Homework policy: I encourage you to try to solve the problems by yourself. However, you may collaborate as long as you do the writeup yourself and list the people you talked with.

Do 4 out of 6.

1. Orthogonal decomposition. Given any \( f : \{−1, 1\}^n \rightarrow \mathbb{R} \), consider \( f_S : \{−1, 1\}^n \rightarrow \mathbb{R} \) defined by \( f_S(x) = \hat{f}(S) \chi_S(x) \). We have \( f = \sum_{S \subseteq [n]} f_S \) as functions, and this “orthogonal decomposition” has the following three properties:

   (i) \( f_S(x) \) depends only on the coordinates of \( x \) in \( S \);

   (ii) \( \mathbb{E}_x[f^S(x)f^T(x)] = 0 \) if \( S \neq T \);

   (iii) \( \sum_{T \subseteq S} f^T \), denoted \( f^\leq S \), gives the conditional expectation of \( f \) conditioned on the coordinates in \( S \).

(a) Prove property (iii); i.e., \( f^\leq S(x) = \mathbb{E}_x[f_{\bar{x} \rightarrow S}] \), where the expectation is over the bits in \( \bar{S} = [n] \setminus S \). (Here the notation is that \( x \in \{−1, 1\}^n \), but in the expression \( f_{\bar{x} \rightarrow S} \), we only restrict the \( S \)-coordinates of \( f \) using the \( S \)-bits of \( x \); the \( \bar{S} \)-bits of \( x \) are ignored.)

In the rest of this problem we establish the same kind of decomposition for general real-valued functions on product probability spaces. Specifically, let \( X \) be any finite set and let \( \pi \) be a probability distribution on \( X \). We think of the \( n \)-fold product set \( X^n \) as having the product probability distribution induced by \( \pi \). All \( \Pr[\cdot], \mathbb{E}[\cdot] \) in what follows refer to this product distribution.

(b) We first make property (iii) hold by fiat: For \( S \subseteq [n] \), we define \( f^\leq S : X^n \rightarrow \mathbb{R} \) to be the function depending only on the coordinates in \( S \) giving the conditional expectation; i.e., \( f^\leq S(x) := \mathbb{E}_x[f_{\bar{x} \rightarrow S}] \), where the expectation is over the product probability distribution on the coordinates outside \( S \). Now given this definition, explicitly write how we should define the functions \( f^S, f^T \) from (b), and then use (c).

(c) Show that \( \mathbb{E}_x[f^\leq S(x)f^\leq T(x)] = \mathbb{E}_x[f^\leq (S \cap T)(x)^2] \), straight from our definition of \( f^\leq S \).

(d) Now show property (ii), that \( \mathbb{E}_x[f^S(x)f^T(x)] = 0 \) when \( S \neq T \). (Hint: write your definitions of \( f^S, f^T \) from (b), and then use (c).)

Remark: This “orthogonal decomposition” of functions \( f \) is often a good substitute for Fourier analysis when the domain is a product probability space other than \( \{−1, 1\}^n \).

2. Logarithmic Sobolev Inequality. Consider the Hypercontractive Theorem with \( q = 2, p = 2 - 2\epsilon \), and \( \rho = \sqrt{1 - 2\epsilon} \), where \( \epsilon \in [0, 1/2] \); if we square it, we get

\[
\|T_{\sqrt{1 - 2\epsilon}} f\|_2^2 \leq \|f\|_2^{2 - 2\epsilon}
\]
for any \( f : \{-1, 1\}^n \to \mathbb{R} \).

(a) Show that we have equality at \( \epsilon = 0 \). Explain why we can now conclude that
\[
\frac{\partial}{\partial \epsilon} \| T_{\sqrt{1-2\epsilon}} f \|_2^2 \bigg|_{\epsilon=0} \leq \frac{\partial}{\partial \epsilon} \| f \|_{2-2\epsilon}^2 \bigg|_{\epsilon=0}.
\]
(b) Show that
\[
\frac{\partial}{\partial \epsilon} \| T_{\sqrt{1-2\epsilon}} f \|_2^2 \bigg|_{\epsilon=0} = -2 \mathbb{I}(f).
\]
(c) Show that
\[
\frac{\partial}{\partial \epsilon} \| f \|_{2-2\epsilon}^2 \bigg|_{\epsilon=0} = -\text{Ent}[f^2],
\]
where \( \text{Ent}[g] \) is the functional defined for nonnegative \( g \) by \( \text{Ent}[g] = E[g \ln g] - E[g] \ln E[g] \). \(^1\)

We conclude that for all \( f : \{-1, 1\}^n \to \mathbb{R} \),
\[
\text{Ent}[f^2] \leq 2 \mathbb{I}(f).
\]
This is called the “Logarithmic Sobolev Inequality”, or the “Entropy-Energy Inequality”. (Recall we called \( \mathbb{I}(f) \) the “energy” of \( f \) in Lecture 1.)

(d) Show that if \( f : \{-1, 1\}^n \to \{\text{T}, \text{F}\} \) has \( p = \Pr[f = \text{T}] \leq 1/2 \), then
\[
2p \ln(1/p) \leq \mathbb{I}(f).
\]
This significantly improves on the Poincaré Inequality \( 4p(1-p) \leq \mathbb{I}(f) \) for small \( p \).

3. \( \epsilon \)-biased sets. For every positive integer \( k \), there is a field \( \mathbb{F}_{2^k} \) with exactly \( 2^k \) elements. There is a natural way of encoding the names of the field elements as \( k \)-bit strings, \( \text{enc} : \mathbb{F}_{2^k} \to \{0,1\}^k \), and this encoding has the property that \( \text{enc}(x+y) = \text{enc}(x) + \text{enc}(y) \) for all \( x, y \in \mathbb{F}_{2^k} \) and also \( \text{enc}(0) = (0, \ldots, 0) \). Further, given \( \text{enc}(x) \) and \( \text{enc}(y) \), one can compute \( \text{enc}(xy), \text{enc}(x/y), \text{enc}(x+y), \text{enc}(x-y) \), in deterministic \( \text{poly}(k) \) time.\(^2\)

(a) Let \( R \) denote a random string in \( \mathbb{F}_{2^k}^n \), formed as follows: Pick \( a, b \in \mathbb{F}_{2^k} \) independently and uniformly at random; then let the \( i \)th bit of \( R \) be \( \langle \text{enc}(a^i), \text{enc}(b) \rangle_{\mathbb{F}_2} \), where \( \langle \cdot, \cdot \rangle_{\mathbb{F}_2} \) denotes dot product in \( \mathbb{F}_2 \). Show that for every nonzero string \( S \in \mathbb{F}_2^n \),
\[
\frac{1}{2} - \frac{1}{2} \cdot \frac{n}{2^k} \leq E[\langle R, S \rangle_{\mathbb{F}_2}] \leq \frac{1}{2},
\]
where in the expectation, we’re taking \( \langle R, S \rangle_{\mathbb{F}_2} \) (which is in \( \mathbb{F}_2 \)) and reinterpreting it as a real number. (Hint: every nonzero degree-\( n \) polynomial over a field has at most \( n \) zeroes.)

(b) As needed for Problem 4 on Homework 3, give efficiently constructible \( \epsilon \)-biased sets for \( \{-1, 1\}^n \) of size \((n/\epsilon)^2\), whenever \( n/\epsilon \) is a power of 2.

\(^1\)0 \ln 0 = 0.

\(^2\)Specifically, it is known that for every \( k \) there is an irreducible polynomial \( p(t) \in \mathbb{F}_2[t] \) of degree \( k \); then we may take \( \mathbb{F}_{2^k} \) to be the set of polynomials in \( \mathbb{F}_2[t] \) modulo \( p(t) \). The function \( \text{enc} \) maps \( \sum_{i=0}^{k-1} a_i t^i \) to \( (a_0, \ldots, a_{k-1}) \). It is known (Shoup, 1990) that one can deterministically find an irreducible \( p \) in time \( \text{poly}(k) \). (Also, it’s very easy to find one in time \( 2^{O(k)} \) which is pretty much good enough for us.)
4. Tightness of the Hypercontractivity Theorem. It is known\(^3\) that when \(q\) is an even positive integer, the largest possible ratio of \(\|f\|_q/\|f\|_2\) for degree-\(d\) real-valued boolean functions \(f\) is achieved, in the limit as \(n \to \infty\), by

\[
 f = \sum_{S \subseteq [n] \atop |S| = d} x_S. \tag{1}
\]

I don’t think there is a closed-form formula for the limiting ratio, but it is known to be \(\Theta_q(1) \cdot d^{-1/4} \cdot \sqrt{q - 1}^d\). In this problem we show a slightly weaker lower bound for our favorite case, \(q = 4\).

(a) Let \(f : \{-1, 1\}^n \to \mathbb{R}\) be the function in (1); assume that \(d\) is divisible by 3 and that \(n \geq 2d\). Show that

\[
 \mathbb{E}[f^4] \geq \frac{(d/3,d/3,d/3,d/3,d/3,n-2d)}{\binom{n}{d}^2} \mathbb{E}[f^2]^2,
\]

where the quantity in the numerator is a multinomial coefficient — specifically, the number of ways of choosing six disjoint \(d/3\)-size subsets of \([n]\). (Hint: given six disjoint \(d/3\)-size subsets, consider quadruples of \(d\)-size sets \(S\) that hit each \(d/3\)-set twice.)

(b) Using Stirling’s formula, show that

\[
 \lim_{n \to \infty} \left( \frac{\binom{n}{d}^2}{d/3,d/3,d/3,d/3,d/3,n-2d} \right) = \Theta(d^{-2} \cdot 9^d).
\]

(c) [Extra credit.] Problems (a) and (b) give a lower bound of \(\Omega(d^{-1/2} \cdot \sqrt{3}^d)\) for the largest possible ratio of \(\|f\|_4/\|f\|_2\) for degree-\(d\) \(f\). For extra credit, either: (i) explain Bonami’s original 1970 argument which gives an upper bound of \(O(d^{-1/8} \cdot \sqrt{3}^d)\); or (ii) show that the \(f\) in (1) indeed achieves \(\Theta(d^{-1/4} \cdot \sqrt{3}^d)\) in the limit as \(n \to \infty\). Or both.

5. Generalized Chernoff bounds. A Chernoff bound says that if \(t \geq 1\) and \(\sum_{i=1}^n a_i^2 = 1\), then

\[
 \mathbb{P}[|a_1 x_1 + \cdots + a_n x_n| \geq t] \leq \exp \left(-\Omega(t^2)\right),
\]

where the \(x_i\)’s denote i.i.d. random ±1 bits, as usual. Prove the following generalization: Let \(p(x_1, \ldots, x_n)\) be a multilinear polynomial over the reals of degree at most \(d\), and assume \(\mathbb{E}[p(x_1, \ldots, x_2)^2] = 1\) (i.e., the sum of squares of \(p\)’s coefficients is 1). Then for \(t \geq 1\),

\[
 \mathbb{P}[|p(x_1, \ldots, x_n)| \geq t] \leq \exp \left(-\Omega(t^{2/d})\right).
\]

(Hint: Markov plus \((2, q, (1/\sqrt{q - 1})^d)\)-hypercontractivity with large \(q\).)

\(^3\)Svante Janson has a published proof; I don’t know if it’s the earliest.

(a) Let \( f : \{-1, 1\}^n \rightarrow \{-1, 1\} \) be computable by a depth-\( d \) decision tree. Show that \( \sum_{i=1}^{n} \hat{f}(i) \leq O(\sqrt{d}) \). (Hint: mimic the proof that Majority maximizes \( \sum \hat{f}(i) \) for general \( f \); but take the expectation over a random path first.) Conclude that if \( f \) is monotone, \( I(f) \leq \sqrt{\text{DT-depth}(f)} \).

(b) Suppose one has access to random examples from a monotone function \( f \). Give a learning algorithm which on input \( \tau \), identifies (w.h.p.) a set \( J \) which contains all coordinates \( i \) with \( \text{Inf}_i(f) \geq \tau \). The algorithm should run in time \( \text{poly}(n, 1/\tau) \) and the set \( J \) identified should have size \( O(1/\tau^2) \).

(c) Show that \( C = \{ \text{monotone } f : \text{DT-depth}(f) \leq \log n \} \) is learnable from random examples only in time \( n^{O(1/\epsilon^2)} \). (Hint: use the Main Lemma that implied Friedgut’s Theorem.)