Hypercontractivity Via the Entropy Method

Dariusz Lasecki

July 4, 2019

1 Motivation and history

The Hypercontractive Inequality upper bounds the action of the noise operator T_ρ (linear) on functions $f : \{-1,1\}^n \to \mathbb{R}$ with respect to norms. It turns out to be extremely useful in the analysis of Boolean functions. It plays an essential role in proofs in many areas, e.g. in random graphs, distributed computing, statistical physics or k-SAT. Surprisingly, it often provides the only known proof of theorems in the mentioned disciplines. The inequality itself was first discovered by Bonami [1], [2] and Gross [3]. However, it was in the work of Kahn, Kalai and Linial [7] when its applicability in theoretical computer science and connection to the anlysis of Bollean functions was realized. The inequality has analytic proofs which usually make use of induction. However, due to its importance, alternative proofs of the Hypercontractive Inequality are desired. Friedgut and Rödl were the first to connect the Hypercontractive Inequality with the Shannon Entropy [8]. However, their work only provided results for certain special cases and used overcomplicated, as it turned out, notions like hypergraphs. In this paper we present the proof of the Hypercontractive Inequality by Blais and Tan [6] which is more general than results of Friedgut and Rödl and uses only basic concepts from the Shannon Theory.

2 Introduction

The Hypercontractive Inequality quantifies the action of T_ρ (linear noise operator) on functions $f : \{-1,1\}^n \to \mathbb{R}$. The action of the noise operator is defined as $T_\rho f(x) = \mathbb{E}[f(y)]$, where y is drawn from the product distribution such that $\forall i \in [n]$ we have $\mathbb{E}[y_i x_i] = \rho$. Alternatively, the action of the noise operator can be expressed via the action on the Fourier coefficients $\hat{f}(S)$ of f in the Fourier expansion as

$$T_\rho f = \sum_{S \subseteq [n]} \rho^{|S|} \hat{f}(S) \chi_S, \quad \text{where} \quad \chi_S(x) = \prod_{i \in S} x_i \quad \text{and} \quad \hat{f}(S) = \langle f, \chi_S \rangle = \mathbb{E}[f \cdot \chi_S].$$

The expression above can be obtained by the following calculation.

$$T_\rho \chi_S(x) = \mathbb{E}[\chi_S(y)] = \mathbb{E}\left[\prod_{i \in S} y_i\right] = \mathbb{E}\left[\prod_{i \in S} y_i \left(\prod_{i \in S} x_i\right)^2\right] = \prod_{i \in S} x_i \mathbb{E}\left[\prod_{i \in S} y_i x_i\right] =$$

$$\chi_S(x)\prod_{i\in S}\mathbb{E}\left[y_i x_i\right] = \chi_S(x)\prod_{i\in S}\rho = \chi_S(x)\rho^{|S|},$$

where we used that fact that $\left(\prod_{i\in S} x_i\right)^2 = 1$ since $x_i \in \{-1, 1\}$. Then, by the fact that T_ρ is linear, we can see that

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

We note that every real function f on $\{-1, +1\}$ has a unique Fourier expansion. It other words, its coefficients uniquely determine the function f.
Having made necessary definitions, we can now state the Hypercontractive Inequality.

Theorem 1 (Hypercontractive Inequality)**.** *Let $f : \{-1, 1\}^n \to \mathbb{R}$ and $0 < \rho < 1$. Then,*

$$||T_\rho f||_2 \leq ||f||_{1+\rho^2}.$$

However, via duality, as proved in [10], the Hypercontractive Inequality as defined above is equivalent to the inequality stated in the Theorem that follows.

Theorem 2. *For any $f : \{-1, 1\}^n \to \mathbb{R}$, let $f^{(=m)} = \sum_{|S|=m}\hat{f}(S)\chi_S$ denote the projection of f to its degree m-part. Then for all $q > 2$,*

$$||f^{(=m)}||_q \leq (q-1)^{\frac{m}{2}}||f^{(=m)}||_2.$$

In this paper, due to the nature of the information-theoretic tools used, we will prove the above Theorem for the case of **even** integers q.

3 Basic definitions and lemmas

Definition 1 (Symmetric difference)**.** *Let S_1 and S_2 be sets. The symmetric difference Δ is defined as*

$$S_1 \Delta S_2 = (S_1 \setminus S_2) \cup (S_2 \setminus S_1).$$

Lemma 1 (Universal upper bound)**.** *For any random variable X over the sample space Ω*

$$H(X) \leq \log|supp(X)| \leq \log|\Omega|.$$

Lemma 2 (Chain rule)**.** *The entropy of a sequence $X_1, ..., X_n$ of random variables satisfies*

$$H(X_1, ..., X_n) = \sum_{i=1}^{n} H(X_i|X_1, ..., X_{i-1}).$$

Lemma 3 (Shearer's Lemma)**.** *Let $\boldsymbol{X} \in \Omega^n$ be a vector of random variables and let $S_1, ..., S_m \subseteq [n]$ be a collection of sets that cover each element in $[n]$ at least t times. Then*

$$H(\boldsymbol{X}) \leq \frac{1}{t}\sum_{j=1}^{m} H(\boldsymbol{X}_{S_j}).$$

4 Proof via the entropy method

We recall that we defined $f^{(=m)} = \sum_{|S|=m} \hat{f}(S)\chi_S$. Thus, we will be interested in subsets of indices $S \subseteq [n] : |S| = m$. For each such subset S_k (there are $\binom{n}{m}$ of them), we define the witness set W_{S_k}. Each W_{S_k} consists of $|\hat{f}(S_k)|$ elements called witnesses for the corresponding set S_k. For any two distinct sets $S_i \neq S_j$, witness sets W_{S_i} and W_{S_j} are disjoint. It means that having any element from the witness set W_{S_k} we can uniquely determine the corresponding set S_k. We also define the union of all $\binom{n}{m}$ witness sets as $\mathcal{W} = \cup_i W_{S_i}$. Having the subset of q witnesses from $\mathcal{W}$, we will call it a legal q-tuple if sets $S_1, ..., S_q$ that they witness satisfy $S_1 \Delta ... \Delta S_q = \emptyset$. The reason for introducing this condition will become clear soon.
Let X be a random variable which is uniformly distributed over $\mathcal{W}^q$ which is the set of all ordered legal q-tuples. Having an instance of X, say $x = (w_1, ..., w_q)$ we define $\mathbf{S} = (S_1, ..., S_q)$ to be the tuple of sets which correspond to witnesses from x. Notice that $\mathbf{S}$ is itself a random variable. We consider the entropy $H(X\mathbf{S})$ which by the chain rule for entropy can be expressed as

$$H(X\mathbf{S}) = H(X) + H(\mathbf{S}|X).$$

We note that since X represents witnesses for sets in $\mathbf{S}$, the conditional entropy above equals 0. Therefore, we can write

$$H(X\mathbf{S}) = H(X) = \log \left(\sum_{\substack{S_1,...,S_q \\ S_1 \Delta ... \Delta S_q = \emptyset}} |\hat{f}(S_1)...\hat{f}(S_q)| \right),$$

where we expressed the entropy of a uniform random variable as the logarithm of the size of the sample space for that random variable (recall a witness set for a subset S_k has $|\hat{f}(S_k)|$ elements, thus for each entry of x there are that many possibilities for a witness element). We can bound the equality from below by moving the absolute value outside, which gives

$$H(X\mathbf{S}) \geq \log \left| \sum_{\substack{S_1,...,S_q \\ S_1 \Delta ... \Delta S_q = \emptyset}} \hat{f}(S_1)...\hat{f}(S_q) \right| = \log \left(||f^{(=m)}||_q^q \right).$$

The equality above holds only for even values of q and the calculation can be found below. We start with the fact that for even q, the q-norm of a function can be expressed as an expected value of that function.

$$||f^{(=m)}||_q^q = \mathbb{E}[(f^{(=m)})^q] = \sum_{S_1,...,S_q} \hat{f}(S_1)...\hat{f}(S_q)\mathbb{E}[\chi_{S_1}...\chi_{S_q}].$$

We make use of the following observation from [12]

$$\mathbb{E}[\chi_{S_1}...\chi_{S_q}] = \mathbb{E}[\chi_{S_1 \Delta ... \Delta S_q}] = \prod_{i \in S_1 \Delta ... \Delta S_q} \mathbb{E}[\chi_{\{i\}}] = \begin{cases} 1 & \text{if } S_1 \Delta ... \Delta S_q = \emptyset \\ 0 & \text{otherwise} \end{cases}.$$

Therefore, we have

$$||f^{(=m)}||_q^q = \sum_{\substack{S_1,...,S_q \\ S_1\Delta...\Delta S_q=\emptyset}} \hat{f}(S_1)...\hat{f}(S_q).$$

Now, we would like to obtain the upper bound on $H(X\mathbf{S})$. To do so, we start by defining a sequence $\mathbf{T}$ of sets, each of which having at most m elements and each set being indexed by subsets of $[q]$ of size 2. So, we have the following sequence of $\binom{q}{2}$ sets

$$\mathbf{T} = (\mathbf{T}_{\{1,2\}}, \mathbf{T}_{\{1,3\}}, ..., \mathbf{T}_{\{q-1,q\}}).$$

Every set $\mathbf{T}_{\{i,j\}}$, for $i < j \in [q]$ is defined as

$$\mathbf{T}_{\{i,j\}} = \left\{ x \in [n] : x \in (\mathbf{S_i} \cap \mathbf{S_j}) \setminus (\mathbf{S_{i+1}} \cup ... \cup \mathbf{S_{j-1}}) \text{ and } \#\left\{\ell \in [q] : \ell < i, x \in \mathbf{S}_\ell\right\} \text{ is even } \right\}.$$

We can imagine the set defined as above refers to sets $\mathbf{S_i}, \mathbf{S_{i+1}}, ..., \mathbf{S_j}$ from the q sets indicated by the witnesses. The first condition on x ensures that it belongs to both the first and the last set under consideration but not to the sets between them. The second condition ensures that x belongs to an even number of preceding sets. These conditions reflect the fact that we consider sets $S_1, ..., S_q$ such that $S_1\Delta...\Delta S_q = \emptyset$. As a result, if an element x belongs to k sets which we label $S_{i_1}, ..., S_{i_k}$ for $i_1 \leq ... \leq i_k$, then x belongs to sets $\mathbf{T}_{\{i_1,i_2\}}, \mathbf{T}_{\{i_3,i_4\}}, ..., \mathbf{T}_{\{i_{k-1},i_k\}}$. The whole purpose of what we described is to finally define the sequence $\mathbf{T}_{\{i,*\}}$ which will give us the partition of S_i for every i. We define it as,

$$\mathbf{T}_{\{i,*\}} := (\mathbf{T}_{\{i,j\}})_{j \neq i} = (\mathbf{T}_{\{1,i\}}, \mathbf{T}_{\{2,i\}}, ..., \mathbf{T}_{\{i-1,i\}}, \mathbf{T}_{\{i,i+1\}}, ..., \mathbf{T}_{\{i,q\}}).$$

Once we fix i and consider an element $x \in \mathbf{S}_i$, there is only one j for which it will hold that $x \in \mathbf{T}_{\{i,j\}}$. It follows directly from how we defined $\mathbf{T}_{\{i,j\}}$. Therefore, we indeed see that every $x \in \mathbf{S}_i$ belongs to exactly one set $\mathbf{T}_{\{i,j\}}$ for some $j \neq i$ so $\mathbf{T}_{\{i,*\}}$ defines the partition of the set $\mathbf{S}_i$ under consideration. It implies that when we consider our sequences $\mathbf{S} = (\mathbf{S_1}, ..., \mathbf{S_q})$ and $\mathbf{T} = (\mathbf{T}_{\{1,*\}}, \mathbf{T}_{\{2,*\}}, ..., \mathbf{T}_{\{*,q\}})$, we will have $H(\mathbf{S}) = H(\mathbf{T})$ and by using the chain rule twice we obtain the following.

$$H(X\mathbf{S}) = H(\mathbf{S}) + H(X|\mathbf{S}) = H(\mathbf{T}) + H(X|\mathbf{S}) = H(\mathbf{T}) + \sum_{i=1}^{q} H(X_i|X_1, ..., X_{i-1}, \mathbf{S_1}, ..., \mathbf{S_n}).$$

We recall that when we consider any set $\mathbf{S_i}$, the random variable X_i corresponds to the uniform distribution over all the witnesses for $\mathbf{S_i}$ and is also independent of other X_k and of corresponding S_k for $k \neq i$. It allows us to simplify the previous equation.

$$H(X\mathbf{S}) = H(\mathbf{T}) + \sum_{i=1}^{q} H(X_i|\mathbf{S_i}).$$

Each set in $\left\{\mathbf{T}_{\{i,j\}}\right\}_{i \neq j}$ is covered by the elements of $\mathbf{T}$ twice. In this case, we can invoke the Shearer's Lemma defined earlier with the parameter $t = 2$ and obtain the following.

$$H(\mathbf{T}) \leq \frac{1}{2} \sum_{i=1}^{q} H(\mathbf{T}_{\{i,*\}}).$$

By the partition property we know that $H(\mathbf{T}_{\{i,*\}}) = H(\mathbf{S}_i\mathbf{T}_{\{i,*\}})$. We can also expand the RHS term with the chain rule to obtain $H(\mathbf{T}_{\{i,*\}}) = H(\mathbf{S}_i) + H(\mathbf{T}_{\{i,*\}}|\mathbf{S}_i)$. We can calculate the size of the sample space for $H(\mathbf{T}_{\{i,*\}}|\mathbf{S}_i)$ and upper bound the entropy by the logarithm of the size of that sample space. Given $\mathbf{S}_i$ of size m, we can have $(q-1)^m$ possible partitions because we can place $(q-1)$ separators in m possible places. Therefore, we have

$$H(\mathbf{T}_{\{i,*\}}|\mathbf{S}_i) \leq \log((q-1)^m) = m\log(q-1).$$

Combining our calculations, we obtain

$$H(\mathbf{T}_{\{i,*\}}) \leq H(\mathbf{S}_i) + m\log(q-1),$$

and then,

$$H(\mathbf{T}) \leq \frac{1}{2}\sum_{i=1}^{q}(H(\mathbf{S}_i) + m\log(q-1)) = \frac{qm\log(q-1)}{2} + \frac{1}{2}\sum_{i=1}^{q}H(\mathbf{S}_i).$$

We come back to our expression for $H(X\mathbf{S})$ which can now be written as

$$H(X\mathbf{S}) \leq \frac{qm\log(q-1)}{2} + \frac{1}{2}\sum_{i=1}^{q}H(\mathbf{S}_i) + \sum_{i=1}^{q}H(X_i|\mathbf{S}_i) = \frac{qm\log(q-1)}{2} + \frac{1}{2}\sum_{i=1}^{q}\left(H(\mathbf{S}_i) + 2H(X_i|\mathbf{S}_i)\right).$$

For the right-most sum we can also use the bound based on the size of the sample size. To make it more convenient, we use the following trick.

$$H(\mathbf{S}_i) + 2H(X_i|\mathbf{S}_i) = H(\mathbf{S}_i) + H(X_i|\mathbf{S}_i) + H(X_i'|\mathbf{S}_i) = H(X_i\mathbf{S}_i) + H(X_i'|\mathbf{S}_i) =$$

$$= H(X_i\mathbf{S}_i) + H(X_i'|\mathbf{S}_iX_i) = H(X_iX_i'\mathbf{S}_i),$$

where we introduced another draw of a witness corresponding to the set $\mathbf{S}_i$ which we denoted by another uniformly distributed random variable X_i'. It gives us the following

$$H(\mathbf{S}_i) + 2H(X_i|\mathbf{S}_i) = H(X_iX_i'\mathbf{S}_i).$$

We recall that there are $\hat{f}(S_i)$ possibilities for every witness of the set $\mathbf{S}_i$, therefore we have

$$H(\mathbf{S}_i) + 2H(X_i|\mathbf{S}_i) \leq \log\left(\sum_{S:|S|=m}|\hat{f}(S)|^2\right) = \log\left(\sum_{S:|S|=m}|\hat{f}(S)\chi_S|^2\right) = \log\left(||f^{(=m)}||_2^2\right).$$

Finally, we obtain

$$H(X\mathbf{S}) \leq \frac{qm\log(q-1)}{2} + \frac{1}{2}\sum_{i=1}^{q}\log\left(||f^{(=m)}||_2^2\right) = \frac{qm\log(q-1)}{2} + \frac{q\log\left(||f^{(=m)}||_2^2\right)}{2} =$$

$$= \frac{q}{2}\log\left((q-1)^m||f^{(=m)}||_2^2\right).$$

For convenience, we restate our lower bound on $H(X\mathbf{S})$

$$H(X\mathbf{S}) \geq \log\left(||f^{(=m)}||_q^q\right).$$

It results in the following combined expression

$$\frac{q}{2}\log\left((q-1)^m||f^{(=m)}||_2^2\right) \geq H(X\mathbf{S}) \geq \log\left(||f^{(=m)}||_q^q\right),$$

and therefore

$$\frac{q}{2}\log\left((q-1)^m||f^{(=m)}||_2^2\right) \geq \log\left(||f^{(=m)}||_q^q\right).$$

It can be further simplified

$$\log\left((q-1)^m||f^{(=m)}||_2^2\right) \geq \log\left(||f^{(=m)}||_q^2\right),$$

$$(q-1)^m||f^{(=m)}||_2^2 \geq ||f^{(=m)}||_q^2,$$

$$||f^{(=m)}||_q \leq (q-1)^{\frac{m}{2}}||f^{(=m)}||_2.$$

We completed the proof.

5 Applications

In this section we will briefly characterize two applications of the Hypercontractive Inequality.

Collective Coin Flipping

The Collective Coin Flipping [11] is the scenario with n players, each of whom controls one bit in an n-bit string. Players assign a binary value to their bit to collectively form an input to a balanced function $f : \{0,1\}^n \to \{0,1\}$. A balanced function takes the value of 1 for exactly half of possible inputs and the value of 0 for another half of possible inputs. We can observe that if we assume that each player chooses the value for their bit uniformly at random, the protocol produces a uniformly random output (fair coin flip), hence the name Collective Coin Flipping. However, we may assume that a certain subset of players cheat, i.e. they can see the values assigned by other players and change the value of their own bit hoping to influence the output of the function f. To measure this kind of influence, we define the influence of variable i

$$Inf_i(f) = \Pr[f(x) \neq f(x \oplus e_i)],$$

where $x \oplus e_i$ means changing the value of i-th bit and the probability is uniform over $x \in \{0,1\}^n$. Then, using the Hypercontractive Inequality we obtain the result [10]

$$\sum_{i=1}^{n} \frac{Inf_i(f)}{\log(1/Inf_i(f))} = \Omega(Var[f]),$$

where $Var[f] = \mathbb{E}[f^2] - \mathbb{E}[f]^2$ and we assume that no variable has influence of 0 or 1. For a balanced or close to balanced f, we have $Var[f] = \Omega(1)$. With this assumption from the equation above it follows that

$$\exists i : Inf_i(f) = \Omega\left(\frac{\log n}{n}\right).$$

This special case is actually the result obtained by Kahn, Kalai and Linial in [7]. It is worth noting that it can also be proved that there exists a set of $O(\frac{n}{\log n})$ variables which in most cases determine the value of the function f despite the value of the rest of bits (this is not an immediate result). It means that protocols for the collective coin flipping are not secure in case of this kind of attack.

Random parities over a fixed set

Another application which leads to interesting conclusions for cryptography is the problem of Random parities over a fixed set. We consider a subset $A \subseteq \{0,1\}^n$ from which we will draw an n-bit string $x \in A$ uniformly at random. In the same uniform way we also draw a subset of k indices $S \subseteq [n]$. We are interested in quantifying the parameter β_S which measures the bias defined as

$$\beta_S = \mathbb{E}_{x \in A}[\chi_S(x)].$$

By using the Hypercontractive Inequality we can obtain the following result which bounds the value of the sum of squared biases for k-bit parities.
For any $\delta \in [0,1]$, $A \subseteq \{0,1\}^n$ we have

$$\sum_{S \in \binom{[n]}{k}} \beta_S^2 \leq \frac{1}{\delta^k} \left(\frac{2^n}{|A|} \right)^{2\delta}.$$

An interesting conclusion is that for large A we obtain small biases. Therefore, if we assume the presence of an adversary who knows that $x \in A$ is uniformly distributed over a large A, they cannot predict most parities of bits selected from x.

6 Summary

We presented the Hypercontractive Inequality together with its proof which uses the basic entropic arguments and information-theoretic tools (for the case of even parameters q). As we explained, the inequality plays the crucial role in numerous areas of research and the examples that we summarized include bounds for parity relevant for cryptography and sensitivity of influences for collective behaviors. This new entropic approach for proving the inequality may provide even more insight and possibly widen the spectrum of its applicability in the course of further research.

References

[1] Aline Bonami: Ensembles $\Lambda(p)$ dans le dual de D^{∞}. Annales de lInstitut Fourier, 18(2):193204, 1968.

[2] Aline Bonami: tude des coefficients Fourier des fonctions de $L^{p}(G)$. Annales de lInstitut Fourier, 20(2):335402, 1970.

[3] Leonard Gross: Logarithmic Sobolev inequalities. American Journal of Mathematics, 97(4):10611083, 1975.

[4] Jeff Kahn, Gil Kalai and Nathan Linial: The influence of variables on boolean functions (extended abstract). In Proc. 29th FOCS, pp. 6880. IEEE Comp. Soc. Press, 1988

[5] Ehud Friedgut AND Vojtech Rödl: Proof of a hypercontractive estimate via entropy. Israel J. Math., 125(1):369380, 2001.

[6] Eric Blais and Li-Yang Tan, Hypercontractivity, Via the Entropy Method, Theory of Computing, Volume 9 (29), 2013, pp. 889 - 896.

[7] Jeff Kahn, Gil Kalai and Nathan Linial: The influence of variables on boolean functions (extended abstract). In Proc. 29th FOCS, pp. 6880. IEEE Comp. Soc. Press, 1988

[8] Ehud Friedgut AND Vojtech Rödl: Proof of a hypercontractive estimate via entropy. Israel J. Math., 125(1):369380, 2001.

[9] Ronald de Wolf: A Brief Introduction to Fourier Analysis on the Boolean Cube. Number 1 in Graduate Surveys. Theory of Computing Library, 2008.

[10] Michel Talagrand, On Russos approximate zero-one law, Ann. Probab., 22(3):15761587, 1994.

[11] Michael Ben-Or and Nathan Linial: Collective coin flipping. Randomness and Computation, volume 5 of Advances in Computing Research: A research annual, pp. 91115. JAI Press, 1989.

[12] Analytical Methods in Combinatorics and Computer Science (Lecture 2) Nov. 9, 2005 [http://www.cs.huji.ac.il/ analyt/scribes/L02.pdf, accessed: 10 April 2018]