0, 1\}^n\) of size \(|A| = \alpha 2^n\). Let \((x, y)\) be a \((1/3)\)-correlated pair. Then
\[ \Pr[x, y \in A] \leq \alpha^{3/2}. \]

\textit{Proof.} Let \(f = 1_A\). Then \(\|f\|_{4/3} = \alpha^{3/4}\). We have
\[ \Pr[x, y \in A] = \mathbb{E}[f(x)f(y)] = \sum_{S, T} \hat{f}(S)\hat{f}(T)\mathbb{E}[x^S y^T]. \]

Recall that we showed that \(\mathbb{E}[x^S y^T] = 0\) if \(S \neq T\) and \(\mathbb{E}[x^S y^T] = (1/3)^{|S|}\). Thus
\[ \Pr[x, y \in A] = \mathbb{E}[f(x)f(y)] = \sum_S 3^{-|S|} \hat{f}(S)^2 = \text{Stab}_{1/3}(f) = \|T_{1/\sqrt{3}} f\|_2^2 \leq \|f\|_{4/3}^2 \leq \alpha^{3/2}. \]
\[ \square \]
Note that an equivalent form is that
\[ \Pr[y \in A|x \in A] \leq \alpha^{1/2}. \]
This can be phrased as: pick \( x \in A \) uniformly, and form \( y \) by flipping each bit of \( x \) independently with probability 1/3. Then the probability that also \( y \in A \) is at most \( \alpha^{1/2} \).

How sharp is this estimate? Let's consider two examples. If \( A = \{1\}^k \times \{-1, 1\}^{n-k} \) is an axis parallel subspace of co-dimension \( k \) then \( \alpha = 2^{-k} \) and
\[ \Pr[x, y \in A] = \Pr[x_1 = \ldots = x_k = y_1, \ldots = y_k = 1] = \Pr[x_1 = y_1 = 1]^k = (1/3)^k = \alpha \log_2(3). \]
The other example is a hamming ball, in which case the exponent 3/2 is tight.

More generally, we may ask for a general \( 0 \leq \rho \leq 1 \) what is the probability that \( x, y \in A \), where \( (x, y) \) is a \( \rho \)-correlated pair. Using the same proof and applying the \( (p, 2) \) hypercontractive inequality we have
\[ \Pr[x, y \in A] = \text{Stab}_p(1_A) = \|T_{\sqrt{p}} 1_A\|_2^2 \leq \|1_A\|_{1+\rho}^2 = \alpha^{2/(1+\rho)}. \]

## 4 KKL Theorem

Let \( f : \{-1, 1\}^n \rightarrow \{-1, 1\} \). Recall that \( \text{Inf}_i(f) = \Pr[f(x) \neq f(x^{\oplus i})] \) are the influences of \( f \). Define \( \text{MaxInf}(f) = \max_i \text{Inf}_i(f) \) to be the maximal influence. Kahn, Kalai and Linial proved that any balanced boolean function must have a variable whose influence is \( \Omega(\frac{\log n}{n}) \). This was conjectured by Ben-Or and Linial, and is tight for the tribes function.

**Theorem 4.1 (KKL).** Let \( f : \{-1, 1\}^n \rightarrow \{-1, 1\} \). Then \( \text{MaxInf}(f) \geq \text{Var}(f) \cdot \Omega\left(\frac{\log n}{n}\right) \).

To simplify the presentation, we focus from now on on balanced functions, namely functions \( f \) with \( \mathbb{E}[f] = 0 \). For such functions \( \text{Var}(f) = 1 \). We first describe a variant of the KKL theorem, based on the total influence \( I(f) = \sum_i \text{Inf}_i(f) \).

**Theorem 4.2 (KKL with bounded total influence).** Let \( f : \{-1, 1\}^n \rightarrow \{-1, 1\} \) be a function with \( \mathbb{E}[f] = 0 \). Then
\[ \text{MaxInf}(f) \geq \frac{1}{I(f)^2 \text{Var}(f)} \cdot \text{Var}(f). \]

**Proof.** Fix \( i \in [n] \) and let \( g_i(x) = (D_i f)(x) = f(x) - f(x^{\oplus i}) \) so that \( \text{Inf}_i(f) = \Pr[g_i(x) \neq 0] = \frac{1}{2} \mathbb{E}[g_i(x)^2] \). Observe that as \( g_i : \{-1, 1\}^n \rightarrow \{-1, 0, 1\} \) we have \( \|g_i\|_{4/3} = \Pr[g_i(x) \neq 0]^{3/4} = \text{Inf}_i(f)^{3/4} \). By the \( (4/3, 2) \) hypercontractive inequality we have
\[ \text{Stab}_{1/3}(g_i) = \|T_{1/\sqrt{3}} g_i\|_{2}^2 \leq \|g_i\|_{4/3}^2 = \text{Inf}_i(f)^{3/2}. \]

Recall that \( \hat{g}_i(S) = 2\hat{f}(S) \) if \( i \in S \), and \( \hat{g}_i(S) = 0 \) if \( i \notin S \). So
\[ \text{Stab}_{1/3}(g_i) = \sum_{i \in S} (1/3)^{|S|} \hat{g}_i(S)^2 = 4 \sum_{i \in S} (1/3)^{|S|} \hat{f}(S)^2. \]
Summing this over all $i \in [n]$ gives
\[ 4 \sum (1/3)^{|S|} |S| \hat{f}(S)^2 \leq \sum \text{Inf}_i(f)^{3/2}. \]
As we assume $\mathbb{E}[f] = 0$ we can lower bound the LHS by
\[ \sum (1/3)^{|S|} |S| \hat{f}(S)^2 \geq \sum (1/3)^{|S|} \hat{f}(S)^2 = \text{Stab}_{1/3}(f). \]
By convexity of the function $x \to (1/3)^x$ we can lower bound $\text{Stab}_{1/3}(f)$ by
\[ \text{Stab}_{1/3}(f) = \sum (1/3)^{|S|} \hat{f}(S)^2 \geq (1/3) \sum \hat{f}(S)^2 |S| = (1/3)^{I(f)}. \]
So we obtained that
\[ 3^{-I(f)} \leq \sum \text{Inf}_i(f)^{3/2} \leq \text{MaxInf}(f)^{1/2} I(f). \]
Rearranging the terms gives
\[ \text{MaxInf}(f) \geq \frac{1}{I(f)^2 9^{I(f)}}. \]

We can now deduce the original KKL theorem.

Proof of KKL theorem. If $I(f) \geq c \log(n)$ then by averaging there exists $i \in [n]$ for which $\text{Inf}_i(f) \geq \frac{c \log(n)}{n}$. Otherwise as we just saw, $\text{MaxInf}(f) \geq \frac{1}{(c \log(n))^{9 \log \log n}}$. Choosing $c > 0$ so that $9^c < 2$ gives that $\text{MaxInf}(f) \geq \frac{c \log n}{n}$ for large enough $n$. \hfill \Box

As a corollary, we show that if $f : \{-1,1\}^n \to \{-1,1\}$ is a monotone function, viewed as a voting function, then we can “bribe” $O\left(\frac{1}{\log n}\right)$ fraction of the voters to vote for some $b \in \{-1,1\}$, such that the bias of the function to be very close to $b$.

Lemma 4.3. Let $f : \{-1,1\}^n \to \{-1,1\}$ be a monotone function. Then there exists $b \in \{-1,1\}$ and $S \subset [n]$ of size $|S| \leq O\left(\frac{n}{\log n}\right)$ such that the function $f$ restricted to the sub-cube $\{x : x_i = b \forall i \in S\}$, denoted $f|_{x_S = b}$, satisfies
\[ \mathbb{E}[f|_{x_S = b}] \geq 0.99. \]

Proof. Recall that for monotone functions $\text{Inf}_i(f) = \hat{f}(i)$. Assume without loss of generality that $\mathbb{E}[f] \geq 0$, and take $b = 1$. Consider the following iterative process. Initialize $S_0 = \emptyset$. Given $S_i$ define $f_i = f|_{x_{S_i} = 1}$. As long as $\mathbb{E}[f_i] < 0.99$, pick $j_i \notin S_i$ with maximal influence $\text{Inf}_{j_i}(f_i)$ and set $S_{i+1} = S_i \cup \{j_i\}$ and $f_{i+1} = f_i|_{x_{j_i} = 1}$. Observe that as $f$ is monotone we have
\[ \mathbb{E}[f_{i+1}] = \mathbb{E}[f_i] + \text{Inf}_{j_i}(f_i) \geq \mathbb{E}[f_i] + \Omega\left(\frac{\log(n - i)}{n - i}\right). \]
In particular this process halts after $O(n/ \log n)$ steps, as necessarily $\mathbb{E}[f_i] \leq 1$. \hfill \Box
5 Fridgut’s theorem

The final result we show is Friedgut’s theorem. It states that functions of low influence are close to juntas.

**Theorem 5.1 (Friedgut).** Let \( f : \{-1, 1\}^n \to \{-1, 1\} \) and let \( 0 < \varepsilon < 1 \). Then \( f \) is \( \varepsilon \)-close to a \( 2^{O(I(f)/\varepsilon)} \)-junta. More specifically, there exists a subset \( J \subset [n] \) of size \( |R| \leq 2^{O(I(f)/\varepsilon)} \) such that

\[
\sum_{S \subseteq J \mid |S| \leq O(I(f)/\varepsilon)} \hat{f}(S)^2 \geq 1 - \varepsilon.
\]

In particular, \( f \) is \( \varepsilon \)-close to a junta on \( J \).

The proof relies on the following lemma.

**Lemma 5.2.** Let \( f : \{-1, 1\}^n \to \{-1, 1\} \). Given \( 0 < \varepsilon < 1 \) and \( k \geq 1 \), define

\[
\tau = \frac{\varepsilon^2}{I(f)^2} 9^{-k}, \quad J = \{i : \text{Inf}_i(f) \geq \tau\}.
\]

Then \( f \) is \( \varepsilon \)-concentrated on the set \( \mathcal{F} = \{S : S \subseteq J\} \cup \{S : |S| > k\} \). Namely,

\[
\sum_{S \notin \mathcal{F}} \hat{f}(S)^2 \leq \varepsilon.
\]

**Proof.** Let \( g_i(x) = (D_i f)(x) = f(x) - f(x^{\oplus i}) \). We have seen that

\[
\sum_{S : i \in S} 3^{-|S|} \hat{f}(S)^2 = \text{Stab}_{1/3}(g_i) \leq \|g_i\|_{1/3}^2 = \text{Inf}_i(f)^{3/2}.
\]

Summing this over all \( i \notin J \) gives

\[
\sum_{S \cap J^c} |S| \cdot 3^{-|S|} \hat{f}(S)^2 \leq \sum_{i \notin J} \text{Inf}_i(f)^{3/2} \leq \left( \max_{i \notin J} \text{Inf}_i(f) \right)^{1/2} \cdot I(f) \leq \tau^{1/2} \cdot I(f) \leq 3^{-k} \varepsilon.
\]

In particular, if we restrict to \( S \notin \mathcal{F} \) then \( |S| \leq k \) and \( |S \cap J^c| \geq 1 \), so

\[
\sum_{S \notin \mathcal{F}} |S \cap J^c| 3^{-|S|} \hat{f}(S)^2 \geq 3^{-k} \sum_{S \notin \mathcal{F}} \hat{f}(S)^2.
\]

Putting these together gives

\[
\sum_{S \notin \mathcal{F}} \hat{f}(S)^2 \leq \varepsilon.
\]
Proof of Friedgut’s theorem. Let $k = I(f)/\varepsilon$ so that $f$ is $\varepsilon$-concentrated up to degree $k$. Let $J$ be as in the lemma, and note that

$$|J| \leq \frac{I(f)}{\tau} = \frac{I(f)^3}{\varepsilon^2} g^{I(f)/\varepsilon} \leq 2^{O(I(f)/\varepsilon)}.$$ 

Indeed, as $\sum_{S : S \subseteq J, |S| \geq k} \hat{f}(S)^2 \geq 1 - \varepsilon$ and $\sum_{S : |S| \geq k} \hat{f}(S)^2 \leq \varepsilon$ we have that

$$\sum_{S : S \subseteq J, |S| \leq k} \hat{f}(S)^2 \geq 1 - 2\varepsilon.$$ 

As we already saw earlier in the class, this shows that $f$ is $2\varepsilon$ close to the function

$$g(x) = \text{sign} \left( \sum_{S : S \subseteq J, |S| \leq k} \hat{f}(S)x^S \right).$$

In particular, $g$ is a junta which depends only on $\{x_i : i \in J\}$. \qed