Analysis of Boolean Functions - Solutions

March 21, 2018

1 The Fourier Expansion

Problem 1. Compute the Fourier expansions of the following functions:

(a) $\min_2 : \{\pm 1\}^2 \to \{\pm 1\}$ defined by $\min_2(x_1, x_2) = \min\{x_1, x_2\}$

(b) $\min_3 : \{\pm 1\}^3 \to \{\pm 1\}$ (defined similarly to $\min_2$).

(c) $1_{\{a\}} : \mathbb{F}_2^n \to \{0, 1\}$

(d) $1_{\varphi_{\{a\}}}$ where $\varphi_{\{a\}}$ is the density function on $\{a\}$.

(e) $1_{\varphi_{\{a,a+e_i\}}}$ where $\varphi_{\{a,a,+e_i\}}$ is the density function on $\{a, a, +e_i\}$.

(f) 1_{φ_ρ} where φ_ρ is the density function in which every coordinate chosen with mean ρ.

(g) The inner product modulus 2 of two n bit length strings $(-1)^{\langle x,y\rangle} : \mathbb{F}_2^{2n} \to \{\pm 1\}$.

(h) Equality function,

$$\mathrm{Equ}(x_1, \dots, x_n) = \begin{cases} 1 & x_1 = x_2 = \cdots = x_n \\ -1 & otherwise \end{cases}$$

where $x_i \in \{\pm 1\}$.

(i) Non-equality function, Equality function,

$$\mathrm{NAE}(x_1, \dots, x_n) = \begin{cases} -1 & x_1 = x_2 = \cdots = x_n \\ 1 & otherwise \end{cases}$$

where $x_i \in \{\pm 1\}$.

(j) Selection:

$$\mathrm{Sel}(x_1, x_2, x_3) = \begin{cases} x_2 & x_1 = -1 \\ x_3 & x_1 = 1 \end{cases}$$

(k) $\bmod_3 : \mathbb{F}_2^3 \to \{0,1\}$ defined by,

$$\bmod_3(x_1,x_2,x_3) = \begin{cases} 1 & \sum x_i = 0 \mod 3 \\ 0 & \sum x_i \neq 0 \mod 3 \end{cases}$$

(l) $OXR : \mathbb{F}_2^3 \to \{0,1\}$ defined by $OXR(x_1,x_2,x_3) = x_1 \vee (x_2 \oplus x_3)$.

(m) $sort : \{\pm 1\}^4 \to \{\pm 1\}$ defined by $sort(x_1,x_2,x_3,x_4) = 1$ if $x_1 \leq x_2 \leq x_3 \leq x_4$ and -1 otherwise.

(n) The hemi-icosahedron function $HI : \{\pm 1\}^6 \to \{\pm 1\}$ defined to be the number of facets labeled $(+1,+1,+1)$ in Figure 1.2, minus the number of facets labeled $(1,1,1)$ modulus 3.

(o) $\mathrm{Maj}_5 : \{\pm 1\}^5 \to \{\pm 1\}$,$\mathrm{Maj}_7 : \{\pm 1\}^7 \to \{\pm 1\}$.

(p) The complete quadratic function

Solution 1.

(a) $\min_2(x_1,x_2) = -\frac{1}{2} + \frac{1}{2}x_1 + \frac{1}{2}x_2 + \frac{1}{2}x_1x_2$

(b) $\min_3 = -\frac{3}{4} + \frac{1}{4}(x_1 + x_2 + x_3 + x_1x_2 + x_2x_3 + x_1x_3 + x_1x_2x_3)$.

(c) $1_{\{a\}}(x) = \sum_{\{i:a_i=0\}\subseteq S} (-1)^{\sum_{i\in S} x_i}$

(d) $1_{\varphi_{\{a\}}} = \sum_{\{i:a_i=0\}\subseteq S} 2^n (-1)^{\sum_{i\in S} x_i}$.

(e) $1_{\varphi_{\{a,a+e_i\}}} = \sum_{\{j:a_j=0 \wedge j\neq i\}\subseteq S} 2^n(-1)^{\sum_{i\in S} x_i} + 2^{n-1}(-1)^{\sum_{a_i=0} x_i}$.

(f) $1_{\varphi_\rho} = 2^n \prod_i \frac{1+\rho x_i}{2} = \sum_S \rho^{|S|} \prod_{i\in S} x_i$.

(g) $\mathrm{IP}(x,y) = \sum_{S,T} 2^{-n}(-1)^{|S\cap T|}(-1)^{\sum_{i\in S} x_i + \sum_{i\in T} y_i}$

(h) $\mathrm{Equ}(x_1,\ldots,x_n) = (-2)^{-n}\prod(x_i - 1) + 2^{-n}\prod(x_1+1) = \sum_S 2^{-n}((-1)^n(-1^{|S|}+1)\prod_{i\in S} x_i$.

(i) $\mathrm{NAE}(x_1,\ldots,x_n) = 1 - \mathrm{Equ}(x_1,\ldots,x_n)$.

(j) $\mathrm{Sel}(x_1,x_2,x_3) = \frac{x_1+1}{2}\cdot x_3 - \frac{(x_1-1)}{2}\cdot x_2$

(k) $\bmod_3(x_1,x_2,x_3) = \frac{1}{4}\left(1 + (-1)^{x_1+x_2} + (-1)^{x_2+x_3} + (-1)^{x_1+x_3}\right)$.

(l) $OXR(x_1,x_2,x_3) = 1 - \left(\frac{1+(-1)^{x_1}}{2}\right)\cdot\left(\frac{1+(-1)^{x_2+x_3}}{2}\right) = \frac{1}{4}\left(3 - (-1)^{x_1} - (-1)^{x_2+x_3} - (-1)^{x_1+x_2+x_3}\right)$.

(m) Note that in boolean values $x_i \leq x_{i+1}$ if either $x_i = x_{i+1}$ (equality) or $x_i = -1$, $x_{i+1} = 1$. Thus,

$$\frac{(x_i + x_{i+1})^2}{4} + \frac{(1 - x_i)(x_{i+1} + 1)}{4} = \begin{cases} 1 & x_i \leq x_{i+1} \\ 0 & otherwise \end{cases}$$

Thus the binary sorting function equality,

$$sort(x_1, \ldots, x_n) = \prod \frac{(x_i + x_{i+1})^2}{4} + \frac{(1 - x_i)(x_{i+1} + 1)}{4}$$

which can be translated to ± 1 by a simple transformation.

(n) The hemi-icosahedron function $HI : \{\pm 1\}^6 \to \{\pm 1\}$ defined to be the number of facets labeled $(+1, +1, +1)$ in Figure 1.2, minus the number of facets labeled $(1, 1, 1)$ modulus 3.

(o) $\mathrm{Maj}_5 : \{\pm 1\}^5 \to \{\pm 1\}$,$\mathrm{Maj}_7 : \{\pm 1\}^7 \to \{\pm 1\}$.

(p) The complete quadratic function

Problem 1.2. How many Boolean functions $f : \{1, 1\}^n \to \{1, 1\}$ have exactly 1 nonzero Fourier coefficient?

Solution 1.2. By Parseval, the magnitude of this coefficient must be 1 and so there are $2 \cdot 2^n$ such functions.

Problem 1.3. Let $f : \mathbb{F}_2^n \to \{0, 1\}$ and suppose $\#\{x : f(x) = 1\}$ is odd. Prove that all of f s Fourier coefficients are nonzero.

Solution 1.3. The Fourier coefficients are given by,

$$\widehat{f}(S) = \mathbb{E}_x \left[f(x) \chi_S(x) \right] = 2^{-n} \sum_{f(x)=1} (-1)^{\sum_{i \in S} x_i}$$

Note that the sum is taken over odd number of terms by assumption, all whom are ± 1, hence nonzero.

Problem 1.4. Let $f : \{1, 1\}^n \to \mathbb{R}$ have Fourier expansion $f(x) = \sum_S \hat{f}(S) x^S$. Let $F : \mathbb{R}^n \to \mathbb{R}$ be the extension of f which is also defined by $F(x) = \sum_S \hat{f}(S) x^S$. Show that if $\mu = (\mu_1, ..., \mu_n) \in [1, 1]^n$ then,

$$F(\mu) = \mathbb{E}_y \left[f(y) \right]$$

where y is the random string in $\{1, 1\}^n$ defined by having $\mathbb{E}_{y_i} [=] \mu_i$ independently for all i

Solution 1.4. Let φ be the density function of y. Then,

$$\mathbb{E}_y[f(y)] = \langle f, \varphi \rangle$$

The density function φ is given by $\varphi(x) = \prod_i (1 + x_i \mu_i)$ hence its Fourier expansion is,

$$\varphi(x) = \sum_S \mu^S x^S$$

and so by Parseval,

$$\mathbb{E}_y[f(y)] = \langle f, \varphi \rangle = \sum_S \widehat{f}(S)\widehat{\varphi}(S) = \sum_S \widehat{f}(S)\mu^S$$

which is what we wanted to show.

Problem 1.5. Prove that any $f : \{1, 1\}^n \to \{1, 1\}$ has at most one Fourier coefficient with magnitude exceeding $1/2$. Is this also true for any $f : \{1, 1\}^n \to \mathbb{R}$ with $||f||_2 = 1$?

Solution 1.5. For a boolean function, Fourier coefficients means correlation with characters. Specifically,

$$\widehat{S} = \Pr_x\left[f(x) = x^S\right] - \Pr_x\left[f(x) \neq x^S\right]$$

Thus, $\widehat{S} > 1/2$ means $\Pr_x\left[f(x) = x^S\right] > 3/4$. However, two different characters x^S, x^T agrees on exactly half the points. Thus, if $\Pr_x\left[f(x) = x^S\right], \Pr_x\left[f(x) = x^T\right] > 3/4$ then $\Pr_x\left[x^T = x^S\right] > 1/2$ contradiction. This is not true for function which are not boolean even when $||f||_2 = 1$. For instance, take arbitrary x^S, x^T, x^W distinct characters and set,

$$f(x) = \frac{x^S}{\sqrt{3}} + \frac{x^T}{\sqrt{3}} + \frac{x^W}{\sqrt{3}}$$

By construction $f(x)$ has 3 nonzero Fourier coefficients all with weight $\frac{1}{\sqrt{3}} > 1/2$ and still $||f||_2 = 1$.

Problem 1.6. Use Parsevals Theorem to prove uniqueness of the Fourier expansion.

Solution 1.6. Parsevals theorem yields,

$$\langle \sum_S a_S x^S, x^T \rangle = a_T$$

Problem 1.7. Let $f : \{1, 1\}^n \to \{1, 1\}$ be a random function. Show that for each $S \subseteq [n]$, the random variable $\widehat{f}(S)$ has mean 0 and variance 2^{-n}.

Solution 1.7. First note that $\widehat{f}(S)$ are equi-distributed since multiplying a function by x^T induces a permutation on all boolean functions and shifts the Fourier coefficients. Moreover, by symmetry $\mathbb{E}\left[\widehat{f}(S)\right] = 0$ (multiplying by -1 inverts the sign of $\widehat{f}(S)$ and permutes the set of boolean functions). For the variance use Parseval identity.

Problem 1.8. The (Boolean) dual of $f : \{1, 1\}^n \to \mathbb{R}$ is the function $f^\dagger$ defined by,

$$f^\dagger(x) = -f(-x)$$

The function f is said to be odd if it equals its dual. Equivalently, if $f(x) = f(x)$ for all x. The function f is said to be even if $f(x) = f(x)$ for all x. Given any function $f : \{1, 1\}^n \to \mathbb{R}$, its odd part is the function $f_{odd} : \{1, 1\}^n \to \mathbb{R}$ defined by $f_{odd}(x) = \frac{f(x)-f(-x)}{2}$, and its even part is the function $f_{even} : \{1, 1\}^n \to \mathbb{R}$ defined by $f_{even}(x) = \frac{f(x)+f(x)}{2}$.

(a) Calculate $\widehat{f^\dagger}(S)$ in terms of $\widehat{f}$.

(b) Verify that $f = f_{odd} + f_{even}$ and that f is odd (respectively, even) if and only if $f = f_{odd}$ (respectively, $f = f_{even}$).

(c) Show that,

$$f_{odd} = \sum_{|S| \text{ odd}} \widehat{f}(S)x^S \qquad f_{even} = \sum_{|S| \text{ even}} \widehat{f}(S)x^S$$

Solution 1.8. First note that,

$$f^\dagger(x) = -f(-x) = -\sum \widehat{f}(S)(-x)^S = \sum (-1)^{|S|+1} \widehat{f}(S)x^S$$

thus $\widehat{f^\dagger}(S) = (-1)^{|S|+1}\widehat{f}(S)$. The rest is straightforward calculations.

Problem 1.9. In this problem we consider representing False,True as $0, 1 \in \mathbb{R}$

(a) Using the interpolation method from Section 1.2, show that every $f : \{False, True\}^n \to \{False, True\}$ can be represented as a real multilinear polynomial

$$q(x) = \sum_S c_S \prod_{i \in S} x_i$$

over $\{0, 1\}$, meaning mapping $\{0, 1\}^n \to \{0, 1\}$.

(b) Show that this representation is unique.

(c) Show that all coefficients c_S in the representation will be integers in the range $[2^n, 2^n]$.

(d) Let $f : \{False, True\}^n \to \{False, True\}$. Let $p(x)$ be fs multilinear representation when False,True are $1, 1 \in \mathbb{R}$ (i.e., p is the Fourier expansion of f) and let $q(x)$ be fs multilinear representation when False,True are $0, 1 \in \mathbb{R}$. Show that,

$$q(x) = \frac{1}{2} - \frac{1}{2}p(1 - 2x_1, \dots, 1 - 2x_n)$$

Solution 1.9. The existence follows from the indicator functions,

$$\prod (x_i + a_i) \qquad a_i \in \{0, 1\}$$

which is 1 when $x = (a_1, \dots, a_n)$ and zero elsewhere. For uniqueness, note that it suffices to prove uniqueness for the zero function. Suppose,

$$0 = \sum_S c_S \prod_{i \in S} x_i$$

for any $x \in \{0,1\}^n$. Suppose not all c_W are zero and let $c_W \neq 0$ maximal w.r.t inclusion W (i.e there is no T s.t. $c_T \neq 0$ and $W \subsetneq T$). Substitute the value z with $z_i = 1$ for $i \in W$ and zero in all other coordinates then in particular,

$$0 = \sum_S c_S \prod_{i \in S} z_i = c_W$$

contradiction since $c_W \neq 0$ thus all c_S are zero. Every function is a sum of at most 2^n indicators and each indicator contributes either 1 or 0 to every product $\prod_{i \in S} x_i$ hence (c) follows. For (d) consider the transformation $\psi : \{0,1\} \to \{-1,1\}$,

$$\psi(x) = 1 - 2x$$

maps $\psi(0) = 1$, $\psi(1) = -1$. Note that $\psi^{-1}(t) = \frac{1}{2} - \frac{t}{2}$ and so by uniqueness

$$p(x) = \psi^{-1}(q(\psi(x)))$$

Problem 1.10. Let $f : \{1,1\}^n \to \mathbb{R}$ be not identically 0. The (real) degree of f , denoted $\deg(f)$, is defined to be the degree of its multilinear Fourier expansion.

(a) Show that $\deg(f) = \deg(a + bf)$ for any $a, b \in \mathbb{R}$ assuming $b \neq 0$, $a + bf \neq 0$.

(b) Show that $deg(f) \leq k$ if and only if f is a real linear combination of functions $g_1, \dots, g_s$, each of which depends on at most k input coordinates.

(c) Which functions in Exercise 1.1 have nontrivial degree (i.e $\deg(f) < n$)?

Solution 1.10. For (a) use that add a only change the zero layer of the Fourier expansion of f and multiplying by nonzero constant only multiplies every Fourier coefficient by a nonzero constant, and in particular the nonzero Fourier coefficients remain unchanged.

For (b), clearly if $\deg(f) \leq k$ then its Fourier expansion looks like,

$$f = \sum_{|S| \leq k} \widehat{f}(S) x^S$$

and each x^S depends on at most $|S|$ coordinates. For the other direction it suffice to prove that if a function depends only on k coordinates its degree is at most k. Let $g : \{1,1\}^n \to \mathbb{R}$ that depends only on some k coordinates and consider $S \subseteq [n]$ with $|S| > k$ then there exists $j \in [n]$ s.t. f does not depend on the j'th coodinate. In particular,

$$g(x) = g(x + e_j)$$

for any x. It follows that,

$$\widehat{g}(S) = \mathbb{E}_x \left[g(x) x^S \right] = \frac{1}{2} \mathbb{E}_x \left[g(x) x^{S \setminus \{j\}} : x_j = 1 \right] \frac{1}{2} \mathbb{E}_x \left[g(x) x^{S \setminus \{j\}} \cdot (-1) : x_j = -1 \right] = 0$$

hence $\widehat{g}(S) = 0$ for any $|S| > k$ and so $\deg(g) \leq k$.

Problem 1.11. Suppose that $f : \{1, 1\}^n \to \{1, 1\}$ has $\deg(f) = k \geq 1$.

(a) Show that fs real multilinear representation over $\{0, 1\}$, call it $q(x)$, also has $\deg(q) = k$.

(b) Show that $\widehat{f}(S)$ is an integer multiple of 2^{1-k}.

(c) Show that $\sum_S \left|\widehat{f}\right| \leq 2^{k-1}$

Solution 1.11. First recall that,

$$q(x) = \frac{1}{2} - \frac{1}{2} f(1 - 2x_1, \ldots, 1 - 2x_n)$$

so (a) follows immediately. Similarly we have,

$$f(x_1, \ldots, x_n) = 1 - 2q\left(\frac{1}{2} - \frac{x_1}{2}, \ldots, \frac{1}{2} - \frac{x_n}{2}\right)$$

and so (b) follows. We now prove (c). This type of inequality screams Cauchy-Schwartz. However a naive application yields,

$$\sum \left|\widehat{f}(S)\right| \leq \sqrt{2^n} \cdot \sum_S \widehat{f}^2 = \sqrt{2^n}$$

We may replace $\sqrt{2^n}$ with $\sqrt{N}$ where N is an upper bound on the number of nonzero Fourier coefficients. To bound N we use (b) which implies that,

$$\widehat{f}(S) = \frac{c_S}{2^{k-1}} \quad c_S \in \mathbb{Z}$$

Using that c_S is an integer with Parseval identity,

$$\sum_S \widehat{f}(S)^2 = \frac{c_S^2}{2^{2(k-1)}} = 1 \Rightarrow \sum_S c_S^2 = 2^{2(k-1)}$$

implies that only $2^{2(k-1)}$ coefficients can be nonzero. Plugging $N = 2^{2(k-1)}$ to the Cauchy Schwartz inequality above yields the result.

Problem 1.12. A Hadamard matrix is any $N \times N$ real matrix with ± 1 entries and orthogonal rows. Particular examples are the WalshHadamard matrices H_N, inductively defined for $N = 2^n$ as follows:

$$H_1 = [1] \qquad N_{2^n} = \begin{bmatrix} H_{2^{n-1}} & H_{2^{n-1}} \\ H_{2^{n-1}} & -H_{2^{n-1}} \end{bmatrix}$$

(a) Lets index the rows and columns of H_{2^n} by the integers $\{0, 1, 2, ..., 2^n - 1\}$ rather than $[2^n]$. Further, lets identify such an integer i with its binary expansion $(i_0, i_1, \ldots, i_{n-1}) \in \mathbb{F}_2^n$, where i_0 is the least significant bit and i_{n-1} the most. For example, if $n = 3$, we identify the index $i = 6$ with $(0, 1, 1)$. Now show that the (γ, x) entry of H_{2^n} is $(-1)^{\gamma \cdot x}$.

(b) Show that if $f : \mathbb{F}_2^n \to \mathbb{R}$ is represented as a column vector in $\mathbb{R}^{2^n}$ (according to the indexing scheme from part (a)) then $2^n H^{2^n} f = \widehat{f}$. Here we think of $\widehat{f}$ as also being a function $\mathbb{F}_2^n \to \mathbb{R}$, identifying subsets $S \subseteq \{0, 1, ..., n1\}$ with their indicator vectors.

(c) Show how to compute $H_{2^n} f$ using just $n2^n$ additions and subtractions (rather than 2^{2n} additions and subtractions as the usual matrix-vector multiplication algorithm would require). This computation is called the Fast WalshHadamard Transform and is the method of choice for computing the Fourier expansion of a generic function $f : \mathbb{F}_2^n \to \mathbb{R}$ when n is large.

(d) Show that taking the Fourier transform is essentially an involution: $\widehat{\widehat{f}} = 2^n f$ (using the notations from part (b)).

Solution 1.12. (a) Induction on n.

(b) Induction on n. Decompose f to $f(x_1, \dots, x_n) = f(x_1, \dots, x_{n-1}, 1)x_n + f(x_1, \dots, x_{n-1}, 0)(1-x_n)$ and apply induction hypothesis on $f(x_1, \dots, x_{n-1}, 1)$,$f(x_1, \dots, x_{n-1}, 0)$.

(c) Use a recursive algorithm. Suppose we want to compute $H_{2^{n+1}} f$ then it suffice to compute $H_{2^n} g$,$H_{2^n} h$ where $g = f(x_1, \dots, x_{n-1}, 0)$,$h = f(x_1, \dots, x_{n-1}, 1)$.

(d) Use that $H_{2^n}^2 = 2^{-n} I$ (follows from orthogonality).

Problem 1.13. Let $f : \{1, 1\}^n \to \mathbb{R}$ and let $0 < p \le q \le \infty$. Show that $||f||_p \le ||f||_q$.

Proof. The case $p \le q < \infty$ follows from Jensen inequality. The case $q = \infty$ follows from a limit argument. □

Problem 1.14. Compute the mean and variance of each function from Exercise 1.1.

Solution 1.14. Use the formula for variance given the Fourier coefficients.

Problem 1.15. Let $f : \{1, 1\}^n \to \mathbb{R}$. Let $K \subseteq [n]$ and let $z \in \{1, 1\}^K$. Suppose $g : 1, 1^{n-|K|} \to \mathbb{R}$ is the sub-function of f gotten by restricting the K-coordinates to be z. Show that $\mathbb{E}_x [g] = \sum_{T \subseteq [K]} \widehat{f}(T) z^T$.

Solution 1.15. Write $x = (y, z)$.

$$
\begin{aligned}
\mathbb{E}_x [g] &= 2^{-(n-|K|)} \sum_y f(y, z) \\
&= 2^{-(n-|K|)} \sum_y \sum_S \widehat{f}(S) x^S \\
&= 2^{-(n-|K|)} \sum_y \sum_S \widehat{f}(S) y^S z^S \\
&= 2^{-(n-|K|)} \sum_S \widehat{f}(S) z^S \sum_y y^S \\
&= \sum_{S \subseteq K} \widehat{f}(S) z^S
\end{aligned}
$$

where we used the simple equality,

$$\sum_y y^S = \begin{cases} 0 & S \not\subseteq K \\ 2^{-(n-|K|)} & S \subseteq K \end{cases}$$

Problem 1.16. If $f : \{1, 1\}^n \to \{1, 1\}$, show that $\mathrm{Var}[f] = 4 \cdot \mathrm{dist}(f, 1) \cdot \mathrm{dist}(f, -1)$.

Solution 1.16. The definition of variance is,

$$\mathrm{Var}[f] = \mathbb{E}_x\left[f^2(x)\right] - \mathbb{E}_x[f(x)]^2$$

For boolean functions $\mathbb{E}_{f^2}[=]1$ and so,

$$\mathrm{Var}[f] = (1 - \mathbb{E}_x[f(x)])(1 + \mathbb{E}_x[f(x)])$$

Again, for boolean functions,

$$\mathbb{E}_x[f(x)] = \mathrm{Pr}_x[f(x) = 1] - \mathrm{Pr}_x[f(x) = -1] = 2\mathrm{Pr}_x[f(x) = 1] - 1 = 1 - 2\mathrm{Pr}_x[f(x) = -1]$$

Substituting $\mathbb{E}_x[f(x)]$,

$$\mathrm{Var}[f] = (2\mathrm{Pr}_x[f(x) = 1]) \cdot (2\mathrm{Pr}_x[f(x) = -1]) = 4 \cdot \mathrm{dist}(f, 1) \cdot \mathrm{dist}(f, -1)$$

Problem 1.17. Prove that if F is a $\{1, 1\}$-valued random variable with mean μ then,

$$\mathrm{Var}[f] = \mathbb{E}\left[(F-\mu)^2\right] = \frac{1}{2}\mathbb{E}\left[(F-F')^2\right] = 2\mathrm{Pr}[F \neq F'] = \mathbb{E}[|F-\mu|]$$

where F' is an independent copy of F.

Solution 1.17. Simplifying $\mathbb{E}[(F-F')^2]$,

$$\mathbb{E}\left[(F-F')^2\right] = \mathbb{E}\left[F^2 - 2FF' + (F')^2\right] = \mathbb{E}\left[F^2\right] - 2\mathbb{E}[FF'] + \mathbb{E}\left[(F')^2\right]$$

Note $\mathbb{E}[FF'] = \mu^2$ and that $\mathbb{E}[(F')^2] = \mathbb{E}[F^2]$ so the second equality holds. The third equality follows since,

$$(F-F')^2 = \begin{cases} 0 & F = F' \\ 4 & F \neq F' \end{cases}$$

For the last inequality set $p = \Pr[F = 1]$,$\mu = \Pr[F = 1] - \Pr[F = -1] = 2p - 1$ then,

$$\begin{aligned}
\mathbb{E}[|F-\mu|] &= p\,|1-\mu| + (1-p)\,|1+\mu| \\
&= p\,(1-\mu) + (1-p)\,(1+\mu) \\
&= p - \mu p + (1-p) + (1-p)\mu \\
&= 1 - (1-2p)^2 \\
&= 4p(1-p) \\
&= 4\Pr[F = 1]\Pr[F = -1] \\
&= 2\Pr[F = 1 \wedge F' = -1] + 2\Pr[F' = 1 \wedge F = -1] \\
&= \Pr[F \neq F']
\end{aligned}$$

Problem 1.18. For any $f : \{1, 1\}^n \to \mathbb{R}$, show that

$$\langle f^{=k}, f^{=l} \rangle = \begin{cases} W^k[f] & l = k \\ 0 & l \neq k \end{cases}$$

Solution 1.18. Follows from orthogonality

Problem 1.19. Let $f : \{1, 1\}^n \to \{1, 1\}$. Then,

(a) Suppose that $W^1(f) = 1$. Show that $f(x) = \pm x^S$ for some $S \neq \emptyset$.

(b) Suppose that $W^{\leq 1}(f) = 1$. Show that f depends on at most one coordinate.

(c) Suppose that $W^{\leq 2}(f) = 1$. Must f depend on at most 2 input coordinates? At most 3 input coordinates? What if we assume $W^2(f) = 1$?

Solution 1.19. If $W^1(f) = 1$ then by Parseval,

$$\sum_i \widehat{f}(\{i\})^2 = 1$$

and also choosing $x_i = \text{sgn}\left(\widehat{f}(\{i\})\right)$ and substitute to the Fourier expansion we get,

$$\sum_i \left|\widehat{f}(\{i\})\right| = 1$$

Actually, the right should be ± 1 but as of all terms on the left are nonnegative it must be 1. Together this implies (1) since $\widehat{f}(\{i\}) = \pm 1$ for some i. If $W^{\leq 1}(f) = 1$ then similarly,

$$\widehat{f}(\emptyset)^2 + \sum_i \widehat{f}(\{i\})^2 = 1$$

$$\widehat{f}(\emptyset) + \sum_i \left|\widehat{f}(\{i\})\right| = \pm 1$$

If $\widehat{f}(\emptyset) \geq 0$ then as before either $\widehat{f}(\emptyset) = \pm 1$ or $\widehat{f}(\{i\}) = \pm 1$. If $\widehat{\emptyset} < 0$ we can do a similar trick by substituting $x_i = -\text{sgn}(\widehat{f}(\{i\}))$ and conclude the same. If $W^{\leq 2}(f) = 1$ the function need not to be dependent on 2 coordinates. For example,

$$f(x_1, x_2, x_3) = \frac{1}{2}x_1x_2 + \frac{1}{2}x_2x_3 - \frac{1}{2}x_1x_3 + \frac{1}{2}$$

Verify that this is indeed a boolean function. However, since $\deg(f) \leq 2$ the granularity of its Fourier coefficients is $\frac{1}{2}$ which leaves us exactly 4 Fourier coefficients (or zero, or one which are degenerate cases) all with magnitude $\frac{1}{2}$. Moreover, note that the nonzero Fourier coefficients must be for sets corresponding to linearly dependent vectors (over $\mathbb{F}_2$). This type of consideration leads to the conclude that the function must be dependent on at most 3 variables. The same argument shows that $W^2(f) = 1$ implies $f = \pm x_i x_j$ for some $i \neq j$.

Problem 1.20. Let $f : \{1, 1\}^n \to \mathbb{R}$ satisfy $f = f^{=1}$. Show that $\mathrm{Var}[f^2] = \sum_{i \neq j} \widehat{f}(\{i\})\widehat{f}(\{j\})$.

Solution 1.20. The variance is the sum of nonzero Fourier coefficients. The Fourier expansion of f^2 is,

$$
\begin{aligned}
f^2 &= \left(\sum_S \widehat{f}(S)x^S\right)^2 \\
&= \sum_i \widehat{f}(\{i\})^2 + \sum_{i \neq j} 2\widehat{f}(\{i\})\widehat{f}(\{j\})x_i x_j \\
&= 1 + \sum_{i \neq j} 2\widehat{f}(\{i\})\widehat{f}(\{j\})x_i x_j
\end{aligned}
$$

and the result follows from the formula for the variance.

Problem 1.21. Prove that there are no functions $f : \{1, 1\}^2 \to \{1, 1\}$ with exactly 2 nonzero Fourier coefficients. What about exactly 3 nonzero Fourier coefficients?

Solution 1.21. Impossible. Suppose that $f(x) = Ax_S + Bx_T$. Since the function is boolean $A^2 + B^2 = 1$. Moreover, we can force x_S, x_T taking any sign we like hence $|A| + |B| = 1$ and so either $A = 1, B = 0$ or $A = 0, B = 1$. For 3 nonzero coefficients the same argument of dependency can be applied and so the only possibility is,

$$f(x) = Ax_S + Bx_T + Cx_{S\Delta T}$$

Assume C is positive and force $x_S = \mathrm{sgn}(A)$,$x_T = \mathrm{sgn}(B)$ we get $f(x) = |A| + |B| + C = 1$ and so either $A = 1$, $B = 1$, $C = 1$. If $C < 0$ force $x_S = -\mathrm{sgn}(A)$,$x_T = -\mathrm{sgn}(B)$.

Another argument is to use granularity and the following nice lemma.

Lemma 1.1. *Let A, B, C nonzero then $A^2 + B^2 + C^2$ is never a power of 2.*

Problem 1.22. Verify Propositions 1.25 and 1.26.

Solution 1.22. Use that $2^{-n}\varphi(y) = \Pr_{Y\sim\varphi}[Y = y]$.

Problem 1.23. Let φ and ψ be probability densities on $\mathbb{F}_2^n$.

(a) Show that the total variation distance between φ and ψ defined by,

$$\Delta(\varphi, \psi) = \max_A \{\Pr_{y\sim\varphi}[y \in A] - \Pr_{y\sim\psi}[y \in A]\}$$

equals $||\varphi - \psi||_1$.

(b) Show that the collision probability of φ, defined by,

$$\Pr_{y,y'\sim\varphi}[y = y']$$

equals $\frac{||\varphi||_2^2}{2^n}$.

(c) The χ_2-distance of φ and ψ defined by,

$$d_{\chi_2}(\varphi, \psi) = \mathbb{E}_{y \sim \psi}\left[\left(\frac{\varphi(y)}{\psi(y)} - 1\right)^2\right]$$

assuming ψ has full support. Show that the χ_2-distance of φ from uniform is equal to $\mathrm{Var}[\varphi]$.

(d) Show that the total variation distance of from uniform is at most $\frac{1}{2}\sqrt{\mathrm{Var}[\varphi]}$.

Solution 1.23. (a) Consider the "most separating event" defined by,

$$A = \{a : \Pr_{Y\sim\varphi}[Y = a] \geq \Pr_{Y\sim\psi}[Y = a]\}$$

First that,

$$\begin{aligned}
||\varphi - \psi||_1 &= \sum_y |\varphi(y) - \psi(y)| \\
&= 2^{-n} \sum_{\varphi(y)\geq\psi(y)} (\varphi(y) - \psi(y)) + \sum_{\varphi(y)<\psi(y)} (\psi(y) - \varphi(y)) \\
&= \sum_{\varphi(y)\geq\psi(y)} \varphi(y) - \sum_{\varphi(y)\geq\psi(y)} \psi(y) + \sum_{\varphi(y)<\psi(y)} \psi(y) - \sum_{\varphi(y)<\psi(y)} \varphi(y) \\
&= \Pr_{y\sim\varphi}[y \in A] - \Pr_{y\sim\psi}[y \in A] + \Pr_{y\sim\psi}[y \notin A] - \Pr_{y\sim\varphi}[y \notin A] \\
&= \Pr_{y\sim\varphi}[y \in A] - \Pr_{y\sim\psi}[y \in A] + (1 - \Pr_{y\sim\psi}[y \in A]) - (1 - \Pr_{y\sim\varphi}[y \in A]) \\
&= 2\Pr_{y\sim\varphi}[y \in A] - 2\Pr_{y\sim\psi}[y \in A]
\end{aligned}$$

which proves that $\frac{1}{2}||\varphi - \psi||_1 \leq \Delta(\varphi, \psi)$. For the converse use the triangle inequality. Suppose that the maximum event is A' then,

$$\begin{aligned}
\Delta(\varphi, \psi) &= \sum_{a\in A'} \varphi(a) - \sum_{a\in A'} \psi(a) \\
&= \frac{1}{2}\sum_{a\in A'} \varphi(a) - \frac{1}{2}\sum_{a\in A'} \psi(a) + \frac{1}{2}\sum_{a\in A'} \varphi(a) - \frac{1}{2}\sum_{a\in A'} \psi(a) \\
&= \frac{1}{2}\sum_{a\in A'} \varphi(a) - \frac{1}{2}\left(1 - \sum_{a\notin A'} \psi(a)\right) + \frac{1}{2}\left(1 - \sum_{a\notin A'} \varphi(a)\right) - \frac{1}{2}\sum_{a\in A'} \psi(a) \\
&= \frac{1}{2}\sum_{a\in A'} (\varphi(a) - \psi(a)) + \frac{1}{2}\sum_{a\notin A'} (\psi(a) - \varphi(a)) \\
&\leq \frac{1}{2}\sum_x |\varphi(x) - \psi(x)|
\end{aligned}$$

and so equality holds.

(b)

$$\Pr_{y,y'\sim\varphi}[y = y'] = \sum_x \Pr_{y\sim\varphi}[\varphi(y) = x]^2 = \sum_x \varphi(x)^2$$

(c) This amounts to showing,

$$\mathbb{E}_{y\sim\psi}\left[(\varphi(y) - 1)^2\right] = \mathrm{Var}[\varphi]$$

Follows immediately since $\mathbb{E}_\varphi[=]\,1$.

(d) This amounts to showing,

$$||\varphi - U_n||_1 \leq \sqrt{\mathrm{Var}[\varphi]}$$

where $U_n(x) = 1$ for any x. Equivalently,

$$\mathbb{E}_x[|\varphi(x) - 1|] \leq \sqrt{\mathbb{E}_x[(\varphi(x) - 1)^2]}$$

which follows from Cauchy-Schwartz.

Problem 1.24. Let $A \subseteq 1, 1^n$ have volume δ, meaning $\mathbb{E}_x[1_A(x)] = \delta$. Suppose φ is a probability density supported on A. Show that $||\varphi||_2^2 \leq \frac{1}{\delta}$ with equality if φ is the uniform density on A.

Solution 1.24. Set $\varphi_A = \delta^{-1}1_A$ density function on A. Since φ is supported on A then,

$$\langle \varphi, \varphi_A \rangle = \delta^{-1}\langle \varphi, 1_A \rangle = \delta^{-1}$$

By Cauchy-Schwartz,

$$\frac{1}{\delta} = \langle \varphi, \varphi_A \rangle \leq ||\varphi||_2^2 ||\varphi_A||_2^2 = ||\varphi||_2^2$$

with equality iff $\varphi = \varphi_A$.

Problem 1.19. Show directly from the definition that the convolution operator is associative and commutative.

Solution 1.25. First we prove it is commutative,

$$\begin{aligned}
f * (g * h)(x) &= \mathbb{E}_y[f(x)g * h(x + y)] \\
&= \mathbb{E}_y[f(x)\mathbb{E}_z[g(x + y)h(x + y + z)]] \\
&= \mathbb{E}_{y,z}[f(x)g(x + y)h(x + y + z)] \\
&= \mathbb{E}_{y,z}[f(x)g(x + y)h(x + z)] \\
&= \mathbb{E}_z[\mathbb{E}_y[f(x)g(x + y)]\,h(x + z)] \\
&= (f * g) * h(x)
\end{aligned}$$

For commutative it follows by simple change of variables.

Problem 1.26. Prove that the following are equivalent for functions $f : \mathbb{F}_2^n \to \mathbb{F}_2$:

1. $f(x+y) = f(x) + f(y)$.
2. $f(x) = a \cdot x$ for some $a \in \mathbb{F}_2^n$.

Solution 1.26. (2) implies (1) easily. To see (1) implies (2) note that it follows from (1) that,

$$f(\sum \epsilon_i e_i) = \sum \epsilon_i f(e_i)$$

for any $\epsilon_i \in \mathbb{F}_2$ hence (2) holds with $a_i = f(e_i)$.

Problem 1.27. Suppose an algorithm is given query access to a linear function $f : \mathbb{F}_2^n \to \mathbb{F}_2$ and its task is to determine which linear function f is. Show that querying f on n inputs is necessary and sufficient.

Solution 1.27. Query on a basis. Clearly f is determined iff that values on all basis elements are determined (for any arbitrary fixed basis).

Problem 1.28. Improve the BLR theorem as follows.

(a) Let $f : \mathbb{F}_2^n \to \{1, 1\}$ and suppose that $\operatorname{dist}(f, \chi_{S'}) = \delta$. Show that $\left|\widehat{f}(S)\right| \leq 2\delta$ for all $S \neq S'$

(b) Deduce that the BLR Test rejects f with probability at least $3\delta - 10\delta^2 + 8\delta^3$.

(c) Show that this lower bound cannot be improved to $c\delta - O(\delta^2)$ for any $c > 3$.

Solution 1.28.

(a) Recall that $\operatorname{dist}(f, \chi_{S'}) = \Pr_x [f(x) \neq \chi_{S'}(x)]$ and, Let $S \neq S'$ arbitrary and recall that $\Pr_x [\chi_S(x) = \chi_{S'}(x)] = \frac{1}{2}$. By union bound,

$$\Pr_x [f(x) = \chi_S(x)] \leq \Pr_x [\chi_S(x) = \chi_{S'}(x)] + \Pr_x [f \neq \chi_{S'}(x)] = \frac{1}{2} + \delta$$

Recall that,

$$\widehat{f}(S) = 2\Pr_x [f(x) = \chi_{S'}(x)] - 1 \leq 2\delta$$

To obtain a bound on $\left|\widehat{f}(S)\right|$ do the same union bound but now using $-\chi_{S'}$.

(b) The success probability of the BLR test is given by,

$$\frac{1}{2} + \frac{1}{2} \sum_S \widehat{f}^3(S)$$

Estimate the Fourier coefficients as follows,

$$\begin{aligned}\sum_S \widehat{f}^3(S) &= \widehat{f}^3(S) + \sum_{S\neq S'} \widehat{f}^3(S) \\ &\leq (1-2\delta)^3 + 2\delta \sum_{S\neq S'} \widehat{f}^2 \\ &= 1 - 6\delta + 12\delta^2 - 8\delta^3 + 2\delta(1-(1-2\delta)^2) \\ &= 1 - 6\delta + 12\delta^2 - 8\delta^3 + 8\delta^2 - 8\delta^3 \\ &= 1 - 6\delta + 20\delta^2 - 16\delta^3\end{aligned}$$

which concludes the result.

(c) Consider g some arbitrary linear function and set B of density δ,

$$f(x) = \begin{cases} g(x) & x \notin B \\ -g(x) & x \in B \end{cases}$$

Note that $\operatorname{dist}(f, g) = \delta$. The BLR test rejects if the test hit odd number of times the bad set B (one or three) which amounts to $3\delta(1-\delta)^2 + \delta^3 = 3\delta + O(\delta^2)$.

Problem 1.29. We call $f : \mathbb{F}_2^n \to \mathbb{F}_2$ an affine function if $f(x) = a \cdot x + b$ for some $a \in \mathbb{F}_2^n$, $b \in \mathbb{F}_2$.

(a) Show that f is affine if and only if $f(x)+f(y)+f(z) = f(x+y+z)$ for all $x, y, z, \in \mathbb{F}_n^2$

(b) Let $f : \mathbb{F}_2^n \to \mathbb{R}$. Suppose we choose $x, y, z \sim \mathbb{F}_2^n$ independently and uniformly. Show that,

$$\mathbb{E}_{x,y,z}\left[f(x)f(y)f(z)f(x+y+z)\right] = \sum_S \widehat{f}(S)^4$$

(c) Give a 4-query test for a function $f : \mathbb{F}_2^n \to \mathbb{F}_2$ with the following property: if the test accepts with probability 1ϵ then f is ϵ-close to being affine. All four query inputs should have the uniform distribution on $\mathbb{F}_2^n$ (but of course need not be independent).

(d) Give an alternate 4-query test for being affine in which three of the query inputs are uniformly distributed and the fourth is not random.

Solution 1.29. Let f affine.

(a) If f is affine then clearly the formula holds. If the formula holds then,

$$f(x) + f(y) + f(0) = f(x+y)$$

hence $f(x) + f(0)$ is linear and so $f(x) + f(0) = a \cdot x$ for some a.

(b)

$$\begin{aligned}
\mathbb{E}_{x,y,z}\left[f(x)f(y)f(z)f(x+y+z)\right] &= \mathbb{E}_x\left[f(x)\mathbb{E}_y\left[f(y)\mathbb{E}_z\left[f(z)f(x+y+z)\right]\right]\right] \\
&= \mathbb{E}_x\left[f(x)\mathbb{E}_y\left[f(y)f*f(x+y)\right]\right] \\
&= \mathbb{E}_x\left[f(x)f*f*f(y)\right] \\
&= \langle f, f*f*f\rangle \\
&= \sum_S \widehat{f}(S)\widehat{f*f*f}(S) \\
&= \sum_S \widehat{f}(S)^4
\end{aligned}$$

(c) Set $g(x) = (-1)^{f(x)}$. Test,

$$g(x)g(y)g(z) = g(x+y+z)$$

Moreover, $g(x)g(y)g(z)g(x+y+z)$ is equivalent to $f(x)+f(y)+f(z) = f(x+y+z)$. Therefore, the success probability of passing the test is,

$$\begin{aligned}
\frac{1}{2}+\frac{1}{2}\mathbb{E}_{x,y,z}\left[g(x)g(y)g(z)g(x+y+z)\right] &= \frac{1}{2}+\frac{1}{2}\sum_S \widehat{g}^4(S) \\
&\leq \frac{1}{2}+\frac{1}{2}\max_S \widehat{g}^2(S) \\
&\leq \frac{1}{2}+\frac{1}{2}\max_S |\widehat{g}(S)|
\end{aligned}$$

This implies that,

$$1-2\epsilon \leq |\widehat{g}(S)|$$

Recall that $\widehat{g}(S)$ measures the correlation with χ_S and so g is either $\Pr_x\left[g(x)=\chi_S\right] \geq 1-\epsilon$ or $\Pr_x\left[g(x)=-\chi_S\right] \geq 1-\epsilon$ which is equivalent to either $\Pr_x\left[f(x)=(-1)^{\sum_{i\in S} x_i}\right] \geq 1-\epsilon$ or $\Pr_x\left[f(x)=(-1)^{1+\sum_{i\in S} x_i}\right] \geq 1-\epsilon$.

(d) Test,

$$f(x)+f(y)+f(0) = f(x+y)$$

Problem 1.30. Permutations $\pi \in S_n$ act on strings $x \in \{1,1\}^n$ in the natural way: $(x^\pi)_i = x(\pi(i))$. They also act on functions $f : \{1,1\}^n \to \mathbb{R}$ via $f^\pi(x) = f(x^\pi)$ for all $x \in \{-1,1\}^n$. We say that functions $g, h : \{1,1\}^n \to \{1,1\}$ are (permutation) isomorphic if $g = h^\pi$ for some $\pi \in S_n$. We call $\mathrm{Aut}\,(f) = \{\pi \in S_n : f^\pi = f\}$ the (permutation-)automorphism group of f.

(a) Prove that $\widehat{f^\pi}(S) = \widehat{f}(\pi^{-1}(S))$.

(b) Define a canonical form for $\mathrm{Aut}\,(f)$ as follows... Show that this is well defined.

(c) Show how to compute canonical(f) in $\tilde{O}(2^n)$ given that $\widehat{f}(\{i\})$ are distinct for $1 \leq i \leq n$.

(d) We could more generally consider $g, h : 1, 1^n \to \{1, 1\}$ to be isomorphic if,

$$g(x) = h(\pm x_{\pi(1)}, \ldots, \pm x_{\pi(n)})$$

for some permutation π on $[n]$ and some choice of signs. Extend the results of this exercise to handle this definition.

Solution 1.30. For (a) note that,

$$\begin{aligned}
\widehat{f^\pi}(S) &= \mathbb{E}_x \left[f^\pi(x)\chi_S(x)\right] \\
&= \mathbb{E}_x \left[f(x_{\pi(1)}, \ldots, x_{\pi(n)})\chi_S(x)\right] \\
&= \mathbb{E}_x \left[f(x_1, \ldots, x_n)\chi_{\pi^{-1}(S)}(x)\right] \\
&= \mathbb{E}_x \left[f(x)\chi_{\pi^{-1}(S)}\right] \\
&= \widehat{f}(\pi^{-1}(S))
\end{aligned}$$

The canonical form follows since it can be described equivalently as follows. Arrange all subsets of $[n]$ in the following linear order:

1. Sets of size k come before sets of size $k + 1$.
2. Sets of the same size are ordered according to their lexicographic order.

Consider all possible functions in $\mathrm{Aut}\,(f)$ and take the function with largest Fourier sequence w.r.t the above linear order. That is, first compare the first Fourier coefficient of $\emptyset$, then compare the Fourier coefficient of $\{1\}, \ldots, \{n\}$ in that order etc. This is well defined and moreover canonical since it is defined via $\mathrm{Aut}\,(f)$. Assuming $\widehat{f}(\{i\})$ are distinct the canonical form can be computed as follows:

1. Sort $\widehat{f}(\{i\})$,
$$\widehat{f}(\{i_1\}) < \widehat{f}(\{i_2\}) < \cdots \widehat{f}(\{i_n\})$$
2. Define $\pi(j) = i_j$.
3. Output f^π.

Taking a minus sign only changes the sign of the Fourier coefficients as follows. Suppose,

$$g(x) = h(\epsilon_1 x_{\pi(1)}, \ldots, \epsilon_n x_{\pi(n)})$$

with $\epsilon_i = \pm 1$. Then,

$$\widehat{g}(S) = \widehat{h}(\pi^{-1}(S)) \prod_{i \in S} \epsilon_i$$

The definition of canonical form remains the same with the minor change that we allow to consider permutations with sign changes. The algorithm also remains the same with the following changes,

1. Sort $\left|\widehat{f}(S)\right|$.
2. Output the permutation as before with sign as the sign of the corresponding Fourier coefficient.

2 Social Choice

Problem 1. For each function in Exercise 1.1, determine if it is odd, transitive-symmetric, and/or symmetric.

Solution 1. (a) Not odd, not symmetric, not transitive-symmetric.

(b) Not odd, not symmetric, not transitive-symmetric.

(c) Odd, not symmetric, not transitive-symmetric.

(d) Odd, not symmetric, not transitive-symmetric.

(e) Not odd, not symmetric, not transitive-symmetric.

(f) Not odd, symmetric, transitive-symmetric.

(g) Not odd, not symmetric, transitive-symmetric.

(h) Not odd, symmetric, transitive-symmetric.

(i) Not odd, symmetric, transitive-symmetric.

(j) Not odd, not symmetric, not transitive-symmetric.

(k) Not odd, symmetric, transitive-symmetric.

(l) Not odd, not symmetric, not transitive-symmetric.

(m) Not odd (flipping constant vector doe not change the value), not symmetric, not transitive-symmetric.

(n) ?

(o) Odd, symmetric, transitive-symmetric.

(p) ?

Problem 2. Show that the n-bit functions majority, AND, OR, $\pm\chi_S$, and ± 1 are all linear threshold functions.

Solution 2.

$$\begin{aligned}
AND(x_1, \ldots, x_n) &= \operatorname{sgn}(n - 0.5 + x_1 + x_2 + \cdots + x_n) \\
OR(x_1, \ldots, x_n) &= \operatorname{sgn}(n - 0.5 - x_1 - x_2 - \cdots - x_n) \\
\chi_i(x_1, \ldots, x_n) &= \operatorname{sgn}(x_i) \\
\pm 1(x_1, \ldots, x_n) &= \operatorname{sgn}(\pm 1)
\end{aligned}$$

Problem 3. Prove Mays Theorem:

(a) Show that $f : \{1,1\}^n \to \{1,1\}$ is symmetric and monotone if and only if it can be expressed as a weighted majority with $a_1 = a_2 = \cdots = a_n = 1$.

(b) Suppose $f : \{1,1\}^n \to \{1,1\}$ is symmetric, monotone, and odd. Show that n must be odd, and that $f = \mathrm{Maj}_n$.

Solution 3. (a) If a function is symmetric it is a function of $\sum_{i=1}^n x_i$. Assuming that the function is monotone it suffice to verify,

$$\sum_{i=1}^{n} x_i > T$$

for some $T \in \mathbb{R}$ by considering the minimal value of $\sum x_i$ s.t. $f(x) = 1$. If a function is of that form, clearly it is symmetric and monotone.

(b) Let f symmetric, odd and monotone then by the above f is of the form,

$$\mathrm{sgn}\left(\sum x_i\right) > T$$

If $n = 2k$ then consider the input z with exactly k ones (and k minus ones). Since that f is odd then $f(z) = -f(-z)$, however $-z$ has also k ones thus by symmetry $f(z) = f(-z)$ contradiction. We now have to prove that $-1 < T < 1$. To see this denote $n = 2k + 1$ and consider the input z with exactly $k + 1$ ones, $-z$ has k ones and so by oddity and monotonicity $f(z) = 1$, $f(-z) = -1$. This implies $-1 < T < 1$.

Problem 4. Subset $A \subseteq \{1,1\}^n$ is called a Hamming ball if $A = \{x : \Delta(x, z) < r\}$ for some $z \in \{1,1\}^n$ and real r. Show that $f : \{1,1\}^n \to \{-1,1\}$ is the indicator of a Hamming ball if and only if its expressible as a linear threshold function,

$$f(x) = \mathrm{sgn}(a_0 + a_1x_1 + \cdots + a_nx_n)$$

with $|a_1| = |a_2| = \cdots = |a_n|$.

Solution 4. Every Hamming ball centered at a with radius r can be described via the follows linear threshold,

$$\sum a_i x_i > n - 2r$$

Conversely, normalize to get $|a_i| = 1$ (if all $a_i = 0$ then the set is either empty or the entire boolean cube) and so,

$$\sum a_i x_i > a_0$$

which is equivalent to the above. Note that the sum when taken over $a_i = \pm 1$ equals the number of coordinates s.t. $x_i = a_i$ minus the number of coordinates s.t. $x_i \neq a_i$.

Problem 5. Let $f : \{1,1\}^n \to \{-1,1\}$ and $i \in [n]$. We say that f is unate in the i'th direction if either $f(x^{(i\to-1)}) \leq f(x^{(i\to1)})$ for all x (monotone in the i'th direction) or $f(x^{(i\to1)}) \leq f(x^{(i\to-1)})$ for all x (antimonotone in the i'th direction). We say that f is unate if it is unate in all n directions.

(a) Show that $\left|\widehat{f}(\{i\})\right| \leq \mathrm{Inf}_i\left[f\right]$ with equality if and only if f is unate in the i'th direction.

(b) Show that the second statement of Theorem 2.33 holds even for all unate f.

Solution 5. Note that $\frac{\left|f(x^{(i\to 1)})-f(x^{(i\to -1)})\right|}{2}$ is the indicator for $f(x^{(i\to 1)}) = f(x^{(i\to -1)})$. Hence,

$$\begin{aligned}
\left|\widehat{f}(\{i\})\right| &= \left|\mathbb{E}_x\left[f(x)x_i\right]\right| \\
&= \left|\frac{1}{2}\cdot \mathbb{E}_x\left[f(x^{(i\to 1)}) - f(x^{(i\to -1)})\right]\right| \\
&\leq \mathbb{E}_x\left[\frac{\left|f(x^{(i\to 1)}) - f(x^{(i\to -1)})\right|}{2}\right] \\
&= \mathrm{Inf}_i\left[f\right]
\end{aligned}$$

Equality holds iff the equality holds in the triangle inequality, which holds iff all summands have the same sign. That is either $f(x^{(i\to 1)})-f(x^{(i\to -1)}) \geq 0$ or $f(x^{(i\to 1)}) \leq f(x^{(i\to -1)})$, i.e f is unate. It follows that for unate functions,

$$\begin{aligned}
\mathrm{Inf}\left[f\right] &= \sum_{i=1}^{n} \left|\widehat{f}(\{i\})\right| \\
&= \sum_{i=1}^{n} \mathrm{sgn}\left(\widehat{f}(\{i\})\right)\widehat{f}(\{i\}) \\
&= \sum_{i=1}^{n} \mathrm{sgn}\left(\widehat{f}(\{i\})\right)\mathbb{E}_x\left[f(x)x_i\right] \\
&= \mathbb{E}_x\left[f(x)x_i\sum_{i=1}^{n}\mathrm{sgn}\left(\widehat{f}(\{i\})\right)\right] \\
&\leq \mathbb{E}_x\left[\left|x_1 + x_2 + \cdots + x_n\right|\right] \\
&= \sqrt{\frac{2n}{\pi}} + O(n^{-1/2}) \\
&= \mathrm{Inf}\left[\mathrm{Maj}_n\right] + O(n^{-1/2})
\end{aligned}$$

Problem 6. Show that linear threshold functions are unate.

Solution 6. Linear threshold functions are of the form,

$$\sum a_i > N$$

If $a_i \geq$ the function is monotone in the i'th direction, otherwise it is antimonotone in the i'th direction.

Problem 7. For each function f in Exercise 1.1, compute $\mathrm{Inf}^1\left[f\right]$.

Solution 7. Use formula for influence.

Problem 8. Let $f : \{1, 1\}^n \to \{1, 1\}$. Show that $\text{Inf}_i[f] \le \text{Var}[f]$ for each $i \in [n]$.

Solution 8.

$$\begin{aligned}\text{Inf}_i[f] &= \sum_{i \in S} \widehat{f}^2(S) \\ &\le \sum_{S \neq \emptyset} \widehat{f}^2(S)\end{aligned}$$

with equality iff all nonzero Fourier coefficients contain x_i.

Problem 9. Let $f : \{0, 1\}^6 \to \{1, 1\}$ be given by the weighted majority,

$$f(x) = \text{sgn}\,(58 + 31x_1 + 31x_2 + 28x_3 + 21x_4 + 2x_5 + 2x_6)$$

Compute $\text{Inf}_i[f]$ for all $i \in [6]$.

Solution 9.

$$\begin{aligned}\text{Inf}_1[f] &= \frac{9}{32} \\ \text{Inf}_2[f] &= \frac{9}{32} \\ \text{Inf}_3[f] &= \frac{9}{32} \\ \text{Inf}_4[f] &= \frac{7}{32} \\ \text{Inf}_5[f] &= \frac{1}{32} \\ \text{Inf}_6[f] &= \frac{1}{32}\end{aligned}$$

Problem 10. Say that coordinate i is b-pivotal for $f : \{1, 1\}^n \to \{-1, 1\}$ on input x (for $b \in \{-1, 1\}$) if $f(x) = b$ and $f(x \oplus i) \neq b$. Show that,

$$\text{Pr}_x[i \text{ is } b\text{-pivotal on } x] = \frac{1}{2}\text{Inf}_i[f]$$

. Deduce that $\text{Inf}[f] = 2\mathbb{E}_x[\text{number of } b\text{-pivotal coordinates on } x]$.

Solution 10. The value x contributes to $\text{Inf}_i[f]$ if $f(x) \neq f(x \oplus e_i)$. Note that such x's come in pairs, x and $x \oplus e_i$. However, fixing b pairs together x,$x \oplus e_i$. Formally,

$$\begin{aligned}\text{Pr}_x[i \text{ is } b\text{-pivotal on } x] &= 2^{-n}\left|\{x : f(x) = b, f(x \oplus e_i) \neq b\}\right| \\ &= 2^{-n} \cdot \frac{1}{2} \cdot \left|\{x : f(x) \neq f(x \oplus e_i)\}\right| \\ &= \frac{1}{2}\text{Inf}_i[f]\end{aligned}$$

the corollary follows immediately.

Problem 11. Let $f : \{1,1\}^n \to \{-1,1\}$ and suppose $\widehat{f}(S) \neq 0$. Show that each coordinate $i \in S$ is relevant for f.

Solution 11. $\widehat{f}(S) \neq 0$ implies $\mathrm{Inf}_i[f] \neq 0$ and so there exists x such that $f(x) \neq f(x \oplus e_i)$.

Problem 12. Let $f : \{1,1\}^n \to \{-1,1\}$ be a random function (uniformly at random). Compute $\mathbb{E}_f[\mathrm{Inf}_1[f]]$,$\mathbb{E}_f[\mathrm{Inf}[f]]$.

Solution 12. By linearity,

$$\mathbb{E}_f[\mathrm{Inf}_1[f]] = \sum_{x, x\oplus e_1} \mathbb{E}_f\left[1_{f(x)=f(x+\oplus e_1)}\right]$$

where we sum over the 2^{n-1} possible pairs. Using that f is random we get $\mathbb{E}_f\left[1_{f(x)=f(x+\oplus e_1)}\right] = \frac{1}{2}$ and so,

$$\mathbb{E}_f[\mathrm{Inf}_1[f]] = 2^{n-2}$$

By linearity,

$$\mathbb{E}_f[\mathrm{Inf}[f]] = n2^{n-2}$$

Problem 13. Let $w \in \mathbb{N}$, $n = w2^w$, and write f for $\mathrm{Tribe}_{w,2^w}, 2w : \{1,1\}^n \to 1, 1$.

(a) Compute $\mathbb{E}[f]$,$\mathrm{Var}[f]$ and estimate them asymptotically in terms of n.

(b) Describe the function $D_1 f$.

(c) Compute $\mathrm{Inf}[f]$,$\mathrm{Inf}_1[f]$ and estimate them asymptotically in terms of n.

Solution 13. Let $w \in \mathbb{N}$, $n = w2^w$, and write f for $\mathrm{Tribe}_{w,2^w}, 2w : \{1,1\}^n \to 1, 1$.

(a) Let X_i be the indicator of the i'th tribe. Since the votes are independent then $\mathbb{E}[X_i] = 2^{-w}$. Note that,

$$\mathrm{Tribe}_{w,2^w} = 1 - \prod_i (1 - X_i)$$

Since the tribes are independent,

$$\mathbb{E}[f] = 1 - \prod_i (1 - \mathbb{E}[X_i]) = 1 - (1 - 2^{-w})^{2^w}$$

Repeating the analysis with the easy observation that $\mathrm{Var}[X_i] = 2^{-w}(1 - 2^{-w})$ then,

$$\mathrm{Var}[f] = 1 - \prod_i (1 - \mathbb{E}[X_i]) = 1 - (1 - 2^{-w}(1 - 2^{-w}))^{2^w}$$

Letting n to infinity we get,

$$\mathbb{E}[f], \mathrm{Var}[f] \to 1 - \frac{1}{e}$$

(b)

$$D_1 f(x) \begin{cases} 2 & \text{all } x_i \text{ in the first tribe are 1 } (i \neq 1) \text{ and all other tribes are not 1} \\ 0 & otherwise \end{cases}$$

(c)

$$\begin{aligned} \text{Inf}_1\left[f\right] &= \frac{1}{2}\mathbb{E}\left[D_1 f(x)\right] \\ &= \frac{(2^w-1)^{(2^w-1)}}{2^n} \\ &= \frac{(2^w-1)^{(2^w-1)}}{2^{w2^w}} \\ &= \frac{(2^w)^{2^w}}{2^{w2^w}} \cdot \frac{(2^w-1)^{(2^w-1)}}{(2^w)^{2^w}} \\ &= \frac{(2^w-1)^{2^w}}{(2^w)^{2^w}} \cdot \frac{1}{2^w-1} \\ &= \left(1-2^{-w}\right)^{2^w} \cdot \frac{1}{2^w-1} \end{aligned}$$

and so by symmetry $\text{Inf}\left[f\right] = w2^w \text{Inf}_1\left[f\right]$. Taking $w \to \infty$

$$\text{Inf}_1\left[f\right] \to \frac{1}{e(2^w-1)}$$

$$\text{Inf}\left[f\right] \to \frac{w}{e}$$

Problem 14. Let $f : \{-1,1\}^n \to \mathbb{R}$. Show that $|D_i|f|| \leq |D_i f|$ pointwise. Deduce that $\text{Inf}_i\left[|f|\right] \leq \text{Inf}_i\left[f\right]$ and $\text{Inf}\left[|f|\right] \leq \text{Inf}\left[f\right]$.

Solution 14. The inequality,

$$|D_i|f|| \leq |D_i f|$$

follows from the triangle inequality,

$$||x| - |y|| \leq |x - y|$$

The second inequality follows by the relation $\text{Inf}_i\left[f\right] = \frac{1}{4}\mathbb{E}_x\left[D_i f^2(x)\right]$. The last inequality follows since the total influence is just the sum of influences.

Problem 15. The i'th expectation operator E_i is the linear operator on functions $f : \{1,1\}^n \to \mathbb{R}$ defined by,

$$E_i f(x) = \mathbb{E}_b\left[f(x_1, \ldots, x_{i-1}, b, x_{i+1}, \ldots, x_n)\right]$$

Show that:

(a) $E_i f(x) = \frac{f(x^{i\to 1})+f(x^{i\to -1})}{2}$.

(b) $E_i f(x) = \sum_{i\notin S} \widehat{f}(S)$.

(c) $f(x) = x_i D_i f(x) + E_i f(x)$.

Solution 15. Use the total-expectation to conclude that,

$$E_i f(x) = \frac{f(x^{i\to 1}) + f(x^{i\to -1})}{2}$$

Writing $f = \sum_S \widehat{f}\chi_S$ note that the Fourier coefficients for S s.t. $i \in S$ cancels hence,

$$E_i f(x) = \sum_{i\notin S} \widehat{f}(S)$$

This yields,

$$x_i D_i f + E_i f(x) = x_i \sum i \in S \widehat{f}(S) x^{S\setminus\{i\}} + \sum_{i\notin S} \widehat{f}(S) = \sum_S \widehat{f}(S) = f$$

Problem 16. The i'th expectation operator E_i is the linear operator on functions $f : \{1,1\}^n \to \mathbb{R}$ defined by,

$$L_i f = f - E_i f$$

Show that:

(a) $L_i f(x) = \frac{f(x)-f(x^{\oplus i})}{2}$.

(b) $L_i f(x) = \sum_{i\in S} \widehat{f}(S)$.

(c) $\langle f, L_i f\rangle = \langle L_i f, L_i f\rangle$

Solution 16. Note that,

$$L_i f(x) = f(x) - E_i f(x) = f(x) - \frac{f(x^{i\to 1}) + f(x^{i\to -1})}{2} = \frac{f(x) - f(x^{\oplus i})}{2}$$

Using the Fourier expansion for $E_i f$ we get,

$$L_i f(x) = \sum_{i\in S} \widehat{f}(S)$$

Therefore,

$$\begin{aligned}\langle f, L_i f\rangle = \langle \sum_S \widehat{f} x^S, \sum_{i\in S} \widehat{f} x^S\rangle \\ &= \sum_{i\in S} \widehat{f}(S)^2 \\ &= \langle L_i f, L_i f\rangle\end{aligned}$$

Problem 17. The i'th expectation operator E_i is the linear operator on functions $f : \{1,1\}^n \to \mathbb{R}$ defined by,

$$Lfi(x) = \sum_i L_i f$$

Show that:

(a) $Lf(x) = (n/2)(f(x) - \text{avg}_i f(x^{\oplus i}))$.

(b) Assuming $f : \{-1,1\}^n \to \{-1,1\}$ is boolean then $Lf(x) = f(x)\text{sens}_f(x)$.

(c) $Lf = \sum_S |S| \widehat{f}(S)\chi_S$.

Solution 17. Note that,

$$\begin{aligned} Lf(x) &= \sum_i L_i f(x) \\ &= \sum_i \frac{f(x) + f(x^{\oplus i})}{2} \\ &= (n/2)f(x) + \frac{1}{2}\sum_i f(x^{\oplus i} \\ &= (n/2)(f(x) - \text{avg}_i f(x^{\oplus i})) \end{aligned}$$

Recall that $\text{sens}_f(x)$ equals the number of pivotal coordinates for boolean functions. If i is pivotal in x then $L_i f = 2f(x)$ and otherwise $L_i f = 0$ so,

$$Lf(x) = \sum_i \frac{f(x) - f(x^{\oplus i})}{2} = f(x)\text{sens}_f(x)$$

The Fourier expansion is given by,

$$\begin{aligned} Lf &= \sum_i L_i f \\ &= \sum_i \sum_{i \in S} \widehat{f}(S)x^S \\ &= \sum_{i \in S} |S| \widehat{f}(S)x^S \end{aligned}$$

Problem 18. Let $f : \{-1,1\}^n \to \mathbb{R}$. Show that,

$$Lf = \frac{\mathrm{d}}{\mathrm{d}\rho} T_\rho f\Big|_{\rho=1} = -\frac{\mathrm{d}}{\mathrm{d}t} T_{e^{-t}} f\Big|_{t=0}$$

Solution 18. Substitute $\rho = 1$ in the following identity,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\rho}T_\rho f\big|_{\rho=1} &= \frac{\mathrm{d}}{\mathrm{d}\rho}\sum_S \rho^{|S|}\widehat{f}(S)x^S\big|_{\rho=1}\\ &= \sum_S |S|\,\rho^{|S|-1}\widehat{f}(S)x^S\end{aligned}$$

and so,

$$\frac{\mathrm{d}}{\mathrm{d}\rho}T_\rho f\big|_{\rho=1} = \sum_S |S|\,\widehat{f}(S)x^S = Lf$$

For the second equality,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}T_\rho f\big|_{t=0} &= \frac{\mathrm{d}}{\mathrm{d}\rho}\sum_S e^{-t|S|}\widehat{f}(S)x^S\big|_{\rho=0}\\ &= -\sum_S |S|\,e^{-t|S|}\widehat{f}(S)x^S\end{aligned}$$

Substituting $t = 0$,

$$-\frac{\mathrm{d}}{\mathrm{d}t}T_\rho f\big|_{t=0} = \sum_S |S|\,\widehat{f}(S)x^S = Lf$$

Problem 19. Suppose $f, g : \{-1,1\}^n \to \mathbb{R}$ have the property that f does not depend on the i'th coordinate and g does not depend on the j'th coordinate ($i \neq j$). Show that $\mathbb{E}_x\left[x_i x_j f(x)g(x)\right] = \mathbb{E}_x\left[D_i f(x) D_j g(x)\right]$.

Solution 19. By Parseval and that $\widehat{D_i f}(S) = \widehat{f}(S \setminus \{i\})$,

$$\begin{aligned}\mathbb{E}_x\left[D_i f(x) D_j g(x)\right] &= \langle D_i f, D_j g\rangle\\ &= \sum_S \widehat{D_i f}(S)\widehat{D_i g}(S)\\ &= \sum_S \widehat{f}(S\setminus\{i\})\widehat{g}(S\setminus\{j\})\\ &= \sum_S \widehat{f}(S)\widehat{g}(S)\end{aligned}$$

where we used that if f is not dependent on i then $\widehat{f}(S) = 0$ if $i \in S$. Note that the Fourier expansions of $x_i f(x)$, $x_j g(x)$,

$$x_i f = \sum_S \widehat{f}(S)x^{S\cup\{i\}} = \sum_{i\in S}\widehat{f}(S\setminus\{i\})x^S$$

$$x_j g = \sum_S \widehat{g}(S) x^{S\cup\{j\}} = \sum_{j\in S} \widehat{g}(S \setminus \{j\}) x^S$$

By Parseval,

$$\begin{aligned}
\mathbb{E}_x \left[x_i x_j f(x) g(x)\right] &= \langle x_i f, x_j g\rangle \\
&= \sum_S \widehat{x_i f}(S) \widehat{x_j g}(S) \\
&= \sum_{i,j\in S} \widehat{f}(S\setminus\{i\})\widehat{g}(S\setminus\{j\}) \\
&= \sum_S \widehat{f}(S)\widehat{g}(S)
\end{aligned}$$

Problem 20. For $f : \{-1,1\}^n \to \{-1,1\}$ we have that $\mathbb{E}_x\left[\mathrm{sens}_f(x)\right] = \mathbb{E}_{S\sim\mathcal{S}_f}\left[|S|\right]$. Show that also $\mathbb{E}_x\left[\mathrm{sens}_f(x)^2\right] = \mathbb{E}_{S\sim\mathcal{S}_f}\left[|S|^2\right]$. Is it true that $\mathbb{E}_x\left[\mathrm{sens}_f(x)^3\right] = \mathbb{E}_{S\sim\mathcal{S}_f}\left[|S|^3\right]$?

Solution 20. Using the Laplacian operator,

$$\langle Lf, Lf\rangle = \mathbb{E}_x\left[f(x)^2 \mathrm{sens}_f(x)^2\right] = \mathbb{E}_x\left[\mathrm{sens}_f(x)^2\right]$$

Recall that the Fourier expansion of the Laplacian is,

$$Lf = \sum_S |S|\, \widehat{f}(S) x^S$$

By Parseval,

$$\mathbb{E}_x\left[\mathrm{sens}_f(x)^2\right] = \sum_S |S|^2\, \widehat{f}(S)^2$$

This is true in the cubic version, e.g for $\min(x_1, x_2, x_3)$.

Problem 21. Let $f : \{-1,1\}^n \to \mathbb{R}$ and $i \in [n]$.

(a) Define,

$$\mathrm{Var}_i f(x) = \mathrm{Var}_b[f(x_1,\ldots,x_{i-1},b,x_{i+1},\ldots,x_n)]$$

Show that $\mathrm{Inf}_i\left[f\right] = \mathbb{E}_x\left[\mathrm{Var}_i f(x)\right]$.

(b) Show that,

$$\mathrm{Inf}_i\left[f\right] = \frac{1}{2}\mathbb{E}_{b,b'}\left[||f|_b - f|_{b'}||_2\right]$$

with $f|b$ means $n-1$ variables and fixing the i'th variable to be b.

Solution 21. For boolean functions both assertion are immediate since it is not hard to see that,

$$\mathrm{Var}_i f(x), \frac{1}{2}\mathbb{E}_{b,b'}\left[(f|_b(x) - f|_{b'}(x))^2\right] = \begin{cases} 1 & i \text{ is pivotal for } x \\ 0 & i \text{ is not pivotal for } x \end{cases}$$

In general, we need to use the definition of influence,

$$\mathrm{Inf}_i[f] = \mathbb{E}_x\left[\frac{1}{4}D_i f(x)^2\right]$$

Consider the random variable that attains the values p_1, p_2 uniformly (each with probability $1/2$). Its variance is given by $\frac{1}{4}(p_1 - p_2)^2$. Therefore,

$$\mathrm{Var}_i f(x) = \frac{1}{4}\left(f(x) - f(x^{\oplus i})\right)^2 = \frac{1}{4}D_i f(x)^2$$

Taking the expectation over x yields the result. For (b) change the order of summation,

$$\begin{aligned}
\frac{1}{2}\mathbb{E}_{b,b'}\left[||f|_b - f|_{b'}||_2\right] &= \frac{1}{2}\mathbb{E}_{b,b'}\left[\mathbb{E}_x\left[f|_b(x) - f|_{b'}(x)\right]\right] \\
&= \mathbb{E}_x\left[\frac{1}{2}\mathbb{E}_{b,b'}\left[(f|_b(x) - f|_{b'}(x))^2\right]\right] \\
&= \mathbb{E}_x\left[\frac{1}{2}\left(\frac{1}{4}(f(x) - f(x^{\oplus i}))^2 + \frac{1}{4}(f(x) - f(x^{\oplus i}))^2 + \frac{1}{4}\cdot 0 + \frac{1}{4}\cdot 0\right)\right] \\
&= \mathbb{E}_x\left[\frac{1}{4}(f(x) - f(x^{\oplus i}))^2\right] \\
&= \mathbb{E}_x\left[\frac{1}{4}D_i f(x)^2\right]
\end{aligned}$$

Problem 22. This exercise is concerned with the influence of the majority function.

(a) Show that $\mathrm{Inf}_i[\mathrm{Maj}_n] = 2^{1-n}\binom{n-1}{\frac{n-1}{2}}$

(b) $\mathrm{Inf}_1[\mathrm{Maj}_n]$ is a decreasing function of n.

(c) Show that $\mathrm{Inf}_i[\mathrm{Maj}_n] = \sqrt{\frac{2}{\pi n}} + O(n^{-3/2})$.

(d) Deduce that,

$$\frac{2}{\pi} \leq W^1[\mathrm{Maj}_n] \leq \frac{2}{\pi} + O(n^{-1})$$

(e) Deduce that,

$$\sqrt{\frac{2n}{\pi}} \leq \mathrm{Inf}[\mathrm{Maj}_n] \leq \sqrt{\frac{2n}{\pi}} + O(n^{-1/2})$$

(f) Suppose n is even and $f : \{-1,1\}^n \to \{-1,1\}$ is a majority function. Show that $\mathrm{Inf}[f] = \mathrm{Inf}\left[\mathrm{Maj}_{n-1}\right]$.

Solution 22.

(a) The probability that a single bit can flip the value of the majority function is only if exactly $\frac{n-1}{2}$ are ± 1. There are $2 \cdot \binom{n-1}{\frac{n-1}{2}}$ hence the probability for such x is given by $2^{1-n}\binom{n-1}{\frac{n-1}{2}}$.

(b) We shall show that the quotient is greater than 1

$$\frac{\mathrm{Inf}_1\left[\mathrm{Maj}_{n-2}\right]}{\mathrm{Inf}_1\left[\mathrm{Maj}_n\right]} > 1$$

for all n odd.

$$\begin{aligned}
\frac{\mathrm{Inf}_1\left[\mathrm{Maj}_{n-2}\right]}{\mathrm{Inf}_1\left[\mathrm{Maj}_n\right]} &= \frac{2^{1-(n-2)}\binom{n-3}{\frac{n-3}{2}}}{2^{1-n}\binom{n-1}{\frac{n-1}{2}}} \\
&= 4 \cdot \frac{\frac{(n-3)!}{\left(\frac{n-3}{2}\right)!\left(\frac{n-3}{2}\right)!}}{\frac{(n-1)!}{\left(\frac{n-1}{2}\right)!\left(\frac{n-1}{2}\right)!}} \\
&= 4 \cdot \frac{\frac{n-1}{2} \cdot \frac{n-1}{2}}{(n-1)(n-2)} \\
&= \frac{n-1}{n-2} \\
&> 1
\end{aligned}$$

(c) Use Stirling approximation (tighter bounds, See Wikipedia),

$$\sqrt{2\pi} n^{n+\frac{1}{2}} e^{-n} e^{\frac{1}{12nn+1}} \leq n! \leq \sqrt{2\pi} n^{n+\frac{1}{2}} e^{-n} e^{\frac{1}{12n+1}}$$

and in particular,

$$n! = \left(\sqrt{2\pi n} + O(n^{-1/2})\right)\left(\frac{n}{e}\right)^n$$

Therefore,

$$\begin{aligned}
\mathrm{Inf}_1\left[f\right] &= 2^{1-n}\binom{n-1}{\frac{n-1}{2}} \\
&= 2^{1-n}\frac{n!}{\left(\frac{n-1}{2}\right)!\left(\frac{n-1}{2}\right)!} \\
&= 2^{1-n}\frac{(n-1)^{n-1}}{\left(\frac{n-1}{2}\right)^{n-1}} \cdot \frac{\sqrt{2\pi(n-1)} + O(n^{-1/2})}{\left(\sqrt{\pi(n-1)} + O(n^{-1/2})\right)^2} \\
&= \frac{\sqrt{2\pi n} + O(n^{-1/2})}{\sqrt{\pi n} + O(1)} \\
&= \sqrt{\frac{2}{\pi n}} + O(n^{-3/2})
\end{aligned}$$

Using the tighter version we obtain,

$$
\begin{aligned}
\mathrm{Inf}_1[f] &= 2^{1-n}\binom{n-1}{\frac{n-1}{2}} \\
&= 2^{1-n}\frac{n!}{\left(\frac{n-1}{2}\right)!\left(\frac{n-1}{2}\right)!} \\
&\leq \frac{\sqrt{2\pi(n-1)}}{\pi(n-1)} \cdot e^{\frac{1}{12n+1}-\frac{2}{12n}} \\
&\leq \sqrt{\frac{2}{\pi(n-1)}} \cdot \left(1 - O\left(\frac{1}{n}\right)\right) \\
&\leq \sqrt{\frac{2}{\pi n}}
\end{aligned}
$$

(d) Recall that for monotone functions $\mathrm{Inf}_i[f] = \widehat{f}(\{i\})$ thus squaring the previous bounds and multiplying by n yield the result.

(e) Recall that for monotone functions the total influence is just the sum of degree one Fourier coefficients and so the inequalities follows by multiplying by n the bounds for $\mathrm{Inf}_1[f]$.

(f) We will show that $\widehat{f}1 = \widehat{\mathrm{Maj}_{n-1}}1$ and since f is monotone the result follows.

$$
\begin{aligned}
\widehat{f}1 &= \mathbb{E}_x[f(x)(x_1+x_2+\cdots+x_n)] \\
&= \frac{1}{2}\mathbb{E}_x[f(x)(x_1+x_2+\cdots+x_{n-1}+1)] + \frac{1}{2}\mathbb{E}_x[f(x)(x_1+x_2+\cdots+x_{n-1}-1)] \\
&= \frac{1}{2}\mathbb{E}_x\left[\mathrm{Maj}_{n-1}(x_1,\ldots,x_{n-1})(x_1+\cdots+x_{n-1}+1)\right] + \frac{1}{2}\mathbb{E}_x[f(x_1,\ldots,x_{n-1})(x_1+\cdots+x_{n-1}-1)] \\
&= \mathbb{E}_x[f(x_1,\ldots,x_{n-1})(x_1+\cdots+x_{n-1})] \\
&= \widehat{\mathrm{Maj}_{n-1}}1
\end{aligned}
$$

The only non-trivial equality is where we replace $f(x)$ with the majority on $x_1, \ldots, x_{n-1}$. To see consider the two possible cases:

- If $\sum_{i=1}^{n} x_i \neq 0$ it means that x_n does not effect the input hence the result is anyway the majority of the first $n-1$ variables.
- If $\sum_{i=1}^{n} x_i = 0$ then the equality still holds since we multiply $f(x)$ by zero hence if we replace $f(x)$ when any arbitrary value the equality still holds.

Problem 23. Using only CauchySchwarz and Parseval, give a very simple proof of the following: If $f : \{-1,1\}^n \to \{-1,1\}$ is monotone then $\mathrm{Inf}[f] \leq \sqrt{n}$. Extend also to the case of f unate.

Solution 23. Using Parseval $W^1[f] \leq 1$. If f is monotone then $\sum \widehat{f}(\{i\}) = \mathrm{Inf}[f]$ and so the inequality from applying Parseval to the vector $(\widehat{f}(\{i\}))$. In the case f is unate we have the inequality $\left|\widehat{f}(\{i\})\right| = \mathrm{Inf}_i[f]$ and so,

$$\mathrm{Inf}[f] = \sum \left|\widehat{f}(\{i\})\right|$$

and the result follows form applying Cauchy-Schwartz to the vector $(\left|\widehat{f}(\{i\})\right|)$.

Problem 24. Let $f : \{-1,1\}^n \to \{-1,1\}$ s.t. all $\widehat{f}(\{i\})$ are equal. Prove that,

$$W^1[f] \leq 2/\pi + O(n^{-1})$$

Solution 24. Recall that Maj_n maximizes $\sum_{i=1}^n \widehat{f}(\{i\})$. Thus,

$$\sum_{i=1}^n \widehat{f}(\{i\}) \leq \sum_{i=1}^n \widehat{\mathrm{Maj}_n}(\{i\}) = \mathrm{Inf}[\mathrm{Maj}_n] = \sqrt{\frac{2n}{\pi}} + O(n^{-1/2})$$

Set $t = \widehat{f}(\{1\}) = \widehat{f}(\{2\}) = \cdots = \widehat{f}(\{n\})$. Then,

$$tn \leq \sqrt{\frac{2n}{\pi}} + O(n^{-1/2}) \Rightarrow t \leq \sqrt{\frac{2}{\pi n}} + O(n^{-3/2})$$

This yields the desired result,

$$W^1[f] = nt^2 \leq \frac{2}{\pi} + O(n^{-1})$$

Problem 25. Show that $T_\rho f = \sum \rho^{|S|}\widehat{f}(S)$ using exercise 1.

Solution 25. Follows since $T_\rho f(x) = F(\rho x)$.

Problem 26. For each function f in Exercise 1, compute $\mathrm{Inf}[f]$.

Solution 26. Use the formula for total influence using the Fourier coefficients.

Problem 27. Which functions $f : \{-1,1\}^n \to \{-1,1\}$ with $|\{x : f(x) = 1\}| = 3$ maximize $\mathrm{Inf}[f]$?

Solution 27. Suppose that $f(y) = f(z) = f(w) = 1$ all distinct and $f(x) = -1$ for all $x \neq y, z, w$. The total influence equals n times the fraction of edges in the Hamming cube which are boundary edges. To maximize this quantity we may choose y, z, w such that there is no edge connecting it. In this case there are $6n$ boundary edges and so the total influence is $\frac{6n}{2^{n-1}}$.

Problem 28. Suppose $f : \{-1,1\}^n \to \mathbb{R}$ is an even function. Show the improved Poincar Inequality $\mathrm{Var}[f] \leq \frac{1}{2}\mathrm{Inf}[f]$.

Solution 28. For even function we have $\widehat{f}(S) = 0$ with S with $|S|$ odd and so,

$$\begin{aligned}
2\mathrm{Var}[f] &= \sum_{S\neq\emptyset} 2\widehat{f}(S)^2 \\
&= \sum_{|S|\geq 1} 2\widehat{f}(S)^2 \\
&= \sum_{|S|\geq 2} 2\widehat{f}(S)^2 \\
&\leq \sum_{S\neq\emptyset} |S|\,\widehat{f}(S)^2 \\
&= \mathrm{Inf}[f]
\end{aligned}$$

Problem 29. Let $f : \{-1,1\}^n \to \{-1,1\}$ be unbiased, $\mathbb{E}_x[f(x)] = 0$, and let $\mathrm{MaxInf}[f] = \max_i \mathrm{Inf}_i[f]$.

(a) Use the Poincaré Inequality to show $\mathrm{MaxInf}[f] \geq \frac{1}{n}$.

(b) Prove that $\mathrm{Inf}[f] \geq 2 - n\mathrm{MaxInf}[f]^2$

(c) Deduce that $\mathrm{MaxInf}[f] \geq \frac{2}{n} - \frac{4}{n^2}$.

Solution 29. By Poincaré Inequality,

$$\mathrm{Var}[f] \leq \mathrm{Inf}[f] \leq n\mathrm{MaxInf}[f]$$

For unbiased functions $\mathrm{Var}[f] = 1$ and so the result follows. Similarly we have,

$$\begin{aligned}
\mathrm{Inf}[f] &= \sum_S |S|\,\widehat{f}(S)^2 \\
&= \sum_S ||S| = 1|\,\widehat{f}(S)^2 + \sum_S ||S| \geq 2|\,|S|\,\widehat{f}(S)^2 \\
&\geq W^1[f] + 2(1 - W^1[f]) \\
&= 2 - W^1[f]
\end{aligned}$$

This implies that,

$$2 - W^1[f] \leq \mathrm{Inf}[f]$$

To conclude (b) we need to show that,

$$W^1[f] = \sum_i \widehat{f}(\{i\})^2 \leq n\mathrm{MaxInf}[f]^2$$

Recall that $\left|\widehat{f}(\{i\})\right| \leq \mathrm{Inf}_i[f]$ then $\widehat{f}(\{i\})^2 \leq \mathrm{Inf}_i[f]^2$. This implies,

$$\begin{aligned}
\sum_i \widehat{f}(\{i\})^2 &\leq n \cdot \max_j \widehat{f}(\{j\})^2 \\
&\leq n\mathrm{MaxInf}[f]^2
\end{aligned}$$

and so (b) follows. For (c), assume towards contradiction that $\mathrm{MaxInf}[f] < \frac{2}{n} - \frac{4}{n^2}$ then by (b),

$$\mathrm{Inf}[f] > 2 - n\left(\frac{2}{n} - \frac{4}{n^2}\right)^2 = 2 - \frac{4}{n} + \frac{16}{n^2} + (n^3)$$

Contradiction since,

$$\mathrm{Inf}[f] \leq n\mathrm{MaxInf}[f] < 2 - \frac{4}{n}$$

Problem 30. Use Exercises 1 to conclude that,

(a) $E_f i = \sum_{i \notin S} \widehat{f}(S) x^S$

(b) $T_\rho f = \sum \rho^{|S|} \widehat{f}(S) x^S$

Solution 30. We use that $\langle f, \varphi \rangle$ equals the expected value of $f(x)$ where x is sampled from the distribution corresponding to the density function φ.

(a) Let φ_a^i the density distribution corresponding to the uniform distribution on $\{x, x^{\oplus}\}$. We have seen that $\varphi_a^i = \sum_{i \notin S} a^S x^S$. Using Parseval,

$$\begin{aligned} E_f i(a) &= \langle f, \varphi_a^i \rangle \\ &= \sum_{i \notin S} \widehat{f}(S) a^S \end{aligned}$$

(b) Let φ_ρ he density function corresponding to the product probability distribution on $\{1, 1\}^n$ in which each coordinate has mean $\rho \in [-1, 1]$. We have seen that $\varphi_\rho = \sum_S \rho^{|S|} x^S$. Using Parseval,

$$\begin{aligned} T_\rho f &= \langle f, \varphi_\rho \rangle \\ &= \sum_S \widehat{f}(S) \rho^{|S|} \widehat{f}(S) \end{aligned}$$

Problem 31. Show that T_ρ is positivity-preserving for $\rho \in [-1, 1]$, i.e., $f \geq 0 \Rightarrow T_\rho f \geq 0$. Show that T_ρ is positivity-improving for $\rho \in [-1, 1]$, i.e., $f \geq 0, f \neq 0 \Rightarrow T_\rho f > 0$.

Solution 31. Let $f \geq 0$ then,

$$\begin{aligned} &T_\rho f(x) \mathbb{E}_{y \sim_\rho x}[f(x)] \\ &\geq 0 \end{aligned}$$

Since $f(x) \geq 0$ for all x. If $f(x_0) > 0$ then the above is positive for some x_0 hence for any x there is (maybe small) nonzero probability that $y = x_0$.

Problem 32. Show that $T_{\rho_1\rho_2} = T_{\rho_1}T_{\rho_2}$.

Solution 32. Follows since the following distributions are equivalent for fixed x:

1. Draw y s.t. y is $\rho_1\rho_2$ correlated with x.
2. Draw y s.t. y is ρ_1 correlated with x and then draw y' s.t. y' is ρ_2 correlated with y. Output y'.

To see this note that being ρ-correlated with x means that is each coordinate of x is flipped with probability $\frac{1-\rho}{2}$. The above claim follows since,

$$\frac{1-\rho_1\rho_2}{2} = \frac{1-\rho_1}{2}\cdot\frac{1+\rho_2}{2} + \frac{1+\rho_1}{2}\cdot\frac{1-\rho_2}{2}$$

Problem 33. For $\rho \in [-1,1]$, show that T_ρ is a contraction on $L_p(\{1,1\}^n)$ for all $p \geq 1$, i.e., $||T_\rho f||_p \leq ||f||_p$ for all $f : \{-1,1\}^n \to \mathbb{R}$.

Solution 33. Let $\rho \in [-1,1]$, $p \geq 1$ and $f : \{-1,1\}^n \to \mathbb{R}$. Using Jensen inequality

$$\begin{aligned}
||T_\rho f||_p^p &= \mathbb{E}_x\left[|T_\rho f(x)|^p\right] \\
&= \mathbb{E}_x\left[\left|\sum_S \rho^{|S|}\widehat{f}(S)\right|^p\right] \\
&= \mathbb{E}_x\left[\left|\sum_S \rho^{p|S|}\left|\widehat{f}(S)\right|^p\right|\right] \\
&\leq \mathbb{E}_x\left[\left|\sum_S \left|\widehat{f}(S)\right|^p\right|\right] \\
&= ||f||_p^p
\end{aligned}$$

Problem 34. Show that $|T_\rho f| \leq T_\rho|f|$ pointwise for any $f : \{-1,1\}^n \to \mathbb{R}$. Further show that for $-1 < \rho < 1$, equality occurs if and only if f is everywhere nonnegative or everywhere non-positive.

Solution 34. Let $f : \{-1,1\}^n \to \mathbb{R}$. By the triangle inequality,

$$\begin{aligned}
|T_\rho f| &= \left|\mathbb{E}_{y\sim N_\rho(x)}\left[f(y)\right]\right| \\
&\leq \mathbb{E}_{y\sim N_\rho(x)}\left[|f|(y)\right] \\
&= T_\rho|f|(x)
\end{aligned}$$

The inequality is tight if taking the triangle inequality is tight which happens iff $f(x) \geq 0$ for any x or $f(x) \leq 0$ for any x.

Problem 35. For $i \in [n]$ and $\in \mathbb{R}$, let $T_i\rho$ be the operator on functions $f : \{-1,1\}^n \to \mathbb{R}$ defined by,

$$T_\rho^i f = \rho f + (1-\rho)E_i f = E_i f + L_i f$$

(a) Show that for $\rho \in [-1,1]$ we have,

$$T_\rho^i f = \mathbb{E}_{y_i \sim N_\rho(x_i)}\left[f(x_1,\ldots,x_{i-1},y_i,x_{i+1},\ldots,x_n)\right]$$

(b) Show that $T^i_{\rho_1\rho_2} f = T^i_{\rho_1} T^i_{\rho_2} f$ and that $T^i_{\rho_1}, T^j_{\rho_2}$ commute for all i, j, ρ_1, ρ_2.

(c) For $(\rho_1,\ldots,\rho_n) \in \mathbb{R}^n$ define $T_{(\rho_1,\ldots,\rho_n)} f = T^1_{\rho_1} T^2_{\rho_2} \cdots T^n_{\rho_n}$. Show that $T_{(\rho,\ldots,\rho)} f = T_\rho f$ and $T_{(1,\ldots,1,\rho,1\ldots,1)} f = T^i_\rho f$ (with ρ in the i'th position).

(d) If $(\rho_1,\ldots,\rho_n) \in [-,1]^n$ show that $T_{(\rho_1,\ldots,\rho_n)} f$ is contraction on $L_p(\{-1,1\}^n)$.

Solution 35.

(a) Let $\rho \in [-1,1]$ then,

$$\begin{aligned}
T^i_\rho f &= \rho f + (1-\rho) E_i f \\
&= \rho f(x_1,\ldots,x_n) + (1-\rho)\mathbb{E}_{x_i}\left[f(x_1,\ldots,x_{i-1},x_i,x_{i+1},\ldots,x_n)\right] \\
&= \mathbb{E}_{y_i \sim N_\rho(x_i)}\left[f(x_1,\ldots,x_{i-1},y_i,x_{i+1},\ldots,x_n)\right]
\end{aligned}$$

(b) The argument is the same as for $T_{\rho_1\rho_2} = T_{\rho_1} T_{\rho_2}$.

(c) If $i = j$ then (b) implies the operations commute. If $i \neq j$ then it follows from changing the order of summation in the formula obtained in (a), or in other words, the operations act on different bits so in that sense it is independent.

(d) Suffice to show that $T^i_\rho f$ is a contraction. Follows from Jensen inequality,

$$\begin{aligned}
||T^i_\rho f||_p^p &= \mathbb{E}_x\left[\left|\mathbb{E}_{y_i \sim N_\rho(x_i)}\left[f(x_1,\ldots,x_{i-1},y_i,x_{i+1},\ldots,x_n)\right]\right|^p\right] \\
&\leq \mathbb{E}_x\left[\left|\mathbb{E}_{y_i \sim N_\rho(x_i)}\left[|f(x_1,\ldots,x_{i-1},y_i,x_{i+1},\ldots,x_n)|^p\right]\right|\right] \\
&= \mathbb{E}_{(x_1,\ldots,y_i,\ldots,x_n)}\left[\left|\mathbb{E}_{x_i \sim N_\rho(y_i)}\left[|f(x_1,\ldots,x_{i-1},y_i,x_{i+1},\ldots,x_n)|^p\right]\right|\right] \\
&= \mathbb{E}_{(x_1,\ldots,y_i,\ldots,x_n)}\left[||f(x_1,\ldots,x_{i-1},y_i,x_{i+1},\ldots,x_n)|^p|\right] \\
& ||f||_p^p
\end{aligned}$$

Problem 36. Show that $\mathrm{Stab}_\rho[f] = -\mathrm{Stab}_\rho[f]$ if f is odd and $\mathrm{Stab}_\rho[f] = \mathrm{Stab}_\rho[f]$ if f is even.

Solution 36. Follows from the spectral formula for stability.

Problem 37. For each function f in Exercise 1, compute $\mathrm{Stab}_\rho[f]$.

Solution 37. Use the spectral formula.

Problem 38. Compute the stability of the tribe function.

Solution 38. First we establish the following identities regarding stability of a boolean function.

Lemma 2.1. *Let $f : \{-1,1\}^n \to \{-1,1\}$ boolean function. Then,*

$$\begin{aligned}
Stab_\rho[f] &= Pr_{x,y\sim N_\rho(x)}[f(x)=f(y)] - Pr_{x,y\sim N_\rho(x)}[f(x)\neq f(y)] \\
&= 2Pr_{x,y\sim N_\rho(x)}[f(x)=f(y)] - 1 \\
&= 1 - 2Pr_{x,y\sim N_\rho(x)}[f(x)\neq f(y)] \\
&= 1 - 4Pr_{x,y\sim N_\rho(x)}[f(x)=1 \wedge f(y)=-1] \\
&= 1 - 4Pr_{x,y\sim N_\rho(x)}[f(x)=-1 \wedge f(y)=1] \\
&= 1 - 4Pr_{x,y\sim N_\rho(x)}[f(x)=1]\, Pr_{x,y\sim N_\rho(x)}[f(y)=-1 : f(x)=1] \\
&= 1 - 4Pr_{x,y\sim N_\rho(x)}[f(x)=-1]\, Pr_{x,y\sim N_\rho(x)}[f(y)=1 : f(x)=-1]
\end{aligned}$$

The last two identities follows from the symmetry of x, y.

Lemma 2.2. *Let $g_1, \ldots, g_s : \{-1,1\}^n \to \{-1,1\}$ be s boolean functions on w disjoint variables. Such that.*

- $Pr[g_i(x) = -1] = p_i$
- $Stab_\rho[g_i] = q_i$

Then,

$$Stab_\rho\left[\bigvee_{i=1}^{s} g_i(x^{(i)})\right] = 1 - 4\cdot\left(\prod p_i\right)\cdot\left(1 - \prod_i\left(1 - \frac{1-q_i}{4p_i}\right)\right)$$

Proof. Set $f(x) = \bigvee_{i=1}^{s} g_i(x^{(i)})$. Using the formulas for stability,

$$\mathrm{Stab}_\rho\left[\bigvee_{i=1}^{s} g_i(x^{(i)})\right] = 1 - 4\mathrm{Pr}_{x,y\sim N_\rho(x)}[f(x)=-1]\,\mathrm{Pr}_{x,y\sim N_\rho(x)}[f(y)=1 : f(x)=-1]$$

To have $f(x) = -1$ we need $g_i(x^{(i)})$ for any $x^{(i)}$. Thus,

$$\begin{aligned}
&\mathrm{Pr}_{x,y\sim N_\rho(x)}[f(y)=1 : f(x)=-1] = 1 - \mathrm{Pr}_{x,y\sim N_\rho(x)}[f(y)=-1 : f(x)=-1] \\
&= 1 - \prod \mathrm{Pr}_{x,y\sim N_\rho(x)}\left[g_i(y^{(i)}) = -1 : f(x^{(i)}) = -1\right] \\
&= 1 - \prod\left(1 - \mathrm{Pr}_{x,y\sim N_\rho(x)}\left[g_i(y^{(i)}) = 1 : f(x^{(i)}) = -1\right]\right) \\
&= 1 - \prod\left(1 - \frac{\mathrm{Pr}_{x,y\sim N_\rho(x)}\left[f(x^{(i)}) = -1\right]\mathrm{Pr}_{x,y\sim N_\rho(x)}\left[g_i(y^{(i)}) = 1 : f(x^{(i)}) = -1\right]}{\mathrm{Pr}_{x,y\sim N_\rho(x)}[f(x^{(i)}) = -1]}\right) \\
&= 1 - \prod_i\left(1 - \frac{1-q_i}{4p_i}\right)
\end{aligned}$$

We conclude that,

$$\mathrm{Stab}_\rho\left[\bigvee_{i=1}^{s} g_i(x^{(i)})\right] = 1 - 4\cdot\left(\prod p_i\right)\cdot\left(1 - \prod_i\left(1 - \frac{1-q_i}{4p_i}\right)\right)$$

□

The stability of the tribe function follows by considering g_i being the AND function and so $p_i = 1 - 2^{-w}$ and $q_i = 1 - 4 \cdot 2^{-w}\left(1 - \left(\frac{1+\rho}{2}\right)^w\right)$.

$$
\begin{aligned}
\mathrm{Stab}_\rho\left[\mathrm{Tribe}_{w,s}\right] &= 1 - 4 \cdot (1-2^{-w})^s \cdot \left(1 - \left(1 + \frac{2^{-w}\left(1 - \left(\frac{1+\rho}{2}\right)^w\right)}{(1-2^{-w})}\right)^s\right) \\
&= 1 - 4\left[(1-2^{-w})^s - \left((1-2^{-w}) + 2^{-w}\left(1 - \left(\frac{1+\rho}{2}\right)^w\right)\right)^s\right] \\
&= 1 - 4\left[(1-2^{-w})^s - \left(1 - 2^{-w}\left(\frac{1+\rho}{2}\right)^w\right)^s\right]
\end{aligned}
$$

Problem 39. Let $f : \{-1,1\}^n \to \{-1,1\}$ and suppose $\min\{\mathrm{Pr}_x\left[f(x) = 1\right], \mathrm{Pr}_x\left[f(x) = -1\right]\} = \alpha$. Prove that for any $0 \leq \delta \leq 1$,

$$\mathrm{Stab}_f\left[\delta\right] \leq 2\alpha$$

Solution 39. Using the identities from lemma 2.1,

$$
\begin{aligned}
\mathrm{Stab}_f\left[\delta\right] &= \frac{1}{2} - \frac{1}{2}\mathrm{Stab}_{1-2\delta}\left[f\right] \\
&= 2\mathrm{Pr}_{x,y\sim N_\rho(x)}\left[f(x) = 1\right]\mathrm{Pr}_{x,y\sim N_\rho(x)}\left[f(y) = -1 : f(x) = 1\right] \\
&= 2\mathrm{Pr}_{x,y\sim N_\rho(x)}\left[f(x) = -1\right]\mathrm{Pr}_{x,y\sim N_\rho(x)}\left[f(y) = 1 : f(x) = -1\right]
\end{aligned}
$$

and so the result follows.

Problem 40. Prove that,

$$\mathrm{Inf}^{(\rho)}\left[\rho\right] = \frac{\mathrm{d}}{\mathrm{d}\rho}\mathrm{Stab}_\rho\left[f\right] = \sum_{k=1}^{n} k\rho^{k-1}W^k\left[f\right]$$

Solution 40. Differentiating the spectral formula for stability yield the result.

Problem 41. Fix $f : \{-1,1\}^n \to \mathbb{R}$. Show that $\mathrm{Stab}_\rho\left[f\right]$ is convex in $\rho \in [0,1]$.

Solution 41. The second derivative of $\mathrm{Stab}_\rho\left[f\right]$ is given by,

$$\sum_{k=2}^{n} k(k-1)\rho^{k-2}W^k\left[f\right]$$

positive for $\rho \in [0,1]$.

Problem 42. Let $f : \{-1,1\}^n \to \{-1,1\}$. Show that $\mathrm{Stab}_f\left[\delta\right] \leq \delta\mathrm{Inf}\left[f\right]$ for $\delta \in [0,1]$.

Solution 42. The spectral formula for $\mathrm{Stab}_f[\delta]$ is,

$$\mathrm{Stab}_f[\delta] = \frac{1}{2}\sum_{k=0}^{n}\left(1-(1-2\delta)^k\right)W^k[f]$$

The result follows from the easy inequality $1-(1-2\delta)^k \leq 2\delta k$.

Problem 43. Define the average influence,

$$\mathcal{E}[f] = \frac{1}{n}\mathrm{Inf}[f]$$

Prove the following:

(a)
$$\mathcal{E}[f] = \mathrm{Pr}_{x,i}\left[f(x) \neq f(x^{\oplus i})\right]$$

(b)
$$\frac{1-e^{-2}}{2}\mathcal{E}[f] \leq \mathrm{NS}_{1/n}[f] \leq \mathcal{E}[f]$$

(c) Given $f:\{-1,1\}^n \to \{-1,1\}$ and integer $k \geq 2$ define,

$$A_k = \frac{1}{k}\left(W^{\geq 1}[f] + W^{\geq 2}[f] + \cdots + W^{\geq k}[f]\right)$$

Prove that,

$$\frac{1-e^{-2}}{2}A_k \leq \mathrm{NS}_{1/k}[f] \leq A_k$$

Solution 43.

(a) By using the definition of influence and the law of total probability,

$$\begin{aligned}
\mathcal{E}[f] &= \frac{1}{n}\sum_{j=1}^{n}\mathrm{Inf}_j[f] \\
&= \sum_{j=1}^{n}\frac{1}{n}\cdot \mathrm{Pr}_x\left[f(x) \neq f(x^{\oplus j})\right] \\
&= \sum_{j=1}^{n}\mathrm{Pr}_{i\sim[n]}[i=j]\cdot \mathrm{Pr}_x\left[f(x) \neq f(x^{\oplus j})\right] \\
&= \sum_{j=1}^{n}\mathrm{Pr}_{x,i\sim[n]}\left[f(x) \neq f(x^{\oplus j}) \wedge i=j\right] \\
&= \mathrm{Pr}_{x,i\sim[n]}\left[f(x) \neq f(x^{\oplus j})\right]
\end{aligned}$$

(b) Use result in exercise 2 with $\delta = 1/n$ for the upper bound. The lower bounds follow from the same argument plus the following inequality,

$$(1 - e^{-2}) \cdot \frac{k}{n} \leq 1 - \left(1 - \frac{2}{n}\right)^k$$

which is equivalent to,

$$1 - e^{-2} \leq \frac{n}{k}\left(1 - \left(1 - \frac{2}{n}\right)^k\right)$$

This follows from the inequality,

$$(1 - e^{-2}) \leq (1 - 2/n)^n \leq (1 - 2/n)^k$$

which is true for all $k \leq n$.

(c) The difference between the spectral formula for the average influence and A_k is that Fourier weights with degree $d > k$ are counted k times (and not d). Still, for $d > k$ we have the inequalities,

$$1 - e^{-2} \leq 1 - (1 - 2/d)^d \leq 1$$

so the same proof works (with minor changes).

Problem 44. Suppose $f_1, \ldots, f_s : \{-1, 1\}^n \to \{-1, 1\}$ satisfy $\mathrm{NS}_\delta[f_i] \leq \epsilon_i$. Let $g : \{-1, 1\}^s \to \{-1, 1\}$ and define $h : \{-1, 1\}^n \to \{-1, 1\}$ by $h = g(f_1, \ldots, f_s)$. Show that,

$$\mathrm{NS}_\delta[h] \leq \sum \epsilon_i$$

Solution 44. Changing x with probability δ (i.e, independently flipping each coordinate with probability δ) yields a change to f_i with probability at most ϵ_i. By union bound, the probability that any of the f_i changes is bounded by $\sum \epsilon_i$ and in particular in that case h will not change (regardless of g).

Problem 45. Show that $(1 - \delta)^{k-1}k \leq 1/\delta$.

Solution 45. Using the geometric sum of $(1 - \delta)$ and that $(1 - \delta)^j \leq (1 - \delta)^{k-1}$ for any $j \leq k - 1$,

$$(1 - \delta)^{k-1}k \leq 1 + (1 - \delta) + (1 - \delta)^2 + \ldots + (1 - \delta)^{k-1} \leq \frac{1 - (1 - \delta)^k}{\delta} \leq 1/\delta$$

Problem 46. Fixing $f : \{-1, 1\}^n \to \mathbb{R}$, show the following Lipschitz bound for stability. When $0 \leq \rho \leq \rho \leq 1$,

$$|\mathrm{Stab}_\rho[f] - \mathrm{Stab}_{\rho-\epsilon}[f]| \leq \epsilon \cdot \frac{1}{1 - \rho} \cdot \mathrm{Var}[f]$$

Solution 46. By the Lagrange theorem,

$$\mathrm{Stab}_{\rho}[f] - \mathrm{Stab}_{\rho-\epsilon}[f] = \epsilon \cdot \frac{\mathrm{d}}{\mathrm{d}\rho}\mathrm{Stab}_{\rho}[f]\Big|_{\rho=\rho_0}$$

where $\rho_0 \in [\rho-\epsilon, \rho]$. Using the spectral formula for stability,

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}\rho}\mathrm{Stab}_{\rho}[f] &= \frac{\mathrm{d}}{\mathrm{d}\rho}\sum_{k\geq 0}\rho^k W^k[f] \\
&= \sum_{k>0} k\rho^{k-1}W^k[f] \\
&\leq (1+\rho+\rho^2+\ldots)\cdot\sum_{k>0} kW^k[f] \\
&\leq \frac{1}{1-\rho}\cdot \mathrm{Var}[f]
\end{aligned}$$

and so the estimation follows since $\frac{1}{1-\rho_0} \leq \frac{1}{1-\rho}$.

Problem 47. Let $f : \{-1,1\}^n \to \{-1,1\}$ be a transitive-symmetric function. Show that,

$$\Pr_{\pi\sim\mathrm{Aut}(f)}[\pi(i)=j] = 1/n$$

for all $i, j \in [n]$.

Solution 47. Let i, j, j'. Since $\mathrm{Aut}(f)$ is transitive then there exists π_0 s.t. $\pi_0(j') = j$. Since applying π_0 is an automorphism on $\mathrm{Aut}(f)$,

$$|\{\pi \in \mathrm{Aut}(f) : \pi(i) = j\}| = \left|\{\pi_0^{-1}\pi \in \mathrm{Aut}(f) : \pi(i) = j\}\right| \leq |\{\pi \in \mathrm{Aut}(f) : \pi(i) = j'\}|$$

Similarly for i, i', j,

$$|\{\pi \in \mathrm{Aut}(f) : \pi(i) = j\}| \leq |\{\pi \in \mathrm{Aut}(f) : \pi(i') = j\}|$$

Thus implies that for any i, j, i', j' it must be that,

$$|\{\pi \in \mathrm{Aut}(f) : \pi(i) = j\}| \leq |\{\pi \in \mathrm{Aut}(f) : \pi(i') = j'\}|$$

Applying the above equality while reversing the role of i, j, i', j' we get equality. Thus for any i, j,

$$\Pr_{\pi\sim\mathrm{Aut}(f)}[\pi(i)=j] = c$$

for some constant c. Since for fixed i, running on j, this must amount to 1 it must be that $c = 1/n$.

Problem 48. Suppose that F is a functional on functions $f : \{-1,1\}^n \to \mathbb{R}$ expressible as $F[f] = \sum c_S \widehat{f}(S)^2$ where $c_S \geq 0$ for all $S \subseteq [n]$. Show that F is convex. That is for any $0 \leq \lambda \leq 1$,

$$F(\lambda f + (1-\lambda)g) \leq \lambda F(f) + (1-\lambda)F(g)$$

Solution 48. Using additivity of the Fourier coefficients,

$$
\begin{aligned}
F(\lambda f+(1-\lambda)g) &= \sum_S c_S \widehat{\lambda f+(1-\lambda)g}(S)^2 \\
&= \sum_S c_S(\lambda \widehat{f}(S)+(1-\lambda)\widehat{g}(S))^2 \\
&= \sum_S c_S(\lambda^2 \widehat{f}(S)^2+(1-\lambda)^2\widehat{g}(S)^2+2\lambda(1-\lambda)\widehat{f}(S)\widehat{g}(S)) \\
&= \lambda F(f)+(1-\lambda)F(g)-\sum_S c_S((\lambda-\lambda^2)\widehat{f}(S)^2+((1-\lambda)^2-(1-\lambda)^2)\widehat{g}(S)^2-2\lambda(1-\lambda)\widehat{f}(S)\widehat{g}(S)) \\
&= \lambda F(f)+(1-\lambda)F(g)-\sum_S c_S(\lambda(1-\lambda)\widehat{f}(S)^2+\lambda(1-\lambda)\widehat{g}(S)^2-2\lambda(1-\lambda)\widehat{f}(S)\widehat{g}(S)) \\
&= \lambda F(f)+(1-\lambda)F(g)-\sum_S c_S(\widehat{f}(S)-\widehat{g}(S))^2
\end{aligned}
$$

Using that $c_S \geq 0$ the sum of the right is positive hence the result follows.

Problem 49. Extend the FKN Theorem as follows: Suppose $f:\{-,1,1\}^n \to \{-1,1\}$ has $W^{\leq 1}[f]=1-\delta$. Show that f is $O(\delta)$-close to 1-junta.

Solution 49. Define $g(x_0, x) = x_0 f(x_0, x)$. Note that,

$$\begin{aligned}
\widehat{g}(\emptyset) &= \mathbb{E}_{x_0,x}\left[g(x_0, x)\right] \\
&= \mathbb{E}_{x_0,x}\left[x_0 f(x_0 x)\right] \\
&= \frac{1}{2}\mathbb{E}_x\left[f(x)\right] + \frac{1}{2}\mathbb{E}_x\left[-f(-x)\right] \\
&= \frac{1}{2}\mathbb{E}_x\left[f(x)\right] - \frac{1}{2}\mathbb{E}_x\left[f(-x)\right] \\
&= \frac{1}{2}\mathbb{E}_x\left[f(x)\right] - \frac{1}{2}\mathbb{E}_x\left[f(x)\right] \\
&= 0 \\
\widehat{g}(\{0\}) &= \mathbb{E}_{x_0,x}\left[x_0 g(x_0, x)\right] \\
&= \mathbb{E}_{x_0,x}\left[x_0^2 f(x_0 x)\right] \\
&= \mathbb{E}_{x_0,x}\left[f(x_0 x)\right] \\
&= \frac{1}{2}\mathbb{E}_x\left[f(x)\right] + \frac{1}{2}\mathbb{E}_x\left[f(-x)\right] \\
&= \frac{1}{2}\mathbb{E}_x\left[f(x)\right] + \frac{1}{2}\mathbb{E}_x\left[f(x)\right] \\
&= \widehat{f}(\emptyset) \\
\widehat{g}(\{i\}) &= \mathbb{E}_{x_0,x}\left[x_i g(x_0, x)\right] \\
&= \mathbb{E}_{x_0,x}\left[x_0 x_i f(x_0 x)\right] \\
&= \mathbb{E}_{x_0,x}\left[x_i f(x_0 x)\right] \\
&= \frac{1}{2}\mathbb{E}_x\left[f(x)\right] - \frac{1}{2}\mathbb{E}_x\left[x_i f(-x)\right] \\
&= \frac{1}{2}\mathbb{E}_x\left[f(x)\right] + \frac{1}{2}\mathbb{E}_x\left[x_i f(x)\right] \\
&= \widehat{f}(\{i\})
\end{aligned}$$

It follows that $W^1[g] = W^{\leq 1}[f]$ thus by FKN g is $O(\delta)$-close to some $\pm\chi_i$. Substitute $x_0 = 1$ we get that f is either constant or $O(\delta)$ close to some $\pm\chi_i$.

Problem 50. Compute the precise probability of a Condorcet winner (under impartial culture) in a 3-candidate, 3-voter election using $f = \mathrm{Maj}_3$.

Solution 50. The probability for a Condorcet winner for f is exactly,

$$\frac{3}{4} - \frac{3}{4}\mathrm{Stab}_{-1/3}\left[f\right]$$

Calculating the stability of $f = \mathrm{Maj}_3$ we get,

$$\begin{aligned}
\frac{3}{4} - \frac{3}{4}\mathrm{Stab}_{-1/3}[f] &= \frac{3}{4} - \frac{3}{4}\sum_k \rho^k W^k[f] \\
&= \frac{3}{4} - \frac{3}{4}\left[(-1/3)\cdot\frac{3}{4} + (-1/3)^2\cdot 0 + (-1/3)^3\cdot\frac{1}{4}\right] \\
&= \frac{3}{4} + \frac{3}{16} + \frac{3}{4^2\cdot 27} \\
&= 0.9444\ldots
\end{aligned}$$

Problem 51. Arrows Theorem for 3 candidates is slightly more general than what we stated: it allows for three different unanimous functions $f, g, h : -1, 1^n \to \{-1, 1\}$ to be used in the three pairwise elections. But show that if using f, g, h always gives rise to a Condorcet winner then $f = g = h$. Moreover, extend Arrow's theorem to the case of Condorcet elections with more than 3 candidates.

Solution 51. Suppose $f, g, h : -1, 1^n \to \{-1, 1\}$ always gives rise to a Condorcet winner. We assume all functions are non-constant (if one function is constant, one can achieve Condorcet winner). Note that $x, y = -x, z$ is always a valid voting scheme. Fix x. Since h is non-constant then $\exists z, z'$ s.t. $h(z) = f(x)$ and by considering the voting scheme $(x, -x, z)$ we conclude that $f(x) = -g(-x)$. Similarly (assuming g is non-constant) we get $f(x) = -h(-x)$. Together, it implies that $g(x) = h(x)$. Next, we prove that $g(x) = -g(-x)$. Assume not, then there exists y s.t. $g(y) = g(-y)$. Take x s.t. $f(x) = g(y)$ then the valid voting scheme $(x, y, -y)$ does not give rise to a Condorcet winner. We conclude that $f = g = h$.

To extend Arrow's theorem to more than three candidates, consider $\binom{n}{2}$ voting schemes between all couples. The result (and voting) is not contradictory if the directed graph on n vertices with edges,

$$\{i \to j : \text{vote } i \text{ over } j\}$$

does not contain a cycle. It is not hard to see,

Problem 52. The polarizations of $f : \{-1, 1\}^n \to \mathbb{R}$ (also known as compressions, downshifts, or two-point rearrangements) are defined as follows. For $i \in [n]$, the i-polarization of f is the function $fi : \{-1, 1\}^n \to \mathbb{R}$ defined by,

$$f^{\sigma_i}(x) = \begin{cases} \max\{f(x^{(i\to 1)}), f(x^{(i\to -1)})\} & x_i = 1 \\ \min\{f(x^{(i\to 1)}), f(x^{(i\to -1)})\} & x_i = -1 \end{cases}$$

(a) Show that $\mathbb{E}_x[f(x)] = \mathbb{E}_x[f^{\sigma_i}(x)]$ and $||f||_p = ||f^{\sigma_i}||_p$ for any p.

(b) Show that $\mathrm{Inf}_j[f^{\sigma_i}] \le \mathrm{Inf}_j[f]$.

(c) Show that $\mathrm{Stab}_\rho[f^{\sigma_i}] \ge \mathrm{Stab}_\rho[f]$ for all $0 \le \rho \le 1$.

(d) Show that f^{σ_i} is monotone in the i'th direction. Furthermore, show that if f is monotone in the j'th direction then f^{σ_i} is also monotone in the j'th direction.

(e) Let $f^* = f^{\sigma_1\sigma_2\cdots\sigma_n}$. Show that $\mathbb{E}_x[f(x)] = \mathbb{E}_x[f^*(x)]$, $\mathrm{Inf}_j[f^*] \le \mathrm{Inf}_j[f]$ and $\mathrm{Stab}_\rho[f^*] \ge \mathrm{Stab}_\rho[f]$ for any $0 \le \rho \le 1$.

Solution 52. Let $f : \{-1,1\}^n \to \mathbb{R}$.

(a) Follows from the observation,

$$\{f^{\sigma_i}(x^{(i\to 1)}), f^{\sigma_i}(x^{(i\to -1)})\} = \{f(x^{(i\to 1)}), f(x^{(i\to -1)})\}$$

(b) Follows from the fact that,

$$D_j f^{\sigma_i}(x^{(i\to 1)})^2 + D_j f^{\sigma_i}(x^{(i\to -1)})^2 \le D_j f(x^{(i\to 1)})^2 + D_j f(x^{(i\to -1)})^2$$

To see this, first set

$$\begin{aligned}
A &= \max\{x^{(i\to 1)}j, x^{(i\to 1)^{\oplus j}}\} \\
B &= \min\{x^{(i\to 1)}j, x^{(i\to 1)^{\oplus j}}\} \\
C &= \max\{x^{(i\to -1)}j, x^{(i\to -1)^{\oplus j}}\} \\
D &= \min\{x^{(i\to -1)}j, x^{(i\to -1)^{\oplus j}}\}
\end{aligned}$$

We have that $A \ge B$, $C \ge D$. The above inequality follows from the easy observation that,

$$(A-C)^2 + (B-D)^2 \le (A-D)^2 + (B-C)^2$$

(c) First observe that for $A \ge B$,$C \ge D$ it holds that,

$$D(A-B) \le C(A-B) \Rightarrow AD + BC \le AC + BD$$

We need to show that,

$$\mathbb{E}_x[f(x)T_\rho f] \le \mathbb{E}_x[f(x)T_\rho f^{\sigma_i}]$$

or equivalently,

$$\mathbb{E}_{x,y\sim N_\rho(x)}[f(x)f(y)] \le \mathbb{E}_{x,y\sim N_\rho(x)}[f^{\sigma_i}(x)f^{\sigma_i}(y)]$$

We shall prove this equality by considering the expected-value four terms each time. Let x and y, then we have 4 different events (x,y) , $(x,y^{\oplus i})$,$(x^{\oplus i},y)$,$(x^{\oplus i},y^{\oplus i})$ with the convention that $x_i = y_i$ (this is important since it determines the probabilities in which these events happen). The contribution of these events to the right side is given by,

$$\frac{1+\rho}{2}\left(f^{\sigma_i}(x)f^{\sigma_i}(y) + f^{\sigma_i}(x^{\oplus i})f^{\sigma_i}(y^{\oplus i})\right) + \frac{1-\rho}{2}\left(f^{\sigma_i}(x^{\oplus i})f^{\sigma_i}(y) + f^{\sigma_i}(x)f^{\sigma_i}(y^{\oplus i})\right)$$

The contribution to the left is the same as above and replacing the polarization of f with simply f. Now set,

$$\begin{aligned} A &= \max\{f(x), f(x^{\oplus i})\} \\ B &= \min\{f(x), f(x^{\oplus i})\} \\ C &= \max\{f(y), f(y^{\oplus i})\} \\ D &= \min\{f(y), f(y^{\oplus i})\} \end{aligned}$$

It is important to note that the contribution on the right is now exactly,

$$\frac{1}{2}\left(AC + AD + BC + BD\right) + \frac{\rho}{2}((AC + BD) - (AD + BC))$$

The contribution on the left is the same, with slight change. The rightest term $\frac{\rho}{2}((AC+BD)-(AD+BC))$ may (and also may not) re replaced with $-\frac{\rho}{2}((AC+BD)-(AD+BC))$. Note that $A \geq B$, $C \geq D$ then the above inequality for A, B, C, D implies that in any case the contribution to the expectation of the polarization is always larger.

(d) Show that f^{σ_i} is monotone in the i'th direction. Furthermore, show that if f is monotone in the j'th direction then f^{σ_i} is also monotone in the j'th direction.

Let x. Assuming $x_i = -1$,

$$\begin{aligned} f^{\sigma_i}(x^{(j\to -1)}) &= \min\{f(x^{(i\to -1)^{(j\to -1)}}), f(x^{(i\to 1)^{(j\to -1)}})\} \\ &\leq \min\{f(x^{(i\to -1)^{(j\to 1)}}), f(x^{(i\to 1)^{(j\to 1)}})\} \\ &= f^{\sigma_i}(x^{(j\to 1)}) \end{aligned}$$

The second inequality follows since f is monotone. Replacing minimum by maximum implies the case $x_i = 1$.

(e) Follows from composing the above results.

Problem 53. For $D \geq 1$, we say that the discrete cube can be embedded into ℓ_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:

- $||F(x) - F(y)||_2 \geq \Delta\left(x, y\right)$
- $||F(x) - F(y)||_2 \geq D \cdot \Delta\left(x, y\right)$

for any $x, y \in \{-, 1\}^n$. In this exercise you will show that the least distortion possible is $D = \sqrt{n}$.

(a) Let $f : \{-1, 1\}^n \to \mathbb{R}$ and f_{odd} be the odd part of f. Show that $||f_{odd}||_2^2 \leq \mathrm{Inf}\left[f\right]$ and conclude that,

$$\mathbb{E}_x\left[(f(x) - f(-x))^2\right] \leq \sum_i \mathbb{E}_x\left[(f(x) - f(x^{\oplus i}))^2\right]$$

(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 the above inequality over $i \in [m]$, show that any F with no contraction must have expansion at least $\sqrt{n}$.

(c) Show that there is an embedding F achieving distortion $\sqrt{n}$.

Solution 53.

(a) This is a direct corollary that follows from the spectral formula for $\mathrm{Inf}[f]$ and the Fourier expansion of f_{odd}.

(b) Consider the inequality achieved in (a) and sum over all functions $f_1, \ldots, f_m$,

$$\sum_{j=1}^{m} \mathbb{E}_x \left[(f_j(x) - f_j(-x))^2 \right] \leq \sum_{j=1}^{m} \sum_{i} \mathbb{E}_x \left[(f_j(x) - f_j(x^{\oplus i}))^2 \right]$$

Using that there is no contraction,

$$\sqrt{\sum_{j=1}^{m} (f_j(x) - f_j(-x))^2} \geq n$$

Therefore,

$$n^2 \leq \sum_{j=1}^{m} \sum_{i} \mathbb{E}_x \left[(f_j(x) - f_j(x^{\oplus i}))^2 \right] = \mathbb{E}_x \left[||F(x) - F(x^{\oplus i})||_2^2 \right]$$

Thus, for at least one x and i for which $n \leq ||F(x) - F(x^{\oplus i})||_2^2$ and so the distortion must be at least $\sqrt{n}$.

(c) The natural embedding with $1 \to 1$, $-1 \to 0$ achieves $\sqrt{n}$ (use Cauchy-Schwartz).

Problem 54. Give a Fourier-free proof of the Poincaré Inequality by induction on n

Solution 54. The induction is on the number of variables n. The case $n = 1$ can be easily verified. Let $f : \{-1,1\}^n \to \{-1,1\}$ boolean function and write,

$$f(x_1, \ldots, x_n) = \frac{1 + x_n}{2} f(x_1, \ldots . x_{n-1}, 1) + \frac{1 - x_n}{2} f(x_1, \ldots . x_{n-1}, -1)$$

Set $f^+(x_1, \ldots, x_{n-1}) = f(x_1, \ldots . x_{n-1}, 1)$, $f^-(x_1, \ldots, x_{n-1}) = f(x_1, \ldots . x_{n-1}, -1)$. In this notation,

$$f(x_1, \ldots, x_n) = \frac{1 + x_n}{2} f^+(x_1, \ldots . x_{n-1}) + \frac{1 - x_n}{2} f^-(x_1, \ldots . x_{n-1})$$

By induction hypothesis,

$$\mathrm{Var}\left[f^+\right] \leq \mathrm{Inf}\left[f^+\right] \quad , \quad \mathrm{Var}\left[f^-\right] \leq \mathrm{Inf}\left[f^-\right]$$

We need to show that,

$$\mathrm{Var}\left[f\right] \leq \mathrm{Inf}\left[f\right]$$

By using the induction hypothesis it suffice to show that,

$$\mathrm{Var}\left[f\right] - \frac{1}{2}\mathrm{Var}\left[f^{+}\right] - \frac{1}{2}\mathrm{Var}\left[f^{-}\right] \leq \mathrm{Inf}\left[f^{+}\right] - \frac{1}{2}\mathrm{Inf}\left[f^{+}\right] - \frac{1}{2}\mathrm{Inf}\left[f^{-}\right]$$

Observe that,

$$\mathrm{Inf}\left[f^{+}\right] - \frac{1}{2}\mathrm{Inf}\left[f^{+}\right] - \frac{1}{2}\mathrm{Inf}\left[f^{-}\right] = \mathrm{Inf}_n\left[f\right]$$

and,

$$\mathrm{Var}\left[f\right] - \frac{1}{2}\mathrm{Var}\left[f^{+}\right] - \frac{1}{2}\mathrm{Var}\left[f^{-}\right] = \frac{1}{2}\left(\mathbb{E}\left[f^{+}\right]\right)^2 + \frac{1}{2}\left(\mathbb{E}\left[f^{-}\right]\right)^2 - \frac{1}{2}\left(\mathbb{E}\left[f\right]\right)^2$$

Therefore, it suffice to prove the following inequality,

$$\frac{1}{2}\left(\mathbb{E}\left[f^{+}\right]\right)^2 + \frac{1}{2}\left(\mathbb{E}\left[f^{-}\right]\right)^2 - \frac{1}{2}\left(\mathbb{E}\left[f\right]\right)^2 \leq \mathrm{Inf}_n\left[f\right]$$

To see this define four sets,

$$\begin{aligned}
S_{+,-} &= \{x : f^{+}(x) = 1, f^{-}(x) = -1\}\\
S_{-,+} &= \{x : f^{+}(x) = -1, f^{-}(x) = 1\}\\
S_{-,-} &= \{x : f^{+}(x) = -1, f^{-}(x) = -1\}\\
S_{+,+} &= \{x : f^{+}(x) = 1, f^{-}(x) = 1\}
\end{aligned}$$

with corresponding measure $\mu_{\pm,\pm}$. Observe that,

$$\begin{aligned}
\mathbb{E}\left[f^{+}\right]^2 &= \mu_{+,+} + \mu_{+,-} - \mu_{-,+} - \mu_{-,-}\\
\mathbb{E}\left[f^{-}\right]^2 &= \mu_{-,+} + \mu_{-,-} - \mu_{+,+} - \mu_{+,-}\\
\mathbb{E}\left[f\right]^2 &= \mu_{+,+} - \mu_{-,-}\\
\mathrm{Inf}_n\left[f\right] &= \mu_{+,-} + \mu_{-,+}
\end{aligned}$$

The inequality is then equivalent to,

$$\frac{1}{2}(\mu_{+,+} + \mu_{+,-} - \mu_{-,+} - \mu_{-,-})^2 + \frac{1}{2}(\mu_{+,+} + \mu_{-,+} - \mu_{+,-} - \mu_{-,-}) - (\mu_{+,+} - \mu_{-,-})^2 \leq \mu_{+,-} + \mu_{-,+}$$

Simplifying the left term,

$$\begin{aligned}
&\frac{1}{2}(\mu_{+,+} + \mu_{+,-} - \mu_{-,+} - \mu_{-,-})^2 + \frac{1}{2}(\mu_{+,+} + \mu_{-,+} - \mu_{+,-} - \mu_{-,-}) - (\mu_{+,+} - \mu_{-,-})^2 =\\
&= (\mu_{+,+}^2 + \mu_{+,-}^2 + \mu_{-,+}^2 + \mu_{-,-}^2) - 2\mu_{+,-}\mu_{+,-} - 2\mu_{+,+}\mu_{-,-} - (\mu_{+,+} - \mu_{-,-})^2\\
&= (\mu_{+,+} - \mu_{-,-})^2 + (\mu_{+,-} - \mu_{+,-})^2 - (\mu_{+,+} - \mu_{-,-})^2\\
&= (\mu_{+,-} - \mu_{+,-})^2
\end{aligned}$$

Since $\mu_{+,-} - \mu_{+,-} \leq 1$ we have,

$$(\mu_{+,-} - \mu_{+,-})^2 \leq \mu_{+,-} - \mu_{+,-}$$

and the proof is completed.

Problem 55. Let V be a normed vector space with norm $||\cdot||$ and fix $w_1, \ldots, w_n \in V$ vectors. Define the boolean function $g : \{-1,1\}^n \to \mathbb{R}$

$$g(x) = ||\sum x_i w_i||$$

(a) Show that $Lg \leq g$ pointwise.

(b) Deduce $2\mathrm{Var}[g] \leq \mathbb{E}_x[g(x)^2]$ and thus the following KhintchineKahane Inequality:

$$\mathbb{E}_x\left[||\sum_{i=1}^n x_i w_i||\right] \geq \frac{1}{\sqrt{2}}\mathbb{E}_x\left[||\sum_{i=1}^n x_i w_i||^2\right]^{1/2}$$

(c) Show that the constant $\frac{1}{\sqrt{2}}$ is tight even if $V = \mathbb{R}$.

Solution 55. Let V be a normed vector space with norm $||\cdot||$ and fix $w_1, \ldots, w_n \in V$ vectors.

(a) Recall the definition of the Laplacian operator,

$$\begin{aligned}
Lg &= \sum_{i=j}^{n} L_j g \\
&= \sum_{j=1}^{n} \frac{g(x) - g(x^{\oplus j})}{2} \\
&= \frac{ng(x) - \sum_{j=1}^n ||\sum_{i=1}^n x_i^{\oplus j} w_i||}{2} \\
&\leq \frac{ng(x) - ||\sum_{j=1}^n \sum_{i=1}^n x_i^{\oplus j} w_i||}{2} \\
&= \frac{ng(x) - (n-2)g(x)}{2} \\
&= g(x)
\end{aligned}$$

(b) We shall use the inequality $2\mathrm{Var}[g] \leq \mathrm{Inf}[g]$.

$$\begin{aligned}
2\mathrm{Var}[g] &\leq \mathrm{Inf}[g] \\
&= \langle g, Lg \rangle \\
&= \mathbb{E}_x[g(x)Lg(x)] \\
&\leq \mathbb{E}_x\left[g(x)^2\right]
\end{aligned}$$

This is equivalent to the inequality

$$\mathbb{E}_x\left[||\sum x_i w_i||^2\right] - \mathbb{E}\left[||\sum x_i w_i||\right]^2 \leq \frac{1}{2}\mathbb{E}\left[||\sum x_i w_i||^2\right]$$

Thus Khintchine-Kahane Inequality follows.

(c) Take $w_1 = w_2 = 1$.

Problem 56. In the correlation distillation problem, a source chooses $x \sim \{-1, 1\}$ n uniformly at random and broadcasts it to q parties. We assume that the transmissions suffer from some kind of noise, and therefore the players receive imperfect copies $y^{(1)}, \ldots, y^{(q)}$ of x. The parties are not allowed to communicate, and despite having imperfectly correlated information they wish to agree on a single random bit. In other words, the i'th party will output a bit $f_i(y(i)) \in \{-1, 1\}$, and the goal is to find functions $f_1, \ldots, f_q$ that maximize the probability that $f_1(y^{(1)}) = f_2(y^{(2)}) = \cdots = f_q(y^{(q)})$. To avoid trivial deterministic solutions, we insist that $\mathbb{E}\left[[f_i^{(y(j))}]\right] =$ for all $j \in [q]$.

(a) Suppose $q = 2$, $\rho \in (0, 1)$, and $y^{(j)} \sim N_\rho(x)$ independently for each j. Show that the optimal solution is $f_1 = f_2 = \pm\chi_i$ for some $i \in [n]$.

(b) Show the same result for $q = 3$.

(c) Let $q = 2$ and $\rho \in (1/2, 1)$. Suppose that $y^{(1)} = x$ exactly, but $y^{(2)} \in \{-1, 0, 1\}^n$ has erasures: its formed from x by setting $y_i^{(2)} = x_i$ with probability ρ and $y_i^{(2)} = 0$ with probability $1 - \rho$, independently for all $i \in [n]$. Show that the optimal success probability is $1/2 + \rho/2$ and there is an optimal solution in which $f_1 = \pm\chi_i$ for any $i \in [n]$.

(d) Consider the previous scenario but with $\rho \in (0, 1/2)$. Show that if n is sufficiently large, then the optimal solution does not have $f_1 = \pm\chi_i$

Solution 56.

(a) First lets analyze the solution $f_1 = f_2 = \pm\chi_i$. The players output the same both if the i'th bit was not flipped for both, or flipped for both. The probability for that event is,

$$\left(\frac{1+\rho}{2}\right)^2 + \left(\frac{1-\rho}{2}\right)^2 = \frac{\rho^2+1}{2}$$

Let f_1, f_2 arbitrary boolean functions. We need to bound the probability,

$$\Pr_{y \sim N_\rho(x)}\left[f_1(y) = f_2(y)\right] =$$

Equivalently it suffice to analyze,

$$\langle T_\rho f_1, T_\rho f_2\rangle = \mathbb{E}_x\left[f_1(y) f_2(y)\right] = 2\Pr_x\left[f_1(x) = f_2(x)\right] - 1$$

To prove that dictator is optimal we need to show that,

$$\frac{\langle T_\rho f_1, T_\rho f_2\rangle + 1}{2} \leq \frac{\rho^2+1}{2}$$

By Parseval,

$$\langle T_\rho f_1, T_\rho f_2\rangle = \sum_{k>0} \rho^{2k} \widehat{f_1}(S)\widehat{f_2}(S)$$

Note that we used that f_1, f_2 are unbiased. To get ρ^2 in that expression with coefficient 1 we must have $\sum_{|S|=1} \widehat{f_1}(S)\widehat{f_2}(S) = 1$. By Cauchy Schwartz,

$$1 = \left(\sum_{|S|=1} \widehat{f_1}(S)\widehat{f_2}(S)\right)^2 \leq \left(\sum_{|S|=1} \widehat{f_1}(S)^2\right)\left(\sum_{|S|=1} \widehat{f_2}(S)^2\right) \leq 1$$

Thus, it must be that,

$$\sum_{|S|=1} \widehat{f_1}(S)^2 = \sum_{|S|=1} \widehat{f_2}(S)^2 = 1$$

This means that $f_1 = \pm\chi_i$, $f_1 = \pm\chi_j$. It is not hard to see that we also need to have $i = j$.

(b) The probability that $f_1 = f_2 = f_3 = \pm\chi_i$ succeeds is given by,

$$\left(\frac{1+\rho}{2}\right)^3 + \left(\frac{1-\rho}{2}\right)^3 = \frac{3\rho^2+1}{4}$$

We shall use the Fourier expansion for the equality function,

$$\text{Equal}(x_1, x_2, x_3) = \frac{1}{2}(x_1x_2 + x_2x_3 + x_1x_3 - 1)$$

with $\text{Equal}(x_1, x_2, x_3) = 1$ if $x_1 = x_2 = x_3$ and -1 otherwise. Note that,

$$\mathbb{E}_x\left[\text{Equal}(T_\rho f_1(x), T_\rho f_2(x), T_\rho f_3)(x)\right] = 2\text{Pr}_{y\sim N_\rho(x)}\left[f_1(y) = f_2(y) = f_3(y)\right] - 1$$

Therefore, it suffice to show that,

$$\frac{\mathbb{E}_x\left[\text{Equal}(T_\rho f_1(x), T_\rho f_2(x), T_\rho f_3(x)\right] + 1}{2} \leq \frac{3\rho^2+1}{4}$$

Direct calculation,

$$\begin{aligned}
\text{Pr}_{y\sim N_\rho(x)}\left[f_1(y) = f_2(y) = f_3(y)\right] &= \frac{\mathbb{E}_x\left[\text{Equal}(T_\rho f_1(x), T_\rho f_2(x), T_\rho f_3(x)\right] + 1}{2} \\
&= \frac{1}{4}\cdot\mathbb{E}_x\left[T_\rho f_1(x)T_\rho f_2(x) + T_\rho f_2(x)T_\rho f_3(x) + T_\rho f_1(x)T_\rho f_3(x)\right] + \frac{1}{4} \\
&= \frac{1}{4}\cdot\sum_{k>0}\rho^{2k}(\widehat{f_1}(S)\widehat{f_2}(S) + \widehat{f_2}(S)\widehat{f_3}(S) + \widehat{f_1}(S)\widehat{f_3}(S)) + \frac{1}{4}
\end{aligned}$$

The argument is now just as before, concluding that all Fourier mass of f_1, f_2, f_3 must lie in the first level.

(c) The optimal solution is $f_1 = \pm\chi_i$ and,

$$f_2(y) = \begin{cases} y_i & y_i \neq 0 \\ 1 & y_i = 0 \end{cases}$$

This approach succeeds with probability $\rho + \frac{1}{2}(1-\rho) = \frac{1}{2} + \frac{1}{2}\rho$. Alternatively, f_2 may choose a random bit. This is equivalent to the standard case with $q = 2$ as we may think of the erased bits as "re-sampling". That is, if a bit is erased we replace it with a random bit. This yields a ρ-correlated inputs for y_1, y_2.

(d) Majority? Tribes?

Problem 57. Let $g : \{-1,1\}^n \to \mathbb{R}^{\geq 0}$ such that $\mathbb{E}_x\left[g(x)\right] = \delta$.

(a) Show that for any $\rho \in [0,1]$

$$\rho \sum_{j=1}^{n} |\widehat{g}(\{j\})| \leq \delta + \sum_{k=2}^{n} \rho^k ||g^{=k}||_\infty$$

(b) Assume further that $g : \{-1,1\}^n \to \{0,1\}$. Show that $||g^{=k}||_\infty \leq \sqrt{\delta\binom{n}{k}}$. Deduce that,

$$\rho \sum_{j=1}^{n} |\widehat{g}(\{j\})| \leq \delta + 2\rho^2\sqrt{\delta}n$$

assuming $\rho \leq \frac{1}{2\sqrt{n}}$.

(c) Assume further $\delta \leq 1/4$. Show that $\sum_{j=1}^{n} |\widehat{g}(\{j\})| \leq 2\sqrt{2}\delta^{3/4}\sqrt{n}$. Deduce that $W^1\left[g\right] \leq 2\sqrt{2}\delta^{7/4}\sqrt{n}$.

(d) Suppose $f : \{-1,1\}^n \to \{-1,1\}$ is monotone and $\mathrm{MaxInf}\left[f\right] \leq \delta$. Show that $W^2\left[f\right] \leq \sqrt{2}\delta^{3/4}\mathrm{Inf}\left[f\right] \cdot \sqrt{n}$.

(e) Suppose further that f is unbiased. Show that $\mathrm{MaxInf}\left[f\right] \leq o(n^{-2/3})$ implies $\mathrm{Inf}\left[f\right] \leq 3 - o(1)$. Conclude $\mathrm{MaxInf}\left[f\right] \geq \frac{3}{n} - o(1/n)$.

Solution 57. Let $g : \{-1,1\}^n \to \mathbb{R}^{\geq 0}$ such that $\mathbb{E}_x\left[g(x)\right] = \delta$.

(a) Recall that $g \geq 0$ implies $T_\rho g \geq 0$ pointwise. Using the Fourier expansion,

$$0 \leq \delta + \rho \sum_{j=1}^{n} x_i \widehat{g}(\{j\}) + \sum_{k=2}^{\rho} \rho^k g^{=k}(x)$$

for any $x \in \{-1,1\}^n$. Substitute $x_j = -\mathrm{sgn}\left(\widehat{g}(\{j\})\right)$ then,

$$\rho \sum_{j=1}^{n} |\widehat{g}(\{j\})| \leq \delta + \rho^2 \sum_{k=2}^{\rho} \rho^{k-2} g^{=k}(x)$$

To conclude that for $k \geq 2$, $\rho \geq 0$,

$$\rho^{k-2} g^{=k}(x) \leq ||g^{=k}||_\infty$$

(b) By Parseval,

$$\mathbb{E}_x\left[g(x)^2\right] = \sum_S \widehat{g}(S)^2$$

Using that $g(x) \in \{0,1\}$ we have $g(x) = g(x)^2$ hence,

$$\delta = \mathbb{E}_x\left[g(x)\right] = \sum_S \widehat{g}(S)^2$$

In particular, any Fourier coefficient is not larger in magnitude than $\sqrt{\delta}$. Every Fourier level has $\binom{n}{k}$ coefficients hence the bound,

$$||g^{=k}||_\infty \leq \sqrt{\delta\binom{n}{k}}$$

Revising the argument in (a) we are looking to bound,

$$\begin{aligned}
\rho^{k-2}||g^{=k}||_\infty &\leq \sqrt{\delta}\cdot 2^{-k+2}n^{-k/2+1}\sqrt{\binom{n}{k}} \\
&\leq \sqrt{\delta}\cdot 2^{-k+2}n^{-k/2+1}n^{k/2} \\
&= n\sqrt{\delta}\cdot 2^{-k+2}
\end{aligned}$$

Thus,

$$\rho\sum_{j=1}^{n}|\widehat{g}(\{j\})| \leq \delta + \rho^2\sqrt{\delta}n\sum_{k=2}^{\rho}\rho^{k-2}2^{-k+2} \leq \delta + 2\rho^2\sqrt{\delta}n$$

(c) We have the upper bound,

$$\sum_{j=1}^{n}|\widehat{g}(\{j\})| \leq 2\sqrt{n}\delta + \delta\sqrt{n}$$

It suffice to show that,

$$2\delta + \sqrt{\delta} \leq 2\sqrt{2}\delta^{3/4}$$

Since $\delta \leq 1/4$ then $\delta^{1/4} \leq 1/\sqrt{2}$ and so,

$$2\delta + \sqrt{\delta} = \delta^{3/4}(2\delta^{1/4} + \delta^{-1/4})$$

(d) Suppose $f : \{-1,1\}^n \to \{-1,1\}$ is monotone and $\mathrm{MaxInf}\left[f\right] \leq \delta$. Show that $W^2\left[f\right] \leq \sqrt{2}\delta^{3/4}\mathrm{Inf}\left[f\right]\cdot\sqrt{n}$.

(e) Suppose further that f is unbiased. Show that $\mathrm{MaxInf}\left[f\right] \leq o(n^{-2/3})$ implies $\mathrm{Inf}\left[f\right] \leq 3 - o(1)$. Conclude $\mathrm{MaxInf}\left[f\right] \geq \frac{3}{n} - o(1/n)$.

Problem 58. ....

Solution 58.

3 Spectral Structure And Learning

Problem 1. Let $M : \mathbb{F}_2^n \to \mathbb{F}_2^n$ invertible linear transformation and $f : \mathbb{F}_2^n \to \mathbb{R}$. Define $f \circ M(x) = f(Mx)$. Show that $\widehat{f \circ M}(\gamma) = \widehat{f}\left(M^{-1}\gamma\right)$. What if M is affine? What is M is not invertible?

Solution 1. Let M, f as above.

$$\begin{aligned}
\widehat{f \circ M}(\gamma) &= \langle f \circ M, \chi_\gamma \rangle \\
&= \mathbb{E}_x \left[f(Mx) \cdot (-1)^{\langle \gamma, x \rangle} \right] \\
&= \mathbb{E}_x \left[f(x) \cdot (-1)^{\langle \gamma, M^{-1}x \rangle} \right] \\
&= \mathbb{E}_x \left[f(x) \cdot (-1)^{\langle M^{-T}\gamma, x \rangle} \right] \\
&= \widehat{f}\left(M^{-T}\gamma\right)
\end{aligned}$$

If M is affine, then $M(x) = M'(x) + b$ where $b \in \mathbb{F}_2^n$ then $f \circ M = f \circ M'^{+b}$ and so $\widehat{f \circ M}(\gamma) = (-1)^{\gamma \cdot b} \widehat{f}\left(M^{-T}\gamma\right)$. If M is not invertible, we can transform it to the following canonical form,

$$M \circ T(x_1, \ldots, x_n) = (x_1, \ldots, x_k, 0, \ldots, 0)$$

where k is the rank of M. This is the same as restriction and so we can find its Fourier coefficients easily and hence we can find the Fourier coefficients of $f \circ M \circ T$. We can then find the coefficients of $f \circ M$ by applying T^{-1} and using the above.

Problem 2. Show that $\frac{1-e^{-2}}{2}$ is smallest constant (not depending on δ or n) that can be taken in Proposition 3.3.

Solution 2. Consider the function $f(x) = \prod_{i=1}^k x_i$ with noise stability approaching to $\frac{1-e^{-2}}{2}$ (as $k \to \infty$). This function is not ϵ-concentrated for any $\epsilon < 1$ up to degree k.

Problem 3. Let $f : \{-1, 1\}^n \to \mathbb{R}$. Show that for $\epsilon = (\mathbb{E}_x \left[f(x)^2\right] - \mathrm{Stab}_{1-\delta}\left[f\right])(1 - e^{-1})$, f is ϵ-concentrated on degree up to $1/\delta$.

Solution 3. Observe that,

$$\begin{aligned}
\epsilon &= \mathbb{E}_x \left[f(x)^2 \right] - \mathrm{Stab}_{1-\delta}\left[f\right] \\
&\le \sum_k W^k\left[f\right](1 - (1-\delta)^k) \\
&\le \mathbb{E}_x \left[f(x)^2 \right] - \mathrm{Stab}_{1-\delta}\left[f\right] \\
&= \sum_{k \ge 1/\delta} W^k\left[f\right](1 - (1-\delta)^{1/\delta}) \\
&\le (1 - e^{-1}) \sum_{k \ge 1/\delta} W^k\left[f\right]
\end{aligned}$$

Problem 4. Prove by induction that if $\deg(f) \leq k$ for $f : \{-1,1\}^n \to \mathbb{R}$ nonzero then $\Pr_x [f(x) \neq 0] \geq 2^{-k}$.

Solution 4. For $n = 1$ the theorem is trivial for every k. Let $f : \{-1,1\}^n \to \mathbb{R}$ nonzero with $\deg(f) \leq k$. Define,

$$g^+(x_1, \ldots, x_{n-1}) = f(x_1, \ldots, x_{n-1}, 1)$$

$$g^-(x_1, \ldots, x_{n-1}) = f(x_1, \ldots, x_{n-1}, -1)$$

Then,

$$\Pr_x [f(x) \neq 0] = \frac{1}{2}\Pr_x \left[g^+(x) \neq 0\right] + \frac{1}{2}\Pr_x \left[g^-(x) \neq 0\right]$$

Since f is nonzero either g^+ or g^- is nonzero (or both). If none are nonzero then we are done by induction hypothesis. Otherwise, suppose that g^- is identically zero then,

$$f(x_1, \ldots, x_{n-1}, x_n) = \frac{1 + x_n}{2} g^+(x_1, \ldots, x_{n-1})$$

Note that since we took out one variable then $\deg(g^+) \geq k - 1$ and so the bound following from induction hypothesis.

Problem 5. Verify that $\hat{||} f \hat{||}_p$ is a norm on the vector space of functions $f : \{-1,1\} \to \mathbb{R}$ for any $p \in [1, \infty]$.

Solution 5. As vector spaces, we have that the space of functions $f : \{-1,1\} \to \mathbb{R}$ is isomorphic to $\mathbb{R}^n$ via the isomorphism,

$$f \to \widehat{f}$$

Under this isomorphism, $\hat{||} f \hat{||}_p$ is just the $||\cdot||_p$ on $\mathbb{R}^n$.

Problem 6. Show that $\hat{||} fg \hat{||}_1 \leq \hat{||} f \hat{||}_1 \cdot \hat{||} g \hat{||}_1$.

Solution 6. Recall the formula for the product of Fourier expansion,

$$\begin{aligned}
\hat{||} fg \hat{||}_1 &= \sum_S \left| \widehat{fg}(S) \right| \\
&= \sum_S \left| \sum_{T_1 \Delta T_2} \widehat{f}(T_1)\widehat{g}(T_2) \right| \\
&\leq \sum_S \sum_{T_1 \Delta T_2} \left| \widehat{f}(T_1) \right| \cdot |\widehat{g}(T_2)| \\
&= \left(\sum_{T_1} \left| \widehat{f}(T_1) \right| \right) \cdot \left(\sum_{T_2} \left| \widehat{f}(T_2) \right| \right) \\
&= \hat{||} f \hat{||}_1 \cdot \hat{||} g \hat{||}_1
\end{aligned}$$

Problem 7. Let $f : \{-1,1\}^n \to \mathbb{R}$, $J \subseteq [n]$ and $z \in \{-1,1\}^{\complement J}$.

(a) Show that restriction reduces spectral 1-norm,

$$\hat{||} f_{J|_z} \hat{||}_1 \leq \hat{||} f \hat{||}_1$$

(b) Show that it also reduces Fourier sparsity,

$$\text{sparsity}\left(\widehat{f_{J|_z}}\right) \leq \text{sparsity}\left(\widehat{f}\right)$$

Solution 7. For (a) use the formula for the Fourier coefficients of restrictions,

$$\begin{aligned}
\hat{||} f_{J|_z} \hat{||}_1 &= \sum_{S\subseteq J} \left|\widehat{f_{J|_z}}(S)\right| \\
&= \sum_{S\subseteq J} \left|\sum_{T\subseteq \overline{J}} \widehat{f}(S\cup T) z^T\right| \\
&\leq \sum_{S\subseteq J}\sum_{T\subseteq \overline{J}} \left|\widehat{f}(S\cup T)\right| \\
&= \sum_{W\subseteq [n]} \left|\widehat{f}(W)\right|
\end{aligned}$$

Note that every set $W \subseteq [n]$ can be written as $W = S \cup T$ where $S \subseteq J$, $T \subseteq \overline{J}$. The argument for (b) is as follows. Suppose that $\widehat{f_{J|_z}}(S) \neq 0$ for some $S \subseteq J$, then it follows that $\widehat{f}(W) \neq 0$ for at least one W of the form $W = S \cup T$ where $T \subseteq \overline{J}$.

Problem 8. Let $f : \{-1,1\}^n \to \mathbb{R}$ and $0 < p \leq q \leq \infty$. Show that,

$$\hat{||} f \hat{||}_p \leq \hat{||} f \hat{||}_q$$

Solution 8. From Jensen inequality,

$$\left|\sum |a_i|^p\right|^{1/p} \leq \left|\sum |a_i|^q\right|^{1/q}$$

for any sequence $a_1, \ldots, a_N \in \mathbb{R}$. Taking $q \to \infty$ yield the inequality for $q = \infty$.

Problem 9. Let $f : \{-1,1\}^n \to \mathbb{R}$. Show that $\hat{||} f \hat{||}_\infty \leq ||f||_1$, $||f||_\infty \leq \hat{||} f \hat{||}_1$.

Solution 9. For the first inequality,

$$\begin{aligned}
\hat{||} f \hat{||}_\infty &= \max_S \left|\widehat{f}(S)\right| \\
&= \max_S \left|\mathbb{E}_x\left[f(x)\chi_S(x)\right]\right| \\
&= \max_S \mathbb{E}_x\left[|f(x)|\right] \\
&= \mathbb{E}_x\left[|f(x)|\right] \\
&= ||f||_1
\end{aligned}$$

For the second inequality,

$$\begin{aligned}
||f||_\infty &= \max_x |f(x)| \\
&= \max_x \left| \sum_S \widehat{f}(S)\chi_S(x) \right| \\
&= \max_x \sum_S \left| \widehat{f}(S) \right| \\
&= \sum_S \left| \widehat{f}(S) \right| \\
&\leq \sum_S \left| \widehat{f}(S) \right| \\
&= \hat{|}|f|\hat{|}_1
\end{aligned}$$

Problem 10. Let $f : \{-1,1\}^n \to \{-1,1\}$ be monotone. Show that $\left| \widehat{f}(S) \right| \leq \widehat{f}(\{i\})$ where $i \in S$. Deduce that $\hat{|}|f|\hat{|}_\infty$ is achieved by sets of cardinality either 0 or 1.

Solution 10. Let $f : \{-1,1\}^n \to \{-1,1\}$ be monotone and $i \in [n]$. Since f is monotone then $D_i f \geq 0$. Appolying 3 on $D_i f$ we have,

$$\begin{aligned}
\hat{|}|D_i f|\hat{|}_\infty &= \max_S \left| \widehat{D_i f}(S) \right| \\
&= \max_{i \in S} \left| \widehat{f}(S) \right| \\
&\leq \max_{i \in S} \mathbb{E}_x \left[|D_i f(x) \chi_S(x)| \right] \\
&= \mathbb{E}_x \left[D_i f(x) \right] \\
&= \widehat{f}(\{i\})
\end{aligned}$$

and so,

$$\left| \widehat{f}(S) \right| \leq \widehat{f}(\{i\})$$

for any $i \in S$. As a conclusion, the maximum Fourier coefficient is either in the first level, or in the zero level.

Problem 11. Let $H + a$ affine space of $\mathbb{F}_2^n$ with co-dimension k. Show that,

$$\widehat{1_A}(\gamma) = \begin{cases} \chi_\gamma(a) 2^{-k} & \gamma \in H^\perp \\ 0 & o.w \end{cases}$$

Conclude that $\varphi_A = \sum_{\gamma \in H^\perp} \chi_\gamma(a) \chi_\gamma$, sparsity $\left(\widehat{1_A} \right) = 2^k$, 1_A is 2^{-k}-granular , $\hat{|}|1_A|\hat{|}_\infty = 2^{-k}$,$\hat{|}|1_A|\hat{|}_1 = 1$.

Solution 11. First we claim that,

$$\sum_{x\in H}(-1)^{\langle\gamma,x\rangle}\begin{cases}|H| & \gamma\in H^{\perp}\\ 0 & \gamma\notin H^{\perp}\end{cases}$$

If $\gamma \in H^{\perp}$ then $\langle\gamma, x\rangle = 0$ for any $x \in H$ and so the inequality follows. If $\gamma \notin H^{\perp}$ then there exists $y \in H$ s.t. $\langle\gamma, y\rangle = 1$. We can pair elements in H to $x + y, x \in H$ and note that $\langle\gamma, x\rangle \neq \langle\gamma, x + y\rangle$ thus summing those in pairs leads to cancellation and so the sum is zero. The Fourier expansion follows form simple calculation,

$$\begin{aligned}\widehat{1_A}(\gamma) &= \mathbb{E}_x\left[1_A(x)\chi_\gamma(x)\right]\\ &= 2^{-n}\sum_{x\in H+a}\chi_\gamma(x)\\ &= 2^{-n}\sum_{x\in H}\chi_\gamma(x+a)\\ &= 2^{-n}\chi_\gamma(a)\sum_{x\in H}\chi_\gamma(x)\\ &= \begin{cases}\chi_\gamma(a)2^{-k} & \gamma\in H^{\perp}\\ 0 & o.w\end{cases}\end{aligned}$$

The conclusions follows directly.

Problem 12. Verify Parsevals Theorem for the Fourier expansion of subspaces,

$$1_A = \sum_{\gamma\in A^{\perp}} 2^{-k}\chi_\gamma$$

where A is linear subspace with co-dimension k.

Solution 12. This is as the previous exercise with $a = 0$.

Problem 13. Let $f = 1_A$. We know that if A is affine then $\hat{||}f\hat{||}_1 = 1$. So assume that A is not an affine subspace.

(a) Show that there exists affine subspace B with dimension 2 such that $|A \cap B| = 3$.

(b) Let $b \in B \setminus A$. Define $\psi = \varphi_B - \frac{1}{2}\varphi_b$ and show that $\hat{||}\psi\hat{||}_\infty = 1/2$.

(c) Show that $\langle f, \psi\rangle = \frac{3}{4}$ and conclude that $\hat{||}f\hat{||}_1 \geq 3/2$.

Solution 13. Recall that A is affine iff,

$$x, y, z \in A \Rightarrow x + y + z \in A$$

Thus, we conclude that there exists $x, y, z \in A$ s.t. $x + y + z \notin A$. Consider the affine subspace $B = z + \mathrm{Span}\{x, y\} = \{z, x, y, x + y + z\}$.

(a) Clearly B is with dimension 2 and $|B \cap A| = |\{x, y, z\}| = 3$.

(b) Recall that $\varphi_B(\gamma) = 1$ if $\gamma \in B^\perp$ and zero otherwise. Also, $\varphi_b(\gamma) = (-1)^{\langle \gamma, b \rangle}$. It follows that if $\varphi_B(\gamma) = 1$ then $\varphi_b(\gamma) = 1$ then $\varphi_B(\gamma) - \frac{1}{2}\varphi_b(\gamma) \le 1/2$ for any γ. Taking $\gamma \in A^\perp$ shows that $\hat{||}\psi\hat{||}_\infty = 1/2$.

(c) Since B contains exactly 3 points from A and $|B| = 4$ then $\langle f, \varphi_B \rangle = 3/4$. Similarly, since $b \notin A$ then $\langle f, \varphi_b \rangle = 0$ and so $\langle f, \psi \rangle = 3/4$. Using Parseval,

$$\langle \psi, f \rangle = \sum_S \widehat{f}(S)\widehat{\psi}(S)$$

$$\le \max_S \left|\widehat{\psi}(S)\right| \sum \left|\widehat{f}(S)\right|$$

$$= \hat{||}\psi\hat{||}_\infty \cdot ||f||_1$$

$$= \frac{1}{2} \cdot ||f||_1$$

It follows that,

$$3/4 \le \frac{1}{2} \cdot ||f||_1 \Rightarrow ||f||_1 \ge 3/2$$

Problem 14. Let $f : \{-1, 1\}^n \to \mathbb{R}$ with $\mathbb{E}_x\left[f(x)^2\right] \le 1$. Show that $\hat{||}f\hat{||}_1 \le 2^{n/2}$ and that for n which is even the bound is achieved by a boolean function.

Solution 14. Follows from Cauchy Schwartz inequality. The inequality is tight for the inner product function.

Problem 15. Given $f : \mathbb{F}_2^n \to \mathbb{R}$, define its (fractional) sparsity to be,

$$\text{sparsity}(f) = \Pr_x\left[f(x) \ne 0\right]$$

In this exercise you will prove the uncertainty principle: If f is nonzero, then $\text{sparsity}(f) \cdot \text{sparsity}\left(\widehat{f}\right) \ge 1$.

(a) Show that we may assume $||f||_1 = 1$.

(b) Suppose $\mathscr{F} = \{\gamma : \widehat{f}(\gamma) \ne 0\}$. Show that $\hat{||}f\hat{||}_2^2 \le |\mathscr{F}|$.

(c) Suppose $\mathscr{G} = \{x : f(x) \ne 0\}$. Show that $\hat{||}f\hat{||}_2^2 \ge 2^n / |\mathscr{G}|$. Deduce the uncertainty principle.

(d) Identify all cases of equality.

Solution 15.

(a) We may assume $||f||_1 = 1$ since the equation in question is homogeneous. That is,

$$\text{sparsity}\,(cf) \cdot \text{sparsity}\left(\widehat{cf}\right) = \text{sparsity}\,(f) \cdot \text{sparsity}\left(\widehat{f}\right)$$

for any nonzero c. Henceforth, assume $||f||_1 = 1$.

(b) Note that,

$$\max_x |f(x)| \le \max_x \left| \sum_{S \in \mathscr{F}} \widehat{f}(S) \chi_S(x) \right| \le |\mathscr{F}|$$

Thus,

$$\begin{aligned} \hat{||} f \hat{||}_2^2 &= \mathbb{E}_x \left[f(x)^2 \right] \\ &\le \max_x |f(x)| \, \mathbb{E}_x \left[|f(x)| \right] \\ &\le |\mathscr{F}| \end{aligned}$$

(c) Using Cauchy-Schwartz inequality.

(d) We need both inequalities to be tight. The first inequality is tight iff all nonzero Fourier coefficients have the same magnitude and the second (which is Cauchy-Schwartz) is tight iff all nonzero values of $f(x)$ have equal magnitude. If $f = 0$ then the inequality is not tight of course, then f takes exactly one nonzero value. Thus, (after normalization) $f = 1_A$ for some $A \subseteq \mathbb{F}_2^n$. This shows that the inequality holds for affine spaces and indeed these are the only functions in which the inequality is in fact equality.

Lemma 3.1. *A is an affine subspace.*

Proof. First note that since sparsity $\left(\widehat{f}\right)$ is an integer then sparsity $(f) \cdot$ sparsity $\left(\widehat{f}\right) = 1$ suggests sparsity $\left(\widehat{f}\right) = 2^\ell$ for some non negative integer ℓ, i.e $|A| = 2^k$ for some non negative integer k. Let $x, y, z \in A$ then,

$$\sum_\gamma \widehat{1_A}(\gamma)(-1)^{\langle x, \gamma \rangle} = 1$$

$$\sum_\gamma \widehat{1_A}(\gamma)(-1)^{\langle y, \gamma \rangle} = 1$$

$$\sum_\gamma \widehat{1_A}(\gamma)(-1)^{\langle z, \gamma \rangle} = 1$$

Multiplying,

$$\left(\sum_\gamma \widehat{1_A}(\gamma)(-1)^{\langle x, \gamma \rangle} \right) \cdot \left(\sum_\gamma \widehat{1_A}(\gamma)(-1)^{\langle y, \gamma \rangle} \right) \cdot \left(\sum_\gamma \widehat{1_A}(\gamma)(-1)^{\langle z, \gamma \rangle} \right) = 1$$

Set $\mathscr{F} = \{\gamma : \widehat{1_A}(\gamma) \neq 0\}$ and t the magnitude of the nonzero Fourier coefficients. In fact, the Fourier coefficient of $\emptyset$ is $\mathbb{E}_x[1_A(x)] = \frac{2|A|-2^n}{2^n}$. Splitting the sum,

$$\sum_{\substack{\gamma_1,\gamma_2,\gamma_3\in\mathscr{F}\\ \text{not all equal}}} (\pm t^3)\cdot(-1)^{\langle x,\gamma_1\rangle+\langle y,\gamma_2\rangle+\langle z,\gamma_3\rangle} + \sum_\gamma \widehat{1_A}(\gamma)^3(-1)^{\langle x+y+z,\gamma\rangle} = 1$$

The plus-minus on the left means that t^3 could be either positive or negative (depending on $\gamma_1,\gamma_2,\gamma_3$). Taking out t^2,

$$t^2 \sum_{\substack{\gamma_1,\gamma_2,\gamma_3\in\mathscr{F}\\ \text{not all equal}}} (\pm t)\cdot(-1)^{\langle x,\gamma_1\rangle+\langle y,\gamma_2\rangle+\langle z,\gamma_3\rangle} + t^2\sum_\gamma \widehat{1_A}(\gamma)(-1)^{\langle x+y+z,\gamma\rangle} = 1$$

Or equivalently,

$$t^3 \sum_{\substack{\gamma_1,\gamma_2,\gamma_3\in\mathscr{F}\\ \text{not all equal}}} (\pm 1)\cdot(-1)^{\langle x,\gamma_1\rangle+\langle y,\gamma_2\rangle+\langle z,\gamma_3\rangle} + t^2\cdot 1_A(x+y+z) = 1$$

Assume towards contradiction that the right term is zero, then we have the equality,

$$Ct^3 = 1$$

For some integer $C \in \mathbb{Z}$ so we must have,

$$C = \frac{1}{t^3} = \left(\frac{2^{n-1}}{2^k - 2^{n-1}}\right)^3$$

In order for C to be integer we need either $k = n$ (and so $1_A = 1$ is the constant function) or $k = n-2$. We can ignore the case where $1_A = 1$ is the constant function (since then $A = \mathbb{F}_2^n$ is an affine subspace) so lets assume $k = n-2$. The number of summands in the left term is exactly $|A|^3 - |A|$ and so C is bounded by,

$$4\cdot 2^{3(n-1)} = |C| \leq |A|^3 - |A| < 2^{3k} = 2^{3(n-2)}$$

contradiction to the assumption that the right term is zero. We conclude that,

$$\sum_\gamma \widehat{1_A}(\gamma)(-1)^{\langle x+y+z,\gamma\rangle} = 1$$

and therefore $x+y+z \in A$. This is true for any x, y, z and so A is an affine subspace. $\square$

Problem 16. Let $f : \{-1,1\}^n \to \mathbb{R}$ and $\epsilon > 0$. Show that f is ϵ-concentrated on $\mathscr{F}$ with $|\mathscr{F}| \leq \hat{||}f\hat{||}_1^2/\epsilon$.

Solution 16. Take $\mathscr{F} = \{S : \left|\widehat{f}(S)\right| \geq \epsilon/\hat{||}f\hat{||}_1\}$. We claim that,

$$|\mathscr{F}| \leq \hat{||}f\hat{||}_1^2/\epsilon$$

Otherwise, $\mathscr{F}$ contributes to the Fourier norm more than,

$$\left(\hat{||}f\hat{||}_1^2/\epsilon\right) \cdot \left(\epsilon/\hat{||}f\hat{||}_1\right) = \hat{||}f\hat{||}_1$$

contradiction. Moreover,

$$\sum_{S \in \overline{\mathscr{F}}} \widehat{f}(S)^2 \leq \left(\epsilon/\hat{||}f\hat{||}_1\right) \cdot \sum_{S \in \overline{\mathscr{F}}} \left|\widehat{f}(S)\right| \leq \epsilon$$

Problem 17. Suppose that the Fourier spectrum of $f : \{-1,1\}^n \to \mathbb{R}$ is ϵ_1-concentrated on $\mathscr{F}$ and that $g : \{-1,1\}^n \to \mathbb{R}$ satisfies $||f - g||_2^2 \leq \epsilon_2$. Then, g is $2(\epsilon_1 + \epsilon_2)$ is concentrated

Solution 17. Consider the following inequality for any a, b

$$\begin{aligned} b^2 &\leq b^2 + 4\left(a - \frac{b}{2}\right)^2 \\ &= b^2 + 4a^2 - 4ab + b^2 \\ &= 2 \cdot (a^2 + (a-b)^2) \end{aligned}$$

Applying this inequality with $a = \widehat{f}(S)$, $b = \widehat{g}(S)$ and summing for all $S \in \mathscr{F}$

$$\begin{aligned} \sum_{S \notin \mathscr{F}} \widehat{g}(S)^2 &\leq \sum_{S \notin \mathscr{F}} 2 \cdot \left(\widehat{f}(S)^2 + \left(\widehat{f}(S) - \widehat{g}(S)\right)^2\right) \\ &= 2 \cdot \sum_{S \notin \mathscr{F}} \widehat{f}(S)^2 + 2 \cdot \sum_{S \notin \mathscr{F}} (\widehat{f}(S) - \widehat{g}(S))^2 \\ &\leq 2(\epsilon_1 + \epsilon_2) \end{aligned}$$

Problem 18. Show that any function $f : \{-1,1\}^n \to \mathbb{R}$ is computed by a descision tree of depth n and size 2^n.

Solution 18. Construct the complete binary tree and so every leaf corresponds to a single value of $f(x)$.

Problem 19. Let $f : \mathbb{F}_2^n \to \mathbb{R}$ computable by a descision tree with depth k and size s. Show that $-f$, $f^\dagger$ can be computed by a descision tree with depth k and size s.

Solution 19. Let T be a descision tree for f. For $-f$ just take T with minus signs on the leaves. For $f^\dagger$ switch the roles of 1 and 0 and change the signs at the leaves.

Problem 20. For each function in Exercise 1 with 4 or fewer inputs, give a decision tree computing it. Try primarily to use the least possible depth, and secondarily to use the least possible size.

Solution 20. ...

Problem 21. Let $f : \mathbb{F}_2^n \to \mathbb{R}$ computable by a descision tree with depth k and size s. Show that,

(a) $\deg(f) \leq k$

(b) sparsity $\left(\widehat{f}\right) \leq s2^k \leq 4^k$

(c) $\hat{||}f\hat{||}_1 \leq ||f||_\infty \cdot s \leq ||f||_\infty \cdot 2^k$

(d) Assume further that $f : \mathbb{F}_2^n \to \mathbb{Z}$, then f is 2^{-k} granular.

Solution 21. First note that since the descision tree is binary then $s \leq 2^k$. The important observation is that,

$$f(x) = \sum_P f(P) \cdot 1_P$$

where the summation is over all paths to leaves in the tree, $f(P)$ is the value at the leaf and 1_P is the indicator function corresponding to the sub-cube induced by the path P.

(a) The indicator function P corresponding to the sub-cube $\{x : x_{i_1} = a_1 \ , \ \ldots \ , \ a_{i_m} = a_m\}$ is given by,

$$1_P(x_1, \ldots, x_n) = \prod_{j=1}^{m} \frac{1 + (-1)^{a_j + x_{i_j}}}{2}$$

Its Fourier expansion is with degree exactly m which is also the depth of the path P. Therefore, f can be expressed as a sum of functions determined by at most k variables. We conclude that $\deg(f) \leq k$.

(b) Let P be any path. The Fourier coefficients in the expansion of 1_P has at most 2^k nonzero coefficients and there are at most s paths.

(c) Each path P contributes to the Fourier expansion (more accurately, its indicator function 1_P contributes) Fourier coefficients with absolute value summing to 1. Thus, every path P contributes at most $f(P) \leq ||f||_\infty$ to the total Fourier norm $||f||_1$. The bound follows since there are at most s paths.

(d) Follows from the formula,

$$f(x) = \sum_P f(P) \cdot 1_P$$

and that 1_P is 2^{-k} granular.

Problem 22. Let $f : \mathbb{F}_2^n \to \{-1, 1\}$ and a descision tree T computing f with size s and depth k. Show that for any $\epsilon > 0$, truncating T at depth $\log(s/\epsilon)$ yields a descision tree computing function ϵ-close to f (you may choose the values at the truncation point). Deduce that the spectrum of f is ϵ-concentrated on degree $\log(s/\epsilon)$.

Solution 22. Truncate T and choose values by majority voting (i.e, according to most values in that sub-tree). TLets examine a single truncation. Since the truncation happens at depth $\log(s/\epsilon)$ it corresponds to a sub-cube of variables s/ϵ hence consists of $\frac{\epsilon}{s} \cdot 2^n$ points in $\mathbb{F}_2^n$. This means that a single truncation spoils $\frac{\epsilon}{s}$ fraction of the points (actually half of that since we take the majority) and as there are at most s paths the result follows. To conclude that f is ϵ-concentrated on degree $\log(s/k)$ note that g is 0-concentrated at that level and $|f-g|\,2 = \Pr_x\left[f(x) = g(x)\right] \leq \epsilon$ and use exercise 3.

Problem 22. A decision list is a decision tree in which every internal node has an outgoing edge to at least one leaf. Show that any function computable by a decision list is a linear threshold function.

Solution 22. Decision list corresponds to the following structure:

$$
\begin{aligned}
&\text{If } x_{i_1} = \delta_1 \text{ output } \gamma_1\\
&\text{Else if } x_{i_2} = \delta_2 \text{ output } \gamma_1\\
&\text{Else if } x_{i_3} = \delta_3 \text{ output } \gamma_2\\
&\vdots\\
&\text{Else output } \gamma_N
\end{aligned}
$$

with $\delta_i, \gamma_i = \pm 1$ and $\delta_N = -\delta_{N-1}$. Replace the k'th rule with $\gamma_k(2^{N-k} + \delta_k x_{i_k})$ where the plus-minus is chosen such that $\gamma_k(2^{N-k} + \delta_k x_{i_k}) = \gamma_k 2^{N-k+1}$ in case the "if" condition holds and zero otherwise. The threshold function,

$$\operatorname{sgn}\left(\gamma_k(2^{N-k} + \delta_k x_{i_k})\right)$$

is equivalent to the function computed by the descision list. To see this, note that as long as the "if" condition does not hold, the contribution to the sum is zero. Once the contribution is nonzero, its sign is determined by γ_k and its magnitude is larger than the magnitude of all values ahead summed together.

Problem 24. Recall the following theorem: Let $f : \{-1,1\}^n \to \{-1,1\}$ with $\deg(f) \leq k$ then f is a $k2^{k-1}$ junta. A read-once decision tree is one in which every internal node queries a distinct variable. Bearing this in mind, show that the bound $k2^{k-1}$ in the above theorem cannot be reduced below $2^k - 1$.

Solution 24. Consider the full binary tree with depth k where each node corresponds to a distinct variable. It is $2^k - 1$ nodes hence its boolean function depends on $2^k - 1$ variables and since the tree is with depth k, the degree is also at most k.

Problem 25. Suppose that f is computed by a read-once decision tree in which every root-to-leaf path has length k and every internal node at the deepest level has one child (leaf) labeled -1 one one child labeled 1. Compute the influence of each coordinate on f , and compute $\mathrm{Inf}\,[f]$.

Solution 25. The function depends on exactly $2^k - 1$ variables. Every non-leaf variable has influence 1/2 (since every sub-tree has even number of ± 1) and every leaf has influence 1 (since changing its value changes the value of the function). Thus,

$$\mathrm{Inf}[f] = 2^{k-1} + \frac{1}{2}(2^{k-1} - 1) = 2^{k-1} + 2^{k-2} - \frac{1}{2}$$

Problem 26. The following are generalizations of decision trees:

- Sub-cube Partition: Collections of sub-cubes $C_1, \ldots, C_s$ that partition $\mathbb{F}_2^n$ each assigned a real value $b_1, \ldots, b_s$. Given $x \in \mathbb{F}_2^n$ the sub-cube tree outputs b_i. s is the "size" and $\max_i \operatorname{codim}(C_i)$ is its depth.
- Parity Decision Trees: Same as descision trees but nodes corresponds to the query $\gamma \cdot x = \pm 1$.
- Affine Partition: Collections of affine subspaces $C_1, \ldots, C_s$ that partition $\mathbb{F}_2^n$ each assigned a real value $b_1, \ldots, b_s$. Given $x \in \mathbb{F}_2^n$ the sub-cube tree outputs b_i. s is the "size" and $\max_i \operatorname{codim}(C_i)$ is its depth.

Show that,

(a) Show that subcube partition size/codimension and parity decision tree size/depth generalize normal decision tree size/depth, and are generalized by affine subspace partition size/codimension.

(b) Show that exercise 3 holds also for the generalizations, except that the statement about degree need not hold for parity decision trees and affine subspace partitions.

(c) Show that the class of functions with affine subspace partition size at most s is learnable from queries with error ϵ in time $\mathrm{poly}(n, s, 1/\epsilon)$.

Solution 26.

(a) Every path in descision tree is equivalent to a subcube (and its depth is the codimension). Note that in terms of size there is a slight difference since the size of subcube partition corresponds to the number of paths. However, in most cases all we needed is to bound the number of paths (and we gave the trivial upper bound, which is the size of the tree). For parity trees, we can replace the queries on x_i to the parity query $e_i \cdot x$ where e_i is the standard unit vector. Affine partition generalize both partition and parity trees, since every subcube and conjunction of parity queries corresponds to an affine subspace. There is no natural correspondence between depth/size of affine subspaces trees or parity trees to the standard descision tree.

(b) As remarked, subcube trees is almost the same as standard descision trees (with slight difference in the size complexity that does not affect the result). For affine trees (also applied to parity trees), we can use that the indicator of an affine subspaces has norm 1 (just as the indicator of the subcube) and its degree is exactly the codimension. From there, the result follows just the same.

(c) Use Goldreich-Levin (we assume that the function is boolean).

Problem 27. Recall that equality function $\text{Equ}(x_1, x_2, x_3)$ which is 1 iff $x_1 = x_2 = x_3$ and -1 otherwise.

(a) Show that $\deg(\text{Equ}) = 2$.

(b) Show that $\text{DT}(\text{Equ}) = 3$

(c) Show that Equ has a parity tree of codimension 2.

(d) For $d \in \mathbb{N}$ define $f : \{-1,1\}^{3^d} \to \{-1,1\}$ by $f = \text{Equ}^{\otimes d}$ (recursive equality). Show that $\deg(f) = 2^d$ and $\text{DT}(d) = 3^d$

Solution 27.

(a) Recall the Fourier expansion from exercise 1.

(b) Clearly there is a descision tree of depth 3 (since the function has only 3 variables). Depth 2 does not suffice (e.g, verify by hand or see general answer in (d)).

(c) First check parity of x_1, x_2, and then parity of x_2, x_3 (this is even a parity descision list).

(d) First note that $\deg(f) = 2^d$ by composing the Fourier expansion of Equ d times yields polynomial with degree 2^d. As, before clearly there is a descision tree for f with depth 3^d as f has 3^d variables. Assume towards contradiction that f has a descision tree with depth $< 3^d$, then there is a path corresponding to a cube with codimension larger than 0. Consider the subcube corresponding to the point $(1, 1, \ldots, 1)$ and so the tree accept every point in that subcube. If the codimension is not zero, it contains at least one "free-variable" and we can flip it to obtain a point that should not be accepted.

Problem 28. Let $f : \{-1,1\}^n \to \mathbb{R}$, $J \subseteq [n]$ and define,

$$f^{\subseteq J}(x) = \mathbb{E}_{y \sim \{-1,1\}^{\overline{J}}} \left[f(x_J, y) \right]$$

where x_J is the projection of x to the coordinates of J. Verify the Fourier expansion,

$$f^{\subseteq J} = \sum_{S \subseteq J} \widehat{f}(S)$$

Solution 28. Recall the notation $f_{J|_z} : \{-1,1\}^J \to \mathbb{R}$ for fixing $\overline{J}$ to z. This time we fix J and so,

$$f^{\subseteq J}(x) = \mathbb{E}_{y \sim \{-1,1\}^{\overline{J}}} \left[f_{\overline{J}|_{x_J}} \right]$$

Using the formula for $S \subseteq \overline{J}$,

$$\widehat{f_{\overline{J}|_{x_J}}}(S) = \sum_{T \subseteq J} \widehat{f}(S \cup T) x_J^T$$

Therefore,

$$\begin{aligned}
f^{\subseteq J}(x) &= \mathbb{E}_{y\sim\{-1,1\}^{\overline{J}}}\left[f(x_J, y)\right] \\
&= \mathbb{E}_{y\sim\{-1,1\}^{\overline{J}}}\left[f_{\overline{J}|x_J}\right] \\
&= \mathbb{E}_{y\sim\{-1,1\}^{\overline{J}}}\left[\sum_{S\subseteq\overline{J}} \widehat{f_{\overline{J}|x_J}}(S) y^S\right] \\
&= \mathbb{E}_{y\sim\{-1,1\}^{\overline{J}}}\left[\sum_{S\subseteq\overline{J}}\sum_{T\subseteq J} \widehat{f}(S\cup T) x_J^T y^S\right] \\
&= \sum_{S\subseteq\overline{J}}\sum_{T\subseteq J} \mathbb{E}_{y\sim\{-1,1\}^{\overline{J}}}\left[\widehat{f}(S\cup T) x_J^T y^S\right] \\
&= \sum_{T\subseteq J} \widehat{f}(T) x_J^T
\end{aligned}$$

Problem 29. Let $\varphi : \mathbb{F}_2^n \to \mathbb{R}$ probability density function and $\phi : \mathbb{F}_2^n \to \mathbb{R}$ its corresponding probability distribution. Let $J \subseteq [n]$.

(a) Express the density function of the marginal distribution on J (in terms of φ).

(b) Consider the probability distribution ϕ conditioned on $z \in \{0,1\}^{\overline{J}}$. Assuming it is well defined, what is the probability density on that distribution (in terms of φ)?

Solution 29. The relation between the probability distribution and density function and the probability distribution is a factor of 2^n. The marginal distribution is given by $\varphi^{\subseteq J}$ from exercise 3. The probability density conditioned on a string is $\varphi_{J|_z}$.

Problem 30. Suppose $f : \{-1,1\}^n \to \mathbb{R}$ is computable by a descision tree that has a leaf with depth k labeled b. Show that $\hat{||}f\hat{||}_\infty \geq b/2^k$.

Solution 30. Let C be the subcube corresponding to the leaf with depth k labeled b. Consider $f^{\subseteq J}$ from exercise 3 with J corresponding to the coordinates of the subcube C and $z \in C$ (arbitrary). Three observations:

- $f^{\subseteq J}(z) = b$.
- $\deg(f) \leq k$.
- The Fourier expansion of $f^{\subseteq J}$,
$$f^{\subseteq J} = \sum_{S\subseteq J} \widehat{f}(S)$$

In particular we have,

$$f^{\subseteq J}(z) = \sum_{S \subseteq J} \widehat{f}(S) z^S = b$$

As $\deg(f) \leq k$ there are 2^k coefficients and the above expression hence one of those must be at least $b/2^k$.

Problem 31. Show that $\widehat{f^{+z}}(\gamma) = (-1)^{\langle \gamma, z \rangle} \widehat{f}(\gamma)$.

Solution 31. By a simple calculation,

$$\begin{aligned}
\widehat{f^{+z}}(\gamma) &= \mathbb{E}_x \left[f(x+z)(-1)^{\langle \gamma, x \rangle} \right] \\
&= \mathbb{E}_x \left[f(x)(-1)^{\langle \gamma, x+z \rangle} \right] \\
&= (-1)^{\langle \gamma, z \rangle} \widehat{f}(\gamma)
\end{aligned}$$

Another proof is that $f^{+z} = f * \varphi_{\{z\}}$ so,

$$\widehat{f^{+z}}(\gamma) = \widehat{f}(\gamma) \cdot \widehat{\varphi_{\{z\}}}(\gamma)$$

Problem 32.

(a) Suppose $f : \mathbb{F}_2^n \to \mathbb{R}$ has nontrivial sparsity $\text{sparsity}\left(\widehat{f}\right) < 2^{-n}$. Show that for any γ s.t. $\widehat{f}(\gamma) \neq 0$ there exists $\beta \in \widehat{\mathbb{F}_2^n}$ satisfying that $f_{\beta^\perp}$ has Fourier coefficient $\widehat{f}(\gamma)$ (the coefficient need not to be of the Fourier character χ_β).

(b) Prove by induction on n that if $f : \mathbb{F}_2^n \to \{-1, 1\}$ has Fourier sparsity $s = \text{sparsity}\left(\widehat{f}\right)$ then $\widehat{f}(\gamma)$ is $2^{1-\lfloor \log s \rfloor}$ granular.

(c) Prove that there are no functions with sparsity $\text{sparsity}\left(\widehat{f}\right) \in \{2, 3, 5, 6, 7, 9\}$.

Solution 32.

(a) Let $f : \mathbb{F}_2^n \to \mathbb{R}$ with sparsity $\text{sparsity}\left(\widehat{f}\right) < 2^{-n}$ and γ s.t. $\widehat{f}(\gamma) \neq 0$. Recall that $f_{\beta^\perp}$ has Fourier coefficients summing the coefficients under the quotient induced by β on $\mathbb{F}_2^n$ (two elements are equivalent if their difference is β). Using $\text{sparsity}\left(\widehat{f}\right) < 2^{-n}$ there exists $\alpha \in \widehat{\mathbb{F}_2^n}$ satisfying $\widehat{f}(\alpha) = 0$. Choose $\beta = \gamma + \alpha$ then under the quotient of β we have the coset $\{\gamma, \alpha\}$ and so restricted to $\beta^\perp$, the Fourier coefficient corresponding to the coset $\{\gamma, \alpha\}$ is $\widehat{f}(\gamma)$.

(b) To be precise, we will strengthen the hypothesis in the following sense: It holds for $f : H \to \{-1,1\}$ where H is a linear subspace and n is replaced by its dimension. However, we shall ignore this by simply stating that after invertible linear transformation (sparsity is invariant under linear transformations), we have a function on $\dim(H)$ variables.

We shall start by proving that the assertion is true if sparsity $\left(\widehat{f}\right) = 2^n$. In that case, the assertion holds since $\deg(f) \leq n$ trivially and we have shown that $\deg(f) \leq k$ implies that $\widehat{f}$ is 2^{1-k} granular. Moreover, the theorem holds trivially for $n = 1$. Now assume sparsity $\left(\widehat{f}\right) < 2^n$ and let $\gamma \in \widehat{\mathbb{F}_2^n}$ s.t. $\widehat{f}(\gamma) \neq 0$. By the previous section, we have a new function with $\widehat{f}(\gamma)$ is a Fourier coefficient. This function is with one variable less (specifically, the function is defined on a subspace with dimension $n-1$) and so we can use the induction hypothesis.

(c) We distinguish between the possible granularities:

- Case $\{2,3\}$: By the previous section, the granularity is 1, i.e all nonzero Fourier coefficients are integers but by Parseval there is exactly one such coefficient with magnitude equal to one.
- Case $\{5,6,7\}$: In this case the granularity is $1/2$, thus the possible values for the nonzero Fourier coefficients square are $1/4$, 1. By Parseval we either have 4 nonzero coefficients with magnitude $1/2$ or exactly one with magnitude 1.
- Case $\{9\}$: In this case the granularity is $1/4$, thus the possible values for the nonzero Fourier coefficients square are $1/16$,$1/4$, $9/16$, 1. Again, use Parseval. We may have only a single nonzero coefficient with magnitude 1 thus we may assume that no nonzero coefficient is with magnitude 1. Denote the number of Fourier nonzero coefficients square with a, b, c for the values $1/16$,$1/4$, $9/16$ respectively. We thus have the following equations,

$$a + b + c = 9$$
$$a + 4b + 9c = 16$$

  Subtract to obtain $3b + 8c = 7$ which does not have solution with positive integers.

Problem 33. Show that one can learn any target $f : \{-1,1\}^n \to \{-1,1\}$ with error 0 from random examples only in time $\tilde{O}(2^n)$.

Solution 33. That probability that a single point is not sampled in t random examples is $(1 - 2^{-n})^t$. For $t = 2^n$ this tends to the constant $1/e$ and for $t = n2^n$ the probability is roughly e^{-n}. By union bound the probability that a single point is not sampled in $n2^n$ random examples is bounded by $\left(\frac{2}{e}\right)^n$ which is exponentially small. After sampling all points we can simply output the truth table of the function.

Problem 34. Let $f : \{-1,1\}^n \to \{-1,1\}$, $g : \{-1,1\}^n \to \mathbb{R}$ such that $||f - g||_1 \leq \epsilon$. Pick $\theta \in [-1,1]$ uniformly at random and define $h : \{-1,1\}^n \to \{-1,1\}$ by,

$$h_\theta(x) = \operatorname{sgn}(g(x) - \theta)$$

Show that $\mathbb{E}_\theta [\operatorname{dist}(f, h_\theta)] \leq \epsilon/2$.

Solution 34. First note the pointwise inequality,

$$\mathbb{E}_\theta\left[1_{f(x)\neq \operatorname{sgn}(g(x)-\theta)}\right] \leq \frac{1}{2}\cdot|f(x)-g(x)|$$

Without the loss of generality assume $f(x)=1$ then the inequality is actually,

$$\mathbb{E}_\theta\left[1_{1\neq \operatorname{sgn}(g(x)-\theta)}\right] \leq \frac{1}{2}\cdot|1-g(x)|$$

For $|1-g(x)|\geq 2$ this is trivial. If $|1-g(x)|\leq 1$ then in order to have $1\neq \operatorname{sgn}(g(x)-\theta)$ we need $\theta<-g(x)$ and this happens with probability $\frac{1}{2}\cdot|1-g(x)|$. If $|1-g(x)|\geq 1$ (note this means $g(x)<0$) then we need $\theta>-g(x)$ which happens with probability $\frac{1}{2}\cdot|1-g(x)|$. We now prove the inequality,

$$\begin{aligned}
\mathbb{E}_\theta\left[\operatorname{dist}(f,h_\theta)\right] &= \mathbb{E}_\theta\left[\Pr_x\left[f(x)\neq \operatorname{sgn}(g(x)-\theta)\right]\right] \\
&= \mathbb{E}_\theta\left[\mathbb{E}_x\left[1_{f(x)\neq \operatorname{sgn}(g(x)-\theta)}\right]\right] \\
&= \mathbb{E}_x\left[\mathbb{E}_\theta\left[1_{f(x)\neq \operatorname{sgn}(g(x)-\theta)}\right]\right] \\
&\leq \frac{1}{2}\cdot\mathbb{E}_x\left[|f(x)-g(x)|\right] \\
&= \frac{1}{2}\cdot||f-g||_1
\end{aligned}$$

Problem 35. (a) For n even find a function $f:\{-1,1\}^n\to\{-1,1\}$ that is not 1/2-concentrated on any set with size $<2^{n-1}$.

(b) Let $f:\{-1,1\}^n\to\{-1,1\}$ uniformly random function. Show that f is not 1/4-concentrated on degree up to $\lfloor\frac{n}{2}\rfloor$ with probability at least 1/2.

Solution 35. (a) Take any boolean function which has all of its Fourier coefficients nonzero (e.g, inner product).

(b) There are roughly 2^{n-1} monomials with degree $\leq n/2$. We now that $\widehat{f}(S)$ has mean zero and variance 2^{-n} for random function. By linearity of expectation,

$$\mathbb{E}_f\left[\sum_{|S|\leq n/2}\widehat{f}(S)^2\right] = 2^{n-1}\cdot 2^{-n} = \frac{1}{2}$$

The result follows from Markov inequality.

Problem 36. Show that $\mathscr{C}=\{f:\{-1,1\}^n\to\{-1,1\}:\deg(f)\leq k\}$ is learnable with error 0 in time $n^k\cdot\operatorname{poly}(2^k,n)$.

Solution 36. We have seen that degree $\leq k$ have Fourier coefficients with granularity 2^{1-k}. Consider the following learning algorithm:

- Estimate Fourier coefficients with degree $\leq k$ with error $\epsilon < 2^{-k}$.

- Round the Fourier coefficient so that it is an integer multiple of 2^{1-k}.

- Output the polynomial.

There are $O(n^k)$ possible Fourier coefficients. For each, we take roughly $O(2^{2k}n)$ random samples so that it has error larger than 2^{-k} with probability at most e^{-n}. By union bound, with exponentially small probability, all Fourier coefficients are estimated with error less than 2^{-k} and so granularity promises are calculation to be correct for every Fourier coefficient.

Problem 37. Prove that the following classes can be learned with zero error and $\text{poly}(n, 2^k)$ time and queries (not random examples).

(a) $\mathscr{C} = \{f : \{-1,1\}^n \to \{-1,1\} : f \text{ is } k\text{-junta}\}$.

(b) $\mathscr{C} = \{f : \{-1,1\}^n \to \{-1,1\} : \mathrm{DT}(f) \leq k\}$.

(c) $\mathscr{C} = \{f : \{-1,1\}^n \to \{-1,1\} : \text{sparsity}\left(\widehat{f}\right) \leq 2^{O(k)}\}$.

Solution 37. In particular k-juntas are functions with degree $\leq k$ so we can use the algorithm for functions with degree $\leq k$. For (b),(c) use granularity and the same algorithm as for degree $\leq k$ functions. For (a), note that the influence of every variable is either zero or at least 2^{-k} thus by sampling we can identify those

Problem 38. Prove that $\{f : \{-1,1\}^n \to \{-1,1\} : \hat{||}f\hat{||}_1 \leq s\}$ is learnable from queries with error ϵ in time $\text{poly}(n, s, 1/\epsilon)$.

Solution 38. Follows from Kushilevitz-Mansour algorithm and that every $f : \{-1,1\}^n \to \mathbb{R}$ is concentrated on set of size $\hat{||}f\hat{||}_1^2/\epsilon$.

Problem 39. Prove that the concept class of boolean functions $\epsilon/4$-concentrated on a collections with size at most M is learnable from queries and error ϵ in time $\text{poly}(M, n, 1/\epsilon)$.

Solution 39. We shall use Goldreich-Levin algorithm.

- Identify heavy Fourier by applying the Goldreich-Levin algorithm with $\tau = \frac{\epsilon}{2M}$. Denote the output list by L.

- Use random queries to sample the Fourier coefficients in the list with sufficient precision (we shall determine the exact parameters later on).

We are given that the the function is $\epsilon/4$-concentrated on a set $\mathscr{F}$ with size at most M.

Lemma 3.2. $\sum_{U \in L} \widehat{f}(U)^2 \geq 1 - \epsilon/2$

Proof. Consider two cases:

1. The Fourier coefficients not in $\mathscr{F}$ contributes at most $\epsilon/4$.

2. Every Fourier coefficients in $\mathscr{F}$ not included in the list contributes at most $\tau/2 = \frac{\epsilon}{4M}$ hence all Fourier coefficients in $\mathscr{F}$ not included in the list contributes at most $\epsilon/4$.

3. Let $g : \{-1,1\}^n \to \mathbb{R}$ be the function which is the sum of the above estimated Fourier coefficients and output $\operatorname{sgn}(()\, g)$.

In total, all Fourier coefficients not in the list contribute at most $\epsilon/2$. $\square$

Next, note that the list L is with size at mos $4/\tau^2 = \mathrm{poly}(M, 1/\epsilon)$. Estimate each coefficient with error $\pm\epsilon$. Using random examples, this takes $\mathrm{poly}(M, 1/\epsilon)$ examples (even much less) to obtain error $(\tau^2/4) \cdot (\epsilon/2)$. By union bound, the total error (in the sampling) is no more than $\epsilon/2$. We analyze the total error in the (Fourier) L_2 norm (with high probability),

- At most $\epsilon/2$ when creating the list L (with high probability).
- At most $\epsilon/2$ when estimating the Fourier coefficients in L (with high probability).

We have $||f - g||_2 \leq \epsilon$ by using Parseval (the bound is on the Fourier norm) and so $\operatorname{dist}(f, \operatorname{sgn}(g)) \leq \epsilon$.

Problem 40. Suppose that A learns $\mathscr{C}$ in random examples, error $\epsilon/2$ and time T with probability at least $9/10$.

(a) Given $f \in \mathscr{C}$, show that A can check if the output hypothesis is indeed $\epsilon/2$-close to f or not in time $\mathrm{poly}(n, T, 1/\epsilon) \cdot \log(1/\delta)$.

(b) Show that there exists a learning algorithm for $\mathscr{C}$ in random examples, error $\epsilon/2$ and time $\mathrm{poly}(n, T, 1/\epsilon) \cdot \log(1/\delta)$ with error probability at least $1 - \delta$.

Solution 40. Suppose that A learns $\mathscr{C}$ in random examples, error $\epsilon/2$ and time T with probability at least $9/10$.

(a) Evaluating $h(x)$ on $O(\log(1/\delta)\log(1/\epsilon)^2)$ random points and comparing to $f(x)$ yields an approximation to $\operatorname{dist}(f, h)$ with error $\epsilon/2$ and with probability $1 - \delta$. To evaluate $h(x)$ note that the hypothesis is a circuit and since A runs at most T then the circuit size is at most T thus computing $h(x)$ takes roughly T time for each point.

(b) Amplify the above. Note that we might say "Yes" if $\operatorname{dist}(f, h) \leq \epsilon$.

   - Algorithm A first output hypothesis $h(x)$ and checks if it is "good".
   - If $h(x)$ is good output $h(x)$.
   - Otherwise continue (stop after n iterations).

   To make a mistake we need that A outputs a "bad" hypothesis, which happens with probability at most $1/10$, and A fails to identify this, which happens with probability at most δ. The probability of making an error is then bounded by,

$$\frac{\delta}{10} + \frac{\delta}{10^2} + \frac{\delta}{10^3} + \cdots \leq \frac{\delta}{9}$$

   Truncating after n iterations leaves a negligible error of 2^{-n}. This will take $\mathrm{poly}(n, T, 1/\epsilon) \cdot \log(1/\delta)$ time.

Problem 41. Our description of the Low-Degree Algorithm with degree k and error ϵ involved using a new batch of random examples to estimate each low-degree Fourier coefficient.

(a) Show that one can instead simply draw a single batch $\mathscr{E}$ of $\mathrm{poly}(n^k, 1/\epsilon)$ examples and use $\mathscr{E}$ to estimate each of the low-degree coefficients.

(b) Show that the output hypothesis is of the form,

$$h(y) = \mathrm{sgn}\left(\sum_{(x,f(x))\in\mathscr{E}} w(\Delta(x,y)) \cdot f(x) \right)$$

where $w : \{0,1,\ldots,n\} \to \mathbb{R}$ some function.

Solution 41. (a) Since we use union bound to bound the total error, there is no requirement for any sort of independence.

(b) Lets examine the contribution of a single sample to a fixed Fourier coefficient $\widehat{f}(S)$. The estimation is given by,

$$est(\widehat{f}(S)) = \sum_{(x,f(x))\in\mathscr{E}} f(x)x^S$$

Thus for input y this contributes to the total sum,

$$y^S \cdot est(\widehat{f}(S)) = \sum_{(x,f(x))\in\mathscr{E}} f(x)(y\cdot x)^S = \sum_{(x,f(x))\in\mathscr{E}} f(x)\cdot(-1)^{\Delta(y_S,x_S)}$$

This suggests that,

$$\begin{aligned}\sum_{|S|\le k} y^S \cdot est(\widehat{f}(S)) &= \sum_{|S|\le k}\sum_{(x,f(x))\in\mathscr{E}} f(x)\cdot(-1)^{\Delta(y_S,x_S)} \\ &= \sum_{(x,f(x))\in\mathscr{E}} f(x) \sum_{|S|\le k}(-1)^{\Delta(y_S,x_S)}\end{aligned}$$

Lemma 3.3. *Suppose that $\Delta(x,y) = D$. Then,*

$$\sum_{|S|\le k}(-1)^{\Delta(y_S,x_S)} = \sum_{j=0}^{D/2}\binom{D}{2j}\binom{n-D}{\le k-2j} - \binom{D}{2j+1}\binom{n-D}{\le k-2j-1}$$

Proof. The parity $(-1)^{\Delta(y_S,x_S)} = 1$ if x, y are different on exactly even number of points in S. Suppose that we are looking for all S satisfying $\Delta(y_S,x_S) = 2j$. We need to choose $2j$ points from $\{i\in[n] : x_i \neq y_i\}$ and another $\le k-2j$ points from $\{i\in[n] : x_i = y_i\}$. There are exactly,

$$\binom{D}{2j}\binom{n-D}{\le k-2j}$$

such sets. Enumerating over all possible j's and doing the same for odd integers gives the desired expression. ☐

Problem 42. Extend the Goldreich Levin Algorithm so that it works also for functions $f : \{-1, 1\}^n \to [-1, 1]$. The learning model for targets $f : \{-1, 1\}^n \to [-1, 1]$ assumes that $f(x)$ is always a rational number expressible by poly(n) bits.

Solution 42. The algorithm still work as is (with minor changes in the constants). The proof works just the same with the following changes:

- By Parseval, the Fourier weight is now ≤ 1 and so the number of active buckets is still bounded (and so the number of iterations is still bounded by $4n/\tau^2$).
- The estimations still holds for functions $f : \{-1, 1\}^n \to \mathbb{R}$ from two reasons. First, the formula for $\sum_{S \in \mathscr{B}} \widehat{f}(S)^2$ holds for real functions (not just boolean). Second (and most importantly), the Chernoff bound holds (with slightly worse constants) for random variables bounded in $[-1, 1]$ and so the analysis of the sampling still holds.

Problem 43. (a) Assume $\gamma, \gamma' \in \widehat{\mathbb{F}_2^n}$ are distinct. Show that $\Pr_x [\langle x, \gamma\rangle = \langle x, \gamma'\rangle] = 1/2$.

(b) Fix γ and suppose $x^{(1)}, \ldots, x^{(m)} \sim \mathbb{F}_2^n$ drawn uniformly at random and independently. Show that for $m = Cn$ where C is sufficiently large constant, the only $\beta \in \widehat{\mathbb{F}_2^n}$ satisfying $\langle x^{(i)}, \beta\rangle = \langle x^{(i)}, \gamma\rangle$ for all $i \in [m]$ is $\beta = \gamma$.

(c) Show that the concept class of all linear functions $\mathbb{F}_n 2 \to \mathbb{F}_2$ can be learned from random examples only, with error 0, in time $O(n^\omega)$ (where n^ω is the time required for multiplying two $n \times n$ matrices).

Solution 43. (a) Let $\gamma \neq \gamma' \in \widehat{\mathbb{F}_2^n}$ then $\gamma_i \neq \gamma'_i$ for some i. Note that $\langle x, \gamma\rangle = \langle x, \gamma'\rangle$ and $\langle x \oplus e_i, \gamma\rangle \neq \langle x \oplus e_i, \gamma'\rangle$ or vise-versa. Considering all points $x \in \mathbb{F}_2^n$ in pairs $x, x \oplus e_i$ and the result follows.

(b) Fix β. Recall that $\Pr_x [\langle x, \gamma\rangle = \langle x, \beta\rangle] = 1/2$ thus the probability that $\langle x^{(i)}, \beta\rangle = \langle x^{(i)}, \gamma\rangle$ for all $i \in [m]$ is 2^{-m}. Take $m = 2n$ and consider the union bound over all $\beta \neq \gamma$. Thus, with probability at least $1 - 2^{-n}$, no β satisfying $\langle x^{(i)}, \beta\rangle = \langle x^{(i)}, \gamma\rangle$ for all $i \in [m]$.

(c) Choose $x^{(1)}, \ldots, x^{(2n)} \sim \mathbb{F}_2^n$ drawn uniformly at random and independently. This defines $2n$ linear constraints on γ, and by the above, has a unique solution with probability at least $1 - 2^{-n}$ which can be found by solving the linear system (takes $O(n^\omega)$ time).

Problem 44. Let $\tau \geq 1/2 + \epsilon$ for some constant $\epsilon > 0$. Give an algorithm simpler than Goldreich and Levins that solves the following problem with high probability: Given query access to $f : \{-1, 1\}^n \to \{-1, 1\}$ in time poly$(n, 1/\epsilon^2)$ find the unique $U \subseteq [n]$ such that $\left|\widehat{f}(U)\right| \geq \tau$, assuming it exists.

Solution 44. Note that $\left|\widehat{f}(U)\right| \geq 1/2 + \epsilon$ means that,

$$\Pr_x [f(x) = \chi_S(x)] \geq 3/4 + \epsilon/2$$

Therefore for a fixed x,

$$\Pr_y\left[f(y)f(xy)=f(x)\right]\geq 1/2+\epsilon$$

Amplifying by taking the majority $\mathrm{Maj}\{f(y^{(i)})f(xy^{(i)}):1\leq i\leq T\}$ where $y^{(1)},\ldots,y^{(T)}\sim\{-1,1\}^n$ uniformly and independently. Taking $T=O(n\log^2 1/\epsilon)$ we have that for any fixed x,

$$\Pr_{y^{(1)},\ldots,y^{(T)}}\left[f(x)\neq\mathrm{Maj}\{f(y^{(i)})f(xy^{(i)}):1\leq i\leq T\}\right]\leq e^{-n}$$

Doing this for $e_1,\ldots,e_n$, then by union bound we computed correctly $f(e_1),\ldots,f(e_n)$ with high probability which gives explicitly the function f,

$$f(x)=\prod x_i f(e_i)$$

Problem 45. Informally: a one-way permutation is a bijective function $f:\mathbb{F}_2^n\to\mathbb{F}_2^n$ that is easy to compute on all inputs but hard to invert on more than a negligible fraction of inputs; a pseudorandom generator is a function $g:\mathbb{F}_2^k\to\mathbb{F}_2^m$ for $m>k$ whose output on a random input looks unpredictable to any efficient algorithm. Goldreich and Levin proposed the following construction of the latter from the former: for $k=2n, m=2n+1$, define

$$g(r,s)=(r,f(s),\langle r,s\rangle)$$

When gs input (r,s) is uniformly random, then so is the first $2n$ bits of its output (using the fact that f is a bijection). The key to the analysis is showing that the final bit, $\langle r,s\rangle$, is highly unpredictable to efficient algorithms even given the first $2n$ bits $(r,f(s))$. This is proved by contradiction.

(a) Show that if A is a determined algorithm such that,

$$\Pr_{r,s}\left[A(r,f(s))=\langle r,s\rangle\right]\geq\frac{1}{2}+\gamma$$

Then, there exists $B\subseteq\mathbb{F}_2^n$ satisfying $|B|/2^n\geq\frac{\gamma}{2}$ and for all $s\in B$,

$$\Pr_r\left[A(r,f(s))=\langle r,s\rangle\right]\geq\frac{1}{2}+\frac{\gamma}{2}$$

(b) Switching to ±1 notation in the output, deduce $A\widehat{(r,f(s)})(s)$ for all $s\in B$ ($A(r,f(s)$ is a function of r and $f(s)$ is fixed).

(c) Show that the adversary can efficiently compute s given $f(s)$ (with high probability) for any $s\in B$. If γ is non-negligible, this contradicts the assumption that f is one-way.

(d) Deduce the same conclusion even if A is a randomized algorithm.

Solution 45. Let $f:\{-1,1\}^n\to\{-1,1\}$ "one-way permutation".

(a) By averaging argument. Consider $B = \{s : \Pr_r[A(r, f(s)) = \langle r, s\rangle] \geq \frac{1}{2} + \frac{\gamma}{2}\}$. Assume towards contradiction that $|B|/2^n > \frac{\gamma}{2}$. Then,

$$\begin{aligned}
\Pr_{r,s}[A(r, f(s)) = \langle r, s\rangle] &\leq \Pr_s[s \in B] \Pr_{r,s}[A(r, f(s)) = \langle r, s\rangle : s \in B] + \\
&+ \Pr_s[s \notin B] \Pr_{r,s}[A(r, f(s)) = \langle r, s\rangle : s \notin B] \\
&< \frac{\gamma}{2} \cdot 1 + 1 \cdot \left(\frac{1}{2} + \frac{\gamma}{2}\right) \\
&\leq \frac{1}{2} + \gamma
\end{aligned}$$

contradiction to the assumption.

(b) Recall that,

$$\widehat{A(r, f(s))}(s) = 2\Pr_r[A(r, f(s)) = \langle r, s\rangle] - 1 \geq \gamma$$

(c) Using Goldreich-Levin we can find all heavy Fourier coefficient (with weight $\geq \gamma$) to obtain a list of heavy Fourier coefficients L. Note that if $s \in B$ then $s \in L$ and as L is not too large a random element gives a non-negligible probability for computing s (assuming γ is non-negligible).

(d) We can fix a randomness to obtain a determinstic algorithm as in (a) with the probability for computing $\langle r, s\rangle$ (i.e, there exists a randomness such that...).

Problem 1. ....

Solution 1.

Problem 1. ....

Solution 1.

Problem 1. ....

Solution 1.

Problem 1. ....

Solution 1.

Problem 1. ....

Solution 1.

Problem 1. ....

Solution 1.

Problem 1. ....

Solution 1.

Problem 1. ....

Solution 1.

Problem 1. ....

Solution 1.

Problem 1. ....

Solution 1.

Problem 1. ....

Solution 1.

Problem 1. ....

Solution 1.

Problem 1. ....

Solution 1.