Hypercontractivity and Friedgut's Junta Theorem (lecture notes)

Course: Analysis of Boolean Functions, Autumn 2025, University of Chicago
Instructor: William Hoza (`williamhoza@uchicago.edu`)

1 The Fourier-Entropy Conjecture

We've seen several interesting examples of classes of Boolean functions with low total influence. What does that buy us?

For starters, we showed previously that any $f\colon \{\pm 1\}^n \to \{\pm 1\}$ is concentrated at degree up to $O(\mathrm{I}[f])$, hence learnable from random examples in time $n^{O(\mathrm{I}[f])}$. For example, constant-width DNFs are learnable from random examples in $\mathrm{poly}(n)$ time. What more can we say?

Conjecture 1.1 (Fourier-Entropy Conjecture). *Let $f\colon \{\pm 1\}^n \to \{\pm 1\}$. Then $H[\mathcal{S}_f] \leq O(\mathrm{I}[f])$, where $H[\cdot]$ denotes Shannon entropy.*

It's okay if you aren't comfortable with Shannon entropy. For our purposes, the important thing is that Theorem 1.1 would imply that f is concentrated on a collection of $2^{O(\mathrm{I}[f])}$ Fourier coefficients, hence learnable from queries in time $2^{O(\mathrm{I}[f])} \cdot \mathrm{poly}(n)$. For example, it would immediately follow that DNFs of width $\log n$ are learnable from queries in $\mathrm{poly}(n)$ time.

The current best bound, by Kelman, Kindler, Lifshitz, Minzer, and Safra, says that any Boolean function f is concentrated on a collection of $2^{O(\mathrm{I}[f] \cdot \log \mathrm{I}[f])}$ Fourier coefficients [KKLMS20].[1] In these notes, we will prove a classic $2^{O(\mathrm{I}[f]^2)}$ bound by Friedgut. For example, this will show that DNFs of width $\sqrt{\log n}$ are learnable from queries in $\mathrm{poly}(n)$ time. The proof is based on *hypercontractivity.*

(Later, we will use other techniques to prove stronger bounds on the Fourier concentration and learnability of DNFs.)

2 Hypercontractivity

Definition 2.1 (p-norm of a function). If $f\colon \{\pm 1\}^n \to \mathbb{R}$ and $p \geq 1$, we define

$$\|f\|_p = \mathop{\mathbb{E}}_x[|f(x)|^p]^{1/p}.$$

Caution: We take an expectation over x, not a sum. Notably, this means that $\|f\|_p$ gets *bigger* as p gets bigger, which is the opposite of how the standard p-norm on $\mathbb{R}^n$ behaves. Indeed,

$$\|f\|_{p\cdot(1+\varepsilon)} = \mathop{\mathbb{E}}_x\left[|f(x)|^{p\cdot(1+\varepsilon)}\right]^{\frac{1}{p\cdot(1+\varepsilon)}} \geq \left(\mathop{\mathbb{E}}_x[|f(x)|^p]^{(1+\varepsilon)}\right)^{\frac{1}{p\cdot(1+\varepsilon)}} = \|f\|_p,$$

by Jensen's inequality.

2.1 Two-function hypercontractivity

Hypercontractivity can be understood as interpolating between two familiar facts about $f, g\colon \{\pm 1\}^n \to \mathbb{R}$.

- Fact 1: If $x, y \in \{\pm 1\}^n$ are uniform and independent, then

$$\mathbb{E}[f(x) \cdot g(y)] = \mathbb{E}[f(x)] \cdot \mathbb{E}[g(y)] \leq \|f\|_1 \cdot \|g\|_1.$$

[1] This result isn't explicitly stated in their paper, but it follows from their results by splitting into cases based on whether $\mathrm{Var}[f] \leq 0.01$.

- Fact 2: If $x \in \{\pm 1\}^n$ is uniform, then $\mathbb{E}[f(x) \cdot g(x)] \leq \sqrt{\mathbb{E}[f(x)^2]} \cdot \sqrt{\mathbb{E}[g(x)^2]} = \|f\|_2 \cdot \|g\|_2$ by the Cauchy-Schwarz inequality. More generally, by Hölder's inequality, for any $r > 0$, we have

$$\mathbb{E}[f(x) \cdot g(x)] \leq \|f\|_{1+r} \cdot \|g\|_{1+1/r}.$$

Theorem 2.2 (Two-Function Hypercontractivity Theorem)**.** *Let $f, g \colon \{\pm 1\}^n \to \mathbb{R}$, let $\rho \in [0,1]$, and sample a ρ-correlated pair $(x, y) \in (\{\pm 1\}^n)^2$. Then for every $r > 0$, we have*

$$\mathbb{E}[f(x) \cdot g(y)] \leq \|f\|_{1+r} \cdot \|g\|_{1+\rho^2/r}.$$

We recover Fact 2 by choosing $\rho = 1$. We recover Fact 1 by choosing $\rho = 0$ and taking $r \to 0$.

2.2 Hypercontractivity in terms of the noise operator

Hypercontractivity is more commonly presented in a slightly different way, based on the following definitions.

Definition 2.3 (Noise distribution)**.** Let $x \in \{\pm 1\}^n$ and $\rho \in [-1, 1]$. The *noise distribution* $N_\rho(x)$ is the distribution over $y \in \{\pm 1\}^n$ in which the coordinates are independent and for each $i \in [n]$, we have $\mathbb{E}[y_i] = \rho x_i$.

Note that if we sample x uniformly at random and then sample $y \sim N_\rho(x)$, then x and y are ρ-correlated.

Definition 2.4 (Noise operator)**.** Let $f \colon \{\pm 1\}^n \to \mathbb{R}$ and $\rho \in [-1, 1]$. We define $(T_\rho f) \colon \{\pm 1\}^n \to \mathbb{R}$ by

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

Observe that if x and y are ρ-correlated, then $\mathbb{E}[f(x) \cdot g(y)] = \langle T_\rho f, g \rangle = \langle f, T_\rho g \rangle$. (The noise operator is "self-adjoint.") Observe also that $T_\rho T_{\rho'} f = T_{\rho\rho'} f$. (This is the "semigroup property.") The hypercontractivity theorem can be re-stated as follows.

Theorem 2.5 (Hypercontractivity Theorem)**.** *For every $f \colon \{\pm 1\}^n \to \mathbb{R}$, $\rho \in [0, 1]$, and $r \in [0, \infty)$, we have $\|T_\rho f\|_{1+r} \leq \|f\|_{1+\rho^2 \cdot r}$.*

Theorem 2.5 says that T_ρ is not only "contractive," meaning it decreases the norm of f, but in fact it is "hypercontractive," meaning it decreases the norm of f even if we switch from the $(1 + \rho^2 \cdot r)$-norm to the $(1 + r)$-norm.

Theorem 2.5 implies Theorem 2.2, because Hölder's inequality gives us $\langle T_\rho f, g \rangle \leq \|T_\rho f\|_{1+r/\rho^2} \cdot \|g\|_{1+\rho^2/r}$. Conversely, Theorem 2.2 implies Theorem 2.5, because Hölder's inequality is tight: there is some g such that $\|T_\rho f\|_{1+r} = \langle T f, g \rangle$ and $\|g\|_{1+1/r} = 1$.

We will not prove the Hypercontractivity Theorem in full generality. However, we will prove a special case, which will be good enough for our applications. The proof uses the decomposition

$$f(x) = \sum_{S \ni i} \widehat{f}(S) \cdot x^S + \sum_{S \not\ni i} \widehat{f}(S) \cdot x^S.$$

The first term can be rewritten as $x_i \cdot D_i f$, where D_i is the derivative operator. Meanwhile, the second term is $E_i f$, where E_i is the *expectation operator*.

Definition 2.6 (Expectation operator)**.** Let $f \colon \{\pm 1\}^n \to \mathbb{R}$ and let $i \in [n]$. We define $(E_i f) \colon \{\pm 1\}^n \to \mathbb{R}$ by

$$(E_i f)(x) = \sum_{S \not\ni i} \widehat{f}(S) \cdot \chi_S(x).$$

Proof of Theorem 2.5 when $r = 3$ and $\rho = 1/\sqrt{3}$. For brevity, let $T = T_\rho$. Our goal is to prove

$$\mathbb{E}[(Tf)^4] \le \mathbb{E}[f^2]^2.$$

We prove it by induction on n. When $n = 0$, the theorem is trivial, so assume $n > 0$. We decompose f into the monomials involving x_n and the monomials that do not involve x_n: $f(x) = x_n \cdot (D_n f)(x) + (E_n f)(x)$. For brevity, let us denote $f = f(x)$, $d = (D_n f)(x)$, and $e = (E_n f)(x)$. Note that neither d nor e depends on x_n. Then

$$Tf = T(dx_n) + Te = \rho \cdot x_n \cdot Td + Te,$$

so

$$\begin{aligned}
\mathbb{E}[(Tf)^4] &= \mathbb{E}\left[(\rho x_n Td)^4 + 4(\rho x_n Td)^3 \cdot Te + 6(\rho x_n Td)^2 \cdot (Te)^2 + 4(\rho x_n Td) \cdot (Te)^3 + (Te)^4\right] \\
&= \frac{1}{9} \cdot \mathbb{E}[(Td)^4] + 2\,\mathbb{E}[(Td)^2 \cdot (Te)^2] + \mathbb{E}[(Te)^4] \\
&\le \mathbb{E}[(Td)^4] + 2\sqrt{\mathbb{E}[(Td)^4] \cdot \mathbb{E}[(Te)^4]} + \mathbb{E}[(Te)^4] \\
&\le \mathbb{E}[d^2]^2 + 2\,\mathbb{E}[d^2] \cdot \mathbb{E}[e^2] + \mathbb{E}[e^2]^2 && \text{(Induction)} \\
&= \left(\mathbb{E}[d^2] + \mathbb{E}[e^2]\right)^2.
\end{aligned}$$

Meanwhile,

$$\mathbb{E}[f^2] = \mathbb{E}[(dx_n)^2 + 2dx_n e + e^2] = \mathbb{E}[d^2] + \mathbb{E}[e^2]. \qquad \square$$

2.3 Bonami's lemma

There is a third way of understanding hypercontractivity. We have the following Fourier formula for the noise operator:

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

Basically, the noise operator dampens the high-degree Fourier coefficients of f. Intuitively, therefore, if f is a low-degree function to begin with, then applying the noise operator shouldn't be necessary. The simplest formalization of this idea is called Bonami's Lemma.

Lemma 2.7 (Bonami's Lemma)**.** *For any $f\colon \{\pm 1\}^n \to \mathbb{R}$, we have $\|f\|_4 \le 3^{\deg(f)/2} \cdot \|f\|_2$.*

Lemma 2.8 (Generalized Bonami's Lemma)**.** *For any $f\colon \{\pm 1\}^n \to \mathbb{R}$ and any $r \ge 1$, we have*

$$\|f\|_{1+r} \le r^{\deg(f)/2} \cdot \|f\|_2.$$

Proof. Extend the definition of T_ρ to the case $\rho > 1$ using the Fourier formula:

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

Then

$$\begin{aligned}
\|f\|_{1+r}^2 = \|T_{1/\sqrt{r}} T_{\sqrt{r}} f\|_{1+r}^2 &\le \|T_{\sqrt{r}} f\|_2^2 && \text{by the Hypercontractivity Theorem} \\
&= \sum_{S \subseteq [n]} (\sqrt{r})^{2|S|} \cdot \widehat{f}(S)^2 \\
&\le r^{\deg(f)} \cdot \|f\|_2^2. && \square
\end{aligned}$$

Bonami's lemma and its generalization bound $\|f\|_{\text{bigger than 2}}$ in terms of $\|f\|_2$. We can also bound $\|f\|_2$ in terms of $\|f\|_{\text{smaller than 2}}$. We will do two versions of this bound. In the first version, instead of *assuming* that f has low degree, we *delete* the high-degree Fourier coefficients. Define $f^{\le k} = \sum_{S \subseteq [n], |S| \le k} \widehat{f}(S) \cdot \chi_S$. Then we have the following.

Theorem 2.9 (Hypercontractivity of the "low-degree part" operator)**.** *For any $f\colon \{\pm 1\}^n \to \mathbb{R}$, $k \in \mathbb{N}$, and $\alpha \in (0,1)$, we have*

$$\|f^{\leq k}\|_2 \leq (1/\alpha)^{k/2} \cdot \|f\|_{1+\alpha}.$$

Proof.

$$\begin{aligned} \|f^{\leq k}\|_2^2 = \langle f^{\leq k}, f^{\leq k}\rangle = \langle f^{\leq k}, f\rangle &\leq \|f^{\leq k}\|_{1+1/\alpha} \cdot \|f\|_{1+\alpha} && \text{by Hölder's inequality} \\ &\leq (1/\alpha)^{k/2} \cdot \|f^{\leq k}\|_2 \cdot \|f\|_{1+\alpha} && \text{by the Generalized Bonami Lemma.} \end{aligned}$$

If $\|f^{\leq k}\|_2 \neq 0$, we can divide both sides by $\|f^{\leq k}\|_2$ to complete the proof. If $\|f^{\leq k}\|_2 = 0$, then the theorem is trivial. □

In the second version, we have to assume that f has low degree, but the benefit is that we can go all the way down to $\alpha = 0$.

Theorem 2.10 (The 1-norm trick)**.** *For any $f\colon \{\pm 1\}^n \to \mathbb{R}$, we have $\|f\|_2 \leq 2^{O(\deg(f))} \cdot \|f\|_1$.*

Proof.

$$\begin{aligned} \|f\|_2^2 = \mathop{\mathbb{E}}_x\left[|f(x)|^{4/3} \cdot |f(x)|^{2/3}\right] = \langle |f|^{4/3}, |f|^{2/3}\rangle &\leq \||f|^{4/3}\|_3 \cdot \||f|^{2/3}\|_{3/2} && \text{by Hölder's inequality} \\ &= \|f\|_4^{4/3} \cdot \|f\|_1^{2/3} \\ &\leq (3^{\deg(f)/2} \cdot \|f\|_2)^{4/3} \cdot \|f\|_1^{2/3} && \text{by Bonami's Lemma.} \end{aligned}$$

If $\|f\|_2 \neq 0$, then we can divide both sides by $\|f\|_2^{4/3}$ to get $\|f\|_2^{2/3} \leq 3^{\deg(f)\cdot 2/3} \cdot \|f\|_1^{2/3}$, which gives us $\|f\|_2 \leq 3^{\deg(f)} \cdot \|f\|_1$.[2] If $\|f\|_2 = 0$, the theorem is trivial. □

Hypercontractivity in terms of degree is a bit messier than hypercontractivity in terms of noise, but hypercontractivity in terms of degree is more useful for applications.

3 Friedgut's junta theorem

Which functions have total influence $O(1)$? Friedgut's junta theorem gives a satisfying answer.

Definition 3.1. Let $f\colon \{\pm 1\}^n \to \mathbb{R}$. We say that f is a *k-junta* if f only depends on at most k variables.

Theorem 3.2 (Friedgut's junta theorem[3])**.** *Let $f\colon \{\pm 1\}^n \to \{\pm 1\}$ and let $\varepsilon \in (0,1)$.*

1. *There is a set $J \subseteq [n]$ of size $2^{O(\mathrm{I}[f]/\varepsilon)}$ such that f is ε-concentrated on the subsets of J.*
2. *The function f is ε-close to a k-junta where $k = 2^{O(\mathrm{I}[f]/\varepsilon)}$.*
3. *The function f is ε-concentrated on a collection of $2^{O(\mathrm{I}[f]^2/\varepsilon^2)}$ Fourier coefficients.*

[2] The bound can be improved to $e^{\deg(f)} \cdot \|f\|_1$.

[3] "Friedgut's junta theorem" typically refers to Item 2 alone, but I would argue that Item 1 is the real meat of the theorem.

Proof. Let $J = \{i \in [n] : \mathrm{Inf}_i \geq \tau\}$, for a suitable number τ that we will choose later. Let $k = 2\,\mathrm{I}[f]/\varepsilon$. Then

$$\begin{aligned}
\sum_{S \not\subseteq J} \widehat{f}(S)^2 = \sum_{\substack{S \not\subseteq J \\ |S| > k}} \widehat{f}(S)^2 + \sum_{\substack{S \not\subseteq J \\ |S| \leq k}} \widehat{f}(S)^2 &\leq \frac{\varepsilon}{2} + \sum_{i \notin J} \sum_{\substack{S \ni i \\ |S| \leq k}} \widehat{f}(S)^2 && \text{by choice of } k \\
&= \frac{\varepsilon}{2} + \sum_{i \notin J} \|(D_i f)^{\leq k-1}\|_2^2 \\
&\leq \frac{\varepsilon}{2} + 3^k \cdot \sum_{i \notin J} \|D_i f\|_{4/3}^2 && \text{by Theorem 2.9} \\
&= \frac{\varepsilon}{2} + 3^k \cdot \sum_{i \notin J} \mathop{\mathbb{E}}_x[|(D_i f)(x)|^{4/3}]^{3/2} \\
&= \frac{\varepsilon}{2} + 3^k \cdot \sum_{i \notin J} \mathrm{Inf}_i[f]^{3/2} \\
&\leq \frac{\varepsilon}{2} + 3^k \cdot \sqrt{\tau} \cdot \mathrm{I}[f] && \text{by choice of } J \\
&\leq \varepsilon,
\end{aligned}$$

provided we choose $\tau = \left(\frac{\varepsilon}{3^k \cdot \mathrm{I}[f]}\right)^2$. Furthermore, $|J| \leq \mathrm{I}[f]/\tau = 2^{O(\mathrm{I}[f]/\varepsilon)}$, completing the proof of Item 1.

The junta is $g(x) = \mathrm{sign}(\sum_{S \subseteq J} \widehat{f}(S) \cdot \chi_S(x))$, completing the proof of Item 2.

Finally, the computations above show that f is ε-concentrated on $\{S \subseteq J : |S| \leq 2\,\mathrm{I}[f]/\varepsilon\}$, which has cardinality at most $(|J|+1)^{2\,\mathrm{I}[f]/\varepsilon} = 2^{O(\mathrm{I}[f]^2/\varepsilon^2)}$. This completes the proof of Item 3. $\square$

References

[KKLMS20] Esty Kelman, Guy Kindler, Noam Lifshitz, Dor Minzer, and Muli Safra. “Towards a proof of the Fourier-entropy conjecture?” In: *Geometric and Functional Analysis* 30.4 (2020), pp. 1097–1138.