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.

**Notation used:**

- \( [n] \): the set \( \{1, 2, \ldots, n\} \)
- \( x^{(i)} \): the \( n \)-bit string \( x \) with its \( i \)th bit flipped, where \( i \in [n] \)
- \( x^{(i=b)} \): the \( n \)-bit string \( x \) with its \( i \)th bit set to \( b \)
- \( \Delta(x, y) \): the Hamming distance between \( x, y \in \{-1, 1\}^n \); i.e., \( |\{i : x_i \neq y_i\}| \)
- \( S \): always a subset of \( [n] \), unless otherwise specified
- \( \|f\|_2 \): = \( \sqrt{\langle f, f \rangle} = \sqrt{E[f^2]} \) when \( f : \{-1, 1\}^n \to \mathbb{R} \) is a function
- \( \|y\|_2 \): when \( y \in \mathbb{R}^n \) is a vector, the usual (Euclidean) length of \( y \); i.e., \( \sqrt{\sum_i y_i^2} \)
- \( f^{\text{odd}} \): when \( f : \{-1, 1\}^n \to \mathbb{R} \), denotes the function \( f^{\text{odd}}(x) = (f(x) - f(-x))/2 \)
- \( \text{Inf}_{i}^{f(\rho)}(f) \): = \( \sum_{S \ni \rho} \rho^{|S|-1} \hat{f}(S)^2 \) when \( f : \{-1, 1\}^n \to \mathbb{R} \) is a function
- \( \text{Inf}_i(f) \): = \( \text{Inf}_{i}^{(1)}(f) \)
- \( I(f) \): = \( \sum_{i=1}^n \text{Inf}_i(f) \) for any \( f : \{-1, 1\}^n \to \mathbb{R} \)
- \( \text{deg}(f) \): max\( \{|S| : \hat{f}(S) \neq 0\} \) for nonzero \( f : \{-1, 1\}^n \to \mathbb{R} \)
- \( \Pr_{x}, E_{x}, \text{etc.} \): always denotes Probability, Expectation, (Variance, Covariance, \ldots) with respect to the *uniform* probability distribution of \( x \) on its range, unless otherwise specified
1. Poincaré Inequality II. For $i \in [n]$, the $i$th derivative operator (AKA $i$th annihilation operator) $D_i$ on functions $f : \{-1, 1\}^n \to \mathbb{R}$ is defined by letting $D_i f : \{-1, 1\}^n \to \mathbb{R}$ be the function given by

$$(D_i f)(x) = \frac{f(x^{(i=1)}) - f(x^{(i=-1)})}{2}.$$ 

(a) Show that $D_i$ acts as the usual derivative with respect to $x_i$ by showing that the Fourier expansion of $D_i f$ is

$$(D_i f)(x) = \sum_{S \ni i} \hat{f}(S) x_{S \setminus \{i\}}.$$ 

Conclude that

$$\inf_i (f) = \|D_i f\|_2^2.$$ 

(b) The gradient of $f$, written $\nabla f : \{-1, 1\}^n \to \mathbb{R}^n$, is defined by $\nabla f = (D_1 f, D_2 f, \ldots, D_n f)$. Show that

$$\mathbb{E}_x [\|\nabla f(x)\|_2^2] = I(f).$$ 

(c) Express $\mathbb{E}_x [\|\nabla f(x)\|_2^2]$ and $\text{Var}[f]$ in terms of weighted sums of squared Fourier coefficients; then conclude that for all $f : \{-1, 1\}^n \to \mathbb{R}$,

$$\text{Var}[f] \leq \mathbb{E} [\|\nabla f\|_2^2].$$ \hspace{1cm} (1)

(d) Show that when $f$’s range is $\{-1, 1\}$, the above result generalizes Problem 1 from Problem Set 1 (when $T$ is treated as $-1$ and $F$ is treated as $1$).

Remark: In analysis, (1) is known as the Poincaré Inequality for the discrete cube.

2. Distortion lower bounds for $\ell_1 \to \ell_2$. Without going into what all the words mean, the discrete cube $\{-1, 1\}^n$ with the Hamming distance $\Delta(\cdot, \cdot)$ is an example of an $\ell_1$ metric space. For $D \geq 1$, we say that the discrete cube can be embedded into $\ell_2$ with distortion $D$ if there is a mapping $F : \{-1, 1\}^n \to \mathbb{R}^m$ for some $m \in \mathbb{N}$ such that:

- (a) ["no contraction"] $\|F(x) - F(y)\|_2 \geq \Delta(x, y)$ for all $x, y$, and
- (b) ["expansion at most $D$"] $\|F(x) - F(y)\|_2 \leq D \cdot \Delta(x, y)$ for all $x, y$.

In this exercise we will show that every embedding has distortion at least $\sqrt{n}$.

(a) Show that for any $f : \{-1, 1\}^n \to \mathbb{R}$,

$$\mathbb{E}_x [(f(x) - f(-x))^2] \leq \sum_{i=1}^n \mathbb{E}_x \left[ \left( f(x^{(i=1)}) - f(x^{(i=-1)}) \right)^2 \right],$$ \hspace{1cm} (2)

by showing that $\|f^{\text{odd}}\|_2^2 \leq I(f)$ and then expanding definitions. (In fact, $\|f^{\text{odd}}\|_2^2 \leq \text{Var}[f]$.)
(b) Suppose \( F : \{-1,1\}^n \to \mathbb{R}^m \), and write \( F(x) = (f_1(x), f_2(x), \ldots, f_m(x)) \) for functions \( f_i : \{-1,1\}^n \to \mathbb{R} \). By summing (2) over \( i = 1 \ldots m \), show that any \( F \) with no contraction must have expansion at least \( \sqrt{n} \).

Remark: Congratulations, you’ve proven the famous Enflo Bound\(^1\): Embedding \( \ell_1 \) metrics on \( N \) points into \( \ell_2 \) can require distortion \( \sqrt{\log N} \). A nearly matching upper bound of \( O(\sqrt{\log N \log \log N}) \) was proven only two years ago, via CS-theory methods.

3. Noise stability. For \( \rho \in [-1,1] \), the noise operator \( T_\rho \) (AKA Bonami-Beckner operator) on functions \( f : \{-1,1\}^n \to \mathbb{R} \) is defined by letting \( T_\rho f : \{-1,1\}^n \to \mathbb{R} \) be the function given by

\[
(T_\rho f)(x) = \mathbb{E}_{y \sim \rho^x} [f(y)],
\]

where the notation \( y \sim \rho^x \) means that each value \( y_i \) is chosen independently via

\[
y_i = \begin{cases} 
  x_i & \text{with probability } \frac{1}{2} + \frac{1}{2} \rho, \\
  -x_i & \text{with probability } \frac{1}{2} - \frac{1}{2} \rho.
\end{cases}
\]

Note that when \( \rho \geq 0 \) this is the same as saying \( y_i \) is set to \( x_i \) with probability \( \rho \) and is set uniformly at random with probability \( 1 - \rho \). We further define the noise stability of \( f \) and \( g \) at \( \rho \) to be

\[
S_\rho(f, g) = \langle f, T_\rho g \rangle = \mathbb{E}_x [f(x)(T_\rho g)(x)],
\]

and define \( S_\rho(f) = S_\rho(f, f) \) to be the noise stability of \( f \) at \( \rho \).

(a) Show that the Fourier expansion of \( T_\rho f \) is

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

Conclude that

\[
S_\rho(f, g) = \sum_{S \subseteq [n]} \rho^{|S|} \hat{f}(S) \hat{g}(S).
\]

(b) Let \( f : \{-1,1\}^n \to \{-1,1\} \) and let \( \epsilon \in [0,1] \). Define the noise sensitivity of \( f \) at \( \epsilon \) to be

\[
NS_\epsilon(f) = \mathbb{P}_{x,y} \left[ f(x) \neq f(y) \right],
\]

where \( x \) and \( y \) are chosen by first choosing \( x \) uniformly at random and then forming \( y \) by flipping each bit of \( x \) with probability \( \epsilon \). Show that

\[
NS_\epsilon(f) = \frac{1}{2} - \frac{1}{2} S_{1-2\epsilon}(f).
\]

(c) Show that for any \( f : \{-1,1\}^n \to \mathbb{R} \) we have \( \text{Inf}_t^{(\rho)}(f) = S_\rho(D_t f) \), and also \( \sum_{i=1}^n \text{Inf}_t^{(\rho)}(f) = \frac{\partial}{\partial \rho} S_\rho(f) \).

\(^1\)Per Enflo, On the nonexistence of uniform homeomorphisms between \( L_p \)-spaces, Arkiv för matematik 8, 1969.
4. Noam and Mario’s bound. Let \( f : \{-1, 1\}^n \rightarrow \mathbb{R} \) be a nonzero function with \( \deg(f) \leq d \).

(a) Generalize Problem 3(b) from Problem Set 1 by showing that \( \Pr_{x} [f(x) \neq 0] \geq 2^{-d} \). (Hint: same.)

(b) Show that if in addition \( f \) maps into \([-1, 1]\) then \( \|f\| \leq d \).

(c) Show that if in addition \( f \) is boolean-valued (maps into \([-1, 1]\)) then \( f \) is a \( d2^{d-1} \)-junta.

(d) The address function with \( k \) address bits is \( \text{Addr}_k : \{-1, 1\}^{k+2k} \rightarrow \{-1, 1\} \) defined by
\[
\text{Addr}_k(x_1, \ldots, x_k, y_1, \ldots, y_{2k}) = y_x,
\]
where \( x = (x_1, \ldots, x_k) \) is identified with a number in \( [2^k] \). Show that \( \deg(\text{Addr}_k) = k + 1 \). Conclude that the junta size in (c) must be at least \( 2^{d-1} + d - 1 \).

5. Quasirandomness implies low correlation with juntas.

(a) Recall that for \( f, g : \{-1, 1\}^n \rightarrow \mathbb{R} \), \( \text{Cov}[f, g] = \mathbb{E}_{x} [f(x)g(x)] - \mathbb{E}_{x} [f(x)] \mathbb{E}_{x} [g(x)] \). Give a formula for \( \text{Cov}[f, g] \) in terms of the Fourier coefficients of \( f \) and \( g \).

(b) Show that if \( h : \{-1, 1\}^n \rightarrow [-1, 1] \) is \((\epsilon, \delta)\)-quasirandom, then \( \text{Cov}[h, f] \leq \sqrt{r/(1-\delta)\sqrt{\epsilon}} \) for every \( r \)-junta \( f : \{-1, 1\}^n \rightarrow \{-1, 1\} \). Explain why this result is trivial if \( r \geq \ln(1/\epsilon)/\delta \).

(Hints: You may need: (a) Cauchy-Schwarz, \( \sum a_i b_i \leq \sqrt{\sum a_i^2} \sqrt{\sum b_i^2} \); (b) the inequality \((1 - x)^y \leq \exp(-xy)\) for \( 0 \leq x \leq 1 \), \( y \geq 0 \).)

6. PCPPs may as well use \( \text{Or}_3 \). Suppose a property \( \mathcal{P} \) of \( m \)-bit strings has PCPPs of length \( \ell(m) \). Show that it has PCPPs of length \( \text{poly}(\ell(m)) \) in which the tester makes 3 queries and then uses one of the 8 possible \( \text{Or}_3 \) predicates: \( v_{i1} \lor v_{i2} \lor v_{i3}, \quad \overline{v}_{i1} \lor v_{i2} \lor v_{i3}, \quad \ldots, \quad \overline{v}_{i1} \lor \overline{v}_{i2} \lor \overline{v}_{i3} \).

7. A hardness reduction that doesn’t work. Suppose we try the following alternate reduction in attempt to prove that the Unique Games Conjecture implies \( 1 - \eta \) vs. \( 1/2 + \eta \) hardness for Max-3Lin for every \( \eta > 0 \). Given a CSP \( \mathcal{G} = (V, E) \) over \([k]\) with unique constraints, we reduce it to a “3Lin” tester over \( (f_v : \{-1, 1\}^n \rightarrow \{-1, 1\})_{v \in V} \) as follows: the tester picks an edge \((v, w) \in E\) uniformly at random and then does the Håstad-Odd test on the collection \( \{f_v \circ \sigma_{v\rightarrow w}, f_w \circ \sigma_{v\rightarrow w}\} \) where \( \sigma_{v\rightarrow w} \) is the edge constraint on \((v, w)\). (Recall that \( \sigma_{v\rightarrow w} \) acts on strings \( x \in \{-1, 1\}^k \) by \( \sigma_{v\rightarrow w}(x) = y \), where \( y_j = x_{\sigma_{v\rightarrow w}^{-1}(j)} \).)

(a) Show that the first part of the proof works out even better: The reduction maps \( \mathcal{G} \) instances with \( \text{val}(\mathcal{G}) \geq 1 - \lambda \) into 3Lin CSPs with value at least \( 1 - \delta - \lambda \).

(b) Show that the second part of the proof can never work. Specifically, show that regardless of what \( \mathcal{G} \) is, the resulting 3Lin system has an assignment with value at least \( 5/8 - \delta/4 \).
**Bonus Problem:** Let $p(x_1, \ldots, x_n)$ be a multilinear polynomial over the reals of degree at most $d$. Let $X_1, \ldots, X_n$ be independent real random variables satisfying $\mathbf{E}[X_i] = 0$, $\mathbf{E}[X_i^2] = 1$, $\mathbf{E}[X_i^4] = 0$, $\mathbf{E}[X_i^6] \leq 9$ for each $i \in [n]$. (For example, $(X_1, \ldots, X_n)$ chosen uniformly at random from $\{-1, 1\}^n$ would be fine.) Show that $\mathbf{E}[p(X_1, \ldots, X_n)^4] \leq 9^d \mathbf{E}[p(X_1, \ldots, X_n)^2]^2$. (Hint: induction.)