Analysis of Boolean Functions Solution Manual

Krish Singal and Ashwin Padaki

May 17, 2023

Preface

These solutions are the culmination of our collective reading of Ryan O'Donnell's *Analysis of Boolean Functions* over the Spring 2023 semester. The hope is for this to be used as a reference for students learning the topic for the first time just like us.

Contents

1 Boolean Functions and the Fourier Expansion

Problem 1.2 Recall that the set of boolean functions $\{f | f : \{-1,1\}^n \to \{-1,1\}\}$ is a vector space of dimension 2^n with parity functions χ_S (for $S \subseteq [n]$) being the basis vectors. Let $\widehat{f}(S)$ be the only non-zero Fourier coefficient. Since the range of f is $\{-1,1\}$, $\widehat{f}(S)$ must be either -1 or 1. By symmetry then, there are $2^n \cdot 2 = 2^{n+1}$ functions with exactly one non-zero Fourier coefficient. □

Problem 1.3 Let $A = f^{-1}(\{1\}) \subset \{-1,1\}^n$. We are given that $|A|$ is odd and $n \geq 2$. Let $S \subseteq [n]$. Then, we have $\widehat{f}(S) = \mathbb{E}_x[f(x) \cdot \chi_S(x)] = 2 \cdot \Pr_x[f(x) = \chi_S(x)] - 1$. To show that $\widehat{f}(S) \neq 0$, then, it suffices to show that $\Pr_x[f(x) = \chi_S(x)] \neq \frac{1}{2}$. Let $B = \chi_S^{-1}(\{1\})$. Observe that $|B| = \begin{cases} 2^{n-1} & \text{if } S \neq \varnothing \\ 2^n & \text{if } S = \varnothing \end{cases}$. Since $n \geq 2$, $|B|$ is even in either case.

Then, we have:

$$\begin{aligned}
\Pr_x[f(x) = \chi_S(x)] &= \Pr_x[f(x) = \chi_S(x) = 1] + \Pr_x[f(x) = \chi_S(x) = -1] \\
&= \frac{|A \cap B|}{2^n} + \frac{|A^c \cap B^c|}{2^n} \\
&= \frac{|A \cap B|}{2^n} + \frac{2^n - |A \cup B|}{2^n} \\
&= \frac{|A \cap B|}{2^n} + \frac{2^n - |A| - |B| + |A \cap B|}{2^n} \\
&= \frac{2|A \cap B| + 2^n - (|A| + |B|)}{2^n}.
\end{aligned}$$

Note that the numerator is odd because $|A| + |B|$ is odd. In particular, the numerator cannot be 2^{n-1}, so $\Pr_x[f(x) = \chi_S(x)] \neq \frac{1}{2}$, which means $\widehat{f}(S) \neq 0$ as desired. □

Problem 1.4

$$\mathbb{E}_y[f(y)] = \mathbb{E}_y[\sum_{S \subseteq [n]} \widehat{f}(S) y^S] = \sum_{S \subseteq [n]} \widehat{f}(S) \cdot \mathbb{E}_y[y^S] \qquad \text{(Linearity of Expectation)}$$

$$= \sum_{S \subseteq [n]} \widehat{f}(S) \cdot \mathbb{E}_y[\Pi_{i \in S} y_i] = \sum_{S \subseteq [n]} \widehat{f}(S) \cdot \Pi_{i \in S} \mathbb{E}_y[y_i] \qquad \text{(Independence)}$$

$$= \sum_{S \subseteq [n]} \widehat{f}(S) \cdot \Pi_{i \in S} \mu_i = F(\mu)$$

□

Problem 1.5

(i) Suppose for the sake of contradiction that $|\widehat{f}(S)|, |\widehat{f}(T)| > \frac{1}{2}$ for $S \neq T$. Then $1 - 2 \cdot \mathrm{dist}(f, \chi_S) > \frac{1}{2}$ and $1 - 2 \cdot \mathrm{dist}(f, \chi_T) > \frac{1}{2}$ by definition, which yields $\mathrm{dist}(f, \chi_S), \mathrm{dist}(f, \chi_T) < \frac{1}{4}$. But then:

$$\begin{aligned}
\mathrm{dist}(\chi_S, \chi_T) &\leq \mathrm{dist}(f, \chi_S) + \mathrm{dist}(f, \chi_T) \\
&< \frac{1}{4} + \frac{1}{4} \\
&= \frac{1}{2}.
\end{aligned}$$

This is a contradiction, since $\mathrm{dist}(\chi_S, \chi_T) = \frac{1}{2}$ (as $\langle \chi_S, \chi_T \rangle = 0$). Hence, no such S, T can exist, as desired.

(ii) Consider $f : \{-1, 1\}^n \to \mathbb{R}$ given by $f(x) = \frac{3}{5} \Pi_{i=1}^n x_i + \frac{4}{5}$

□

Problem 1.7 Fix $S \subseteq [n]$. Then we have:

$$\begin{aligned}
\mathbb{E}_f[\widehat{f}(S)] &= 2^{-n} \sum_f \widehat{f}(S) \\
&= 2^{-n} \sum_f \langle f, \chi_S \rangle \\
&= 2^{-n} \langle \sum_f f, \chi_S \rangle && \text{(bilinearity)} \\
&= 2^{-n} \langle 0, \chi_S \rangle && \text{(can partition function space into } (f, -f) \text{ pairs)} \\
&= 0.
\end{aligned}$$

To clarify the above logic, let A be the set of functions $f\colon \{-1,1\}^n \to \{-1,1\}$ for which $f(1,1,\ldots,1) = 1$ and B be the set where $f(1,1,\ldots,1) = -1$. A and B partition the function space and have the property that $f \in A \iff -f \in B$. Therefore $\sum_f f = \sum_A f + \sum_B f = 0$.

We now compute the variance:

$$\begin{aligned}
\mathbf{Var}_f[\widehat{f}(S)] &= \mathbb{E}_f[\widehat{f}(S)^2] - \mathbb{E}_f[\widehat{f}(S)]^2 \\
&= \mathbb{E}_f[\widehat{f}(S)^2] \\
&= \mathbb{E}_f[\langle f, \chi_S \rangle^2] \\
&= \mathbb{E}_f[\langle f \cdot \chi_S, \chi_S \rangle^2] && \text{(permutation of function space)} \\
&= \mathbb{E}_f[\mathbb{E}_x[f(x) \cdot \chi_S(x)^2] \\
&= \mathbb{E}_f[\langle f, 1 \rangle^2] && (\chi_S^2 = 1) \\
&= \mathbb{E}_f[\widehat{f}(\varnothing)].
\end{aligned}$$

To clarify the above logic, observe that χ_S is fixed and $\{f\} = \{f \cdot \chi_S\}$, so replacing f with χ_S does not change the expectation. But note that there is no dependence on S. Therefore $\mathbb{E}_f[\widehat{f}(S)^2] = \mathbb{E}_f[\widehat{f}(\varnothing)^2]$ for all S. Finally, by Parseval's, we have:

$$\begin{aligned}
1 &= \mathbb{E}_f\left[\sum_T \widehat{f}(T)^2\right] \\
&= \sum_T \mathbb{E}_f\left[\widehat{f}(T)^2\right] \\
&= \sum_T \mathbb{E}_f[\widehat{f}(\varnothing)^2] \\
&= 2^n \cdot \mathbb{E}_f[\widehat{f}(\varnothing)^2].
\end{aligned}$$

We conclude that $\mathbb{E}_f[\widehat{f}(S)^2] = E_f[\widehat{f}(\varnothing)] = 2^{-n}$, as desired. □

Problem 1.8

(a)

$$\widehat{f^\dagger}(S) = \langle f^\dagger(x), \chi_S(x) \rangle = \langle -f(-x), \chi_S(x) \rangle = -\frac{1}{2^n} \sum_{x \in \{-1,1\}^n} f(-x)\chi_S(x)$$

Note that

$$\chi_S(x) = \begin{cases} -\chi_S(-x) \text{ if } |S| \bmod 2 = 1 \\ \chi_S(-x) \text{ if } |S| \bmod 2 = 0 \end{cases}$$

Then,

$$\widehat{f^\dagger}(S) = \begin{cases} \frac{1}{2^n} \sum_{x\in\{-1,1\}^n} f(-x)\chi_S(-x) \text{ if } |S| \bmod 2 = 1 \\ -\frac{1}{2^n} \sum_{x\in\{-1,1\}^n} f(-x)\chi_S(-x) \text{ if } |S| \bmod 2 = 0 \end{cases} = \begin{cases} \widehat{f}(S) \text{ if } |S| \bmod 2 = 1 \\ -\widehat{f}(S) \text{ if } |S| \bmod 2 = 0 \end{cases}$$

(b) $f^{\text{odd}}(x) + f^{\text{even}}(x) = \frac{f(x)-f(-x)}{2} + \frac{f(x)+f(-x)}{2} = f(x)$. If f is odd, then $f(-x) = -f(x)$ and $f^{\text{odd}}(x) = \frac{f(x)+f(x)}{2} = f(x)$. Similarly, if $f(x) = f^{\text{odd}}(x)$ then $f(x) = -f(-x)$ which means f is odd. A correspondingly similar argument can be used to show the case for even f.

(c) Note that

$$f = f^{\text{odd}} + f^{\text{even}}$$
$$f^\dagger = f^{\text{odd}} - f^{\text{even}}$$

Thus,

$$f + f^\dagger = 2f^{\text{odd}} = \sum_{S\subseteq[n]} \widehat{f}(S)\chi_S + \sum_{\substack{S\subseteq[n] \\ |S| \text{ odd}}} \widehat{f}(S)\chi_S - \sum_{\substack{S\subseteq[n] \\ |S| \text{ even}}} \widehat{f}(S)\chi_S = 2 \sum_{\substack{S\subseteq[n] \\ |S| \text{ odd}}} \widehat{f}(S)\chi_S$$

Which means $f^{\text{odd}} = \sum_{\substack{S\subseteq[n] \\ |S| \text{ odd}}} \widehat{f}(S)\chi_S$. Then,

$$f^{\text{even}} = f - f^{\text{odd}} = \sum_{S\subseteq[n]} \widehat{f}(S)\chi_S - \sum_{\substack{S\subseteq[n] \\ |S| \text{ odd}}} \widehat{f}(S)\chi_S = \sum_{\substack{S\subseteq[n] \\ |S| \text{ even}}} \widehat{f}(S)\chi_S$$

□

Problem 1.9

(a) We will use the notation $\{0, 1\}$ rather than $\{\text{F}, \text{T}\}$. Let $f\colon \{0, 1\}^n \to \{0, 1\}$ be arbitrary. For each $a \in \{0, 1\}^n$, let $\mathbf{1}_a(x) = \prod_{i=1}^n (1-a_i-x_i+2a_ix_i)$. By construction, $\mathbf{1}_a(x) = \begin{cases} 1 & \text{if } a = x \\ 0 & \text{if } a \neq x \end{cases}$. Moreover, $\mathbf{1}_a(x)$ is multilinear. Therefore, $f(x) = \sum_{a\in\{0,1\}^n} f(a) \cdot \mathbf{1}_a(x)$ is a multilinear representation for f.

(b) It is enough to show that the zero function has no nonzero multilinear representation $q(x) = \sum_{S\subseteq[n]} c_S x^S$ (if some function f had distinct representations p, p' then consider $q = p - p'$). Suppose for contradiction such a q exists. Let $T \subseteq [n]$ be minimal with $c_T \neq 0$, so that $S \subsetneq T$ does not hold for any S where $c_S \neq 0$. Let $a \in \{0, 1\}^n$ be given by $a_i = \begin{cases} 1 & \text{if } i \in T \\ 0 & \text{if } i \notin T \end{cases}$. Then:

$$q(a) = \sum_{S\subseteq[n]} c_S x^S = \sum_{S\subseteq T} c_S x^S = c_S \neq 0.$$

This is a contradiction, meaning no such q exists, which proves uniqueness of multilinear representation.

(c) Let $\{c_S\}$ be the coefficients for the multilinear representation of $f\colon \{0,1\}^n \to \{0,1\}$ .Let $T \subseteq [n]$ be arbitrary. We aim to show that $c_T \in [-2^n, 2^n] \cap \mathbb{Z}$. We make the following claim:

$$\sum_{S \subseteq T} (-1)^{|S|} \sum_{R \subseteq S} c_R = (-1)^{|T|} c_T$$

To show this, take any $R \subsetneq T$ and consider the coefficient of c_R in the sum:

$$\sum_{S\colon\ R \subseteq S \subseteq T} (-1)^{|S|} = (1-1)^{|T|-|R|} = 0.$$

Meanwhile, c_T only shows up once in the sum with coefficient $(-1)^{|T|} c_T$. This proves our claim. Now, observe that for a set $S \subseteq [n]$, we have $\sum_{R \subseteq S} c_R = f(a_S)$ where $(a_S)_i := \begin{cases} 1 & \text{if } i \in S \\ 0 & \text{if } i \notin S \end{cases}$. Therefore, we have:

$$|c_T| = \left| \sum_{S \subseteq T} (-1)^{|S|} \cdot f(a_S) \right|.$$

Finally, recall that $f(a_S) \in \{0,1\}$, which means that $c_T \in \mathbb{Z}$, and which also gives the following bounds:

$$|c_T| \leq \max \left\{ \sum_{\substack{S \subseteq T \\ |S| \text{ even}}} 1,\ \sum_{\substack{S \subseteq T \\ |S| \text{ odd}}} 1 \right\} = \begin{cases} 2^{|T|-1} & \text{if } T \neq \varnothing \\ 1 & \text{if } T = \varnothing \end{cases}.$$

To clarify, we could make c_T as positive as possible by setting $f(a_S) = 1$ for all even $|S|$, and $f(a_S) = 0$ for all odd $|S|$. Or we could make c_T as negative as possible with the opposite setting. Both cases give us the sum of every other binomial coefficient $\binom{|T|}{k}$, which is $2^{|T|-1} \leq 2^{n-1}$ when $T \neq \varnothing$. Therefore, we have actually shown a slightly stronger statement, that $c_T \in [-2^{n-1}, 2^{n-1}] \cap \mathbb{Z}$ whenever $n \geq 1$ (and $c_\varnothing \in \{0,1\}$ always).

(d) Take $f\colon \{0,1\}^n \to \{0,1\}$ with Fourier expansion p and $\{0,1\}$-Fourier expansion q. If $x \in \{0,1\}^n$ then we have $\frac{1}{2} - \frac{1}{2}p(1-2x_1, \ldots, 1-2x_n) = \frac{1}{2} - \frac{1}{2} \cdot (-1)^{f(x)} = f(x)$. By uniqueness of multilinear representation, it follows that $q(x) = \frac{1}{2} - \frac{1}{2}p(1-2x_1, \ldots, 1-2x_n)$ as desired.

$\square$

Problem 1.10

(a) Suppose $a, b \in \mathbb{R}$ with $b \neq 0$ and $a + bf \neq 0$. Write $f(x) = \sum_{S \subseteq [n]} \widehat{f}(S) x^S$ so that $(a+bf)(x) = a + \sum_{S \subseteq [n]} b\widehat{f}(S) x^S$. If $S \neq \varnothing$ then clearly $\widehat{(a+bf)}(S) = b\widehat{f}(S) = 0 \iff \widehat{f}(s) = 0$ since $b \neq 0$. Likewise, $\widehat{(a+bf)}(\varnothing) = a + b\widehat{f}(\varnothing) = 0$ only if $\deg f \neq 0$, by our assumption that $a + bf \neq 0$. And so either $\deg f = 0$ in which case $\deg(a+bf) = \deg f = 0$ as well, or $\deg f \neq 0$ in which case $\deg(a+bf) = \deg f$, which is what we wanted to show.

(b) ($\Rightarrow$) Suppose $\deg f \leq k$. By definition, we can write:

$$f(x) = \sum_{\substack{S \subseteq [n] \\ |S| \leq k}} \widehat{f}(S) x^S.$$

If we take $g_S := x^S$ then each g_S depends on at most k input coordinates, meaning f is of the desired form.

($\Leftarrow$) Suppose $f = \alpha_1 g_1 + \cdots + \alpha_m g_m$ where $\alpha_i \in \mathbb{R}$ and g_i depend on at most k input coordinates. For a given g_i, let $x_{i_1}, \ldots, x_{i_k}$ be the input coordinates g_i depends on. Then for $a \in \{-1, 1\}^k$ we can write the following indicators

$$\mathbf{1}_a(x_{i_1}, \ldots, x_{i_k}) = \prod_{j=1}^{k} \frac{1 + a_j x_{i_j}}{2} = \begin{cases} 1 & \text{if } a_j = x_{i_j} \forall j \\ 0 & \text{otherwise} \end{cases}$$

we can write $g_i(x) = \sum_{a \in \{-1,1\}^k} \mathbf{1}_a(x_{i_1}, \ldots, x_{i_k})$. But $\deg(g_i) \leq k$ so by applying part (a) finitely many times we conclude $\deg(f) = \deg(\alpha_1 g_1 + \cdots + \alpha_m g_m) \leq k$.

(c) Omitting this part because it is tedious and uninteresting.

$\square$

Problem 1.12

(a) We can show this by induction. For $n = 1$ we have $H_{2^1}[\gamma, x] = \begin{cases} -1 & \text{if } \gamma = x = 1 \\ 1 & \text{otherwise} \end{cases} = (-1)^{\gamma \cdot x}$. Now, supposing $H_{2^n}[\gamma, x] = (-1)^{\gamma \cdot x}$ for a specific $n \in \mathbb{N}$, we aim to show the same for $n+1$. If $\gamma, x \in \mathbb{F}_2^{n+1}$ observe that we can write $\gamma = \gamma' || a$ and $x = x' || b$ (concatenation) where $\gamma', x' \in \mathbb{F}_2^n$ and $a, b \in \mathbb{F}_2$. Since a, b are the most significant bits of γ and x, respectively, we can say $H_{2^{n+1}}[\gamma, x] = \begin{cases} -H_{2^n}[\gamma', x'] & \text{if } a = b = 1 \\ H_{2^n}[\gamma', x'] & \text{otherwise} \end{cases}$. But this is simply equal to $H[\gamma', x'] \cdot (-1)^{a \cdot b}$ which by the inductive hypothesis is $(-1)^{\gamma' \cdot x'} \cdot (-1)^{a \cdot b} = (-1)^{\gamma' \cdot x' + a \cdot b} = (-1)^{\gamma \cdot x}$, as desired.

(b) Let $f \colon \mathbb{F}_2 \to \mathbb{R}$. Let $i \in \mathbb{F}_2^n$ be associated with $S_i \subset \{0, \ldots, n-1\}$ according to the indexing scheme mentioned in the question. Then we have:

$$\begin{aligned} H_{2^n} f[i] &= H_{2^n}[i] \cdot f \\ &= \sum_{x=0}^{2^n - 1} (-1)^{i \cdot x} \cdot f(x). \end{aligned}$$

On the other hand, we have:

$$
\begin{aligned}
2^n \widehat{f}(S_i) &= 2^n \mathbb{E}_x[f(x) \cdot \chi_{S_i}(x)] \\
&= \sum_{x=0}^{2^n-1} f(x) \cdot (-1)^{\sum_{j \in S_i} x_j} \\
&= \sum_{x=0}^{2^n-1} f(x) \cdot \prod_{\ell=1}^{n} (-1) \cdot \mathbf{1}[x_\ell = -1] \cdot \mathbf{1}[\ell \in S_i] \\
&= \sum_{x=0}^{2^n-1} f(x) \cdot (-1)^{i \cdot x}.
\end{aligned}
$$

Hence $2^{-n} H_{2^n} f = \widehat{f}$.

(c) Let $f \in \mathbb{R}^{2^n}$. We can express $f = f_1 || f_2$ where $f_1, f_2 \in \mathbb{R}^{2^{n-1}}$. Then, consider the following algorithm $\mathrm{ALG}(n, f)$ for computing $H_{2^n} f$:

- If $n = 0$: return f.
- If $n \geq 1$: return $\begin{bmatrix} \mathrm{ALG}(n-1, f_1 + f_2) \\ \mathrm{ALG}(n-1, f_1 - f_2) \end{bmatrix}$.

This algorithm is correct because for $n \geq 1$ we have:

$$
\begin{aligned}
H_{2^n} f &= \begin{bmatrix} H_{2^{n-1}} & H_{2^{n-1}} \\ H_{2^{n-1}} & -H_{2^{n-1}} \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \end{bmatrix} \\
&= \begin{bmatrix} H_{2^{n-1}}(f_1 + f_2) \\ H_{2^{n-1}}(f_1 - f_2) \end{bmatrix}.
\end{aligned}
$$

We now evaluate its complexity. Its recursion tree has depth n, and the recursive step at depth k takes $2 \cdot 2^{n-k-1} = 2^{n-k}$ additions/subtractions (where $0 \leq k \leq n-1$), because one needs to compute $f_1 + f_2, f_1 - f_2 \in \mathbb{R}^{2^{n-k-1}}$. No computation needs to be done at the depth n level. Hence, the total complexity is given by:

$$
\begin{aligned}
\sum_{k=0}^{n-1} 2^{n-k} \cdot 2^k &= \sum_{k=0}^{n-1} 2^n \\
&= n 2^n.
\end{aligned}
$$

Hence, we have found an algorithm for computing $H_{2^n} f$ using only $n 2^n$ additions/subtractions, as desired.

(d) Since H_{2^n} is orthogonal and each row vector has magnitude $\sqrt{2^n}$, we have $H_{2^n}^2 = 2^n \cdot I_{2^n}$. Using part (b), we have $\widehat{\widehat{f}} = 2^{-n} H_{2^n} \widehat{f} = 2^{-2n} H_{2^n}^2 f = 2^{-n} f$, as desired.

$\square$

Problem 1.13 Consider the convex function $\varphi(t) = t^{\frac{q}{p}}$ where $q > p$. By Jensen's inequality

$$\begin{aligned}\mathbb{E}_x[\varphi(f(x)^p)] &\geq \varphi(\mathbb{E}_x[f(x)^p]) \\ \mathbb{E}_x[f(x)^q] &\geq \mathbb{E}_x[f(x)^p]^{q/p} \\ \mathbb{E}_x[f(x)^q]^{1/q} &\geq \mathbb{E}_x[f(x)^p]^{1/p} \\ \|f\|_q &\geq \|f\|_p\end{aligned}$$

To show the inequality holds for $q = \infty$ where $||f||_\infty := \max_{x\in\{-1,1\}^n}\{|f(x)|\}$, it suffices to show that for every f, $\lim_{q\to\infty} ||f||_q = ||f||_\infty$. Suppose $M = ||f||_\infty$ so that $|f(x)| \leq M$ for all $x \in \{-1, 1\}^n$. Then, we have:

$$\begin{aligned}\lim_{q\to\infty} ||f||_q &= \lim_{q\to\infty} \mathbb{E}_x\left[|f(x)|^q\right]^{1/q} \\ &\leq \lim_{q\to\infty} \mathbb{E}_x\left[M^q\right]^{1/q} \\ &= M.\end{aligned}$$

And:

$$\begin{aligned}\lim_{q\to\infty} ||f||_q &= \lim_{q\to\infty} \mathbb{E}_x\left[|f(x)|^q\right]^{1/q} \\ &= \lim_{q\to\infty} \frac{\left(\sum_x |f(x)|^q\right)^{1/q}}{2^{n/q}} \\ &\geq \lim_{q\to\infty} \frac{M}{2^{n/q}} \\ &= M & (\lim_{q\to\infty} 2^{n/q} = 1)\end{aligned}$$

We conclude that $\lim_{q\to\infty} ||f||_q = ||f||_\infty$ in general, so we are done. $\square$

Problem 1.15 We denote $k = |K|$. Then,

$$\mathbb{E}[g(x)] = \frac{1}{2^{n-k}} \sum_{x\in\{0,1\}^{n-k}} g(x) = \frac{1}{2^{n-k}} \sum_{x\in\{0,1\}^{n-k}} \sum_{S\subseteq[n]} \widehat{f}(S)\chi_S(xz)$$

where xz denotes the complete n bit string including the fixed z bits. We split this sum into two parts

$$= \frac{1}{2^{n-k}} \sum_{x\in\{0,1\}^{n-k}} \sum_{T\subseteq K} \widehat{f}(T)z^T + \frac{1}{2^{n-k}} \sum_{x\in\{0,1\}^{n-k}} \sum_{S\not\subseteq K} \widehat{f}(S)\chi_S(xz)$$

Notice that $\sum_{T\subseteq K} \widehat{f}(T)z^T$ is a constant for every x since z is fixed and $\widehat{f}(T)$ is determined by f. Thus,

$$= \sum_{T\subseteq K} \widehat{f}(T)z^T + \frac{1}{2^{n-k}} \sum_{S\not\subseteq K} \sum_{x\in\{0,1\}^{n-k}} \widehat{f}(S)\chi_S(xz)$$

where we have also switched the order of summation in the second term. For any $S \not\subseteq K$, $\chi_S(xz)$ is a function of $m = |S| - |S \cap K| \geq 1$ unfixed bits. Note then that the number of $x \in \{0,1\}^{n-k}$ such that $\chi_S(xz)$ has odd parity is exactly equal to those such that $\chi_S(xz)$ has even parity. This is because there are $2^{n-k+m} \sum_{i=0}^{\lfloor \frac{m-1}{2} \rfloor} \binom{m}{2i}$ many x that result in one parity and $2^{n-k+m} \sum_{i=0}^{\lfloor \frac{m-1}{2} \rfloor} \binom{m}{2i+1}$ many that result in the other. Since the sum of the even binomial coefficients is equal to that of the odd, the claim follows.

Then, for each $S \not\subseteq K$, $\sum_{x \in \{0,1\}^{n-k}} \widehat{f}(S)\chi_S(xz) = 0$. Therefore,

$$\mathbb{E}[g(x)] = \sum_{T \subseteq K} \widehat{f}(T) z^T$$

$\square$

Problem 1.16 By definition $\text{dist}(f,1) = \mathbb{P}[f(x) \neq 1] = \mathbb{P}[f(x) = -1]$ and $\text{dist}(f,-1) = \mathbb{P}[f(x) \neq -1] = \mathbb{P}[f(x) = 1]$. For $f : \{-1,1\}^n \to \{-1,1\}$

$$\begin{aligned}
\mathbf{Var}(f) &= 1 - \mathbb{E}[f^2] = 1 - (\mathbb{P}[f(x) = 1] - \mathbb{P}[f(x) = -1])^2 \\
&= 1 - \mathbb{P}[f(x) = 1]^2 - \mathbb{P}[f(x) = -1]^2 + 2\mathbb{P}[f(x) = 1]\mathbb{P}[f(x) = -1] \\
&= (1 + \mathbb{P}[f(x) = 1])\mathbb{P}[f(x) = -1] - \mathbb{P}[f(x) = -1]^2 + 2\mathbb{P}[f(x) = 1]\mathbb{P}[f(x) = -1] \\
&= \mathbb{P}[f(x) = -1](1 + \mathbb{P}[f(x) = 1] - \mathbb{P}[f(x) = -1]) + 2\mathbb{P}[f(x) = 1]\mathbb{P}[f(x) = -1] \\
&= 4\mathbb{P}[f(x) = 1]\mathbb{P}[f(x) = -1] = 4 \ \text{dist}(f,-1)\text{dist}(f,1)
\end{aligned}$$

Notice that $\text{dist}(f,1), \text{dist}(f,-1) \leq 1$. Then,

$$\begin{aligned}
\mathbf{Var}[f] &= 4 \cdot \min(\text{dist}(f,1), \text{dist}(f,-1)) \cdot \max(\text{dist}(f,1), \text{dist}(f,-1)) \\
&= 4\epsilon \cdot \max(\text{dist}(f,1), \text{dist}(f,-1)) \leq 4\epsilon
\end{aligned}$$

Secondly, because $\text{dist}(f,1) + \text{dist}(f,-1) = 1$, it follows that $\max(\text{dist}(f,1), \text{dist}(f,-1)) \geq \frac{1}{2}$. Thus,

$$\begin{aligned}
\mathbf{Var}[f] &= 4 \cdot \min(\text{dist}(f,1), \text{dist}(f,-1)) \cdot \max(\text{dist}(f,1), \text{dist}(f,-1)) \\
&\geq 2 \cdot \min(\text{dist}(f,1), \text{dist}(f,-1)) = 2\epsilon
\end{aligned}$$

Therefore $2\epsilon \leq \mathbf{Var}[f] \leq 4\epsilon$. $\square$

Problem 1.17 We denote $p = \mathbb{P}[f(x) = 1]$ and $q = \mathbb{P}[f(x) = -1]$. Then,

(1) $$\mathbf{Var}[F] = \mathbb{E}[(F-\mu)^2] = \mathbb{E}[F^2] - \mu^2 \qquad \text{(Definition)}$$

(2) $$\begin{aligned}
\mathbb{E}[(F - F')^2] &= \mathbb{E}[F^2] - 2\mathbb{E}[F'F] + \mathbb{E}[F'^2] && \text{(Linearity of Expectation)} \\
&= 2\mathbb{E}[F^2] - 2\mu^2 && \text{(Independence)} \\
&= 2\mathbb{E}[(F-\mu)^2]
\end{aligned}$$

(3)
$$\mathbb{E}[(F-F')^2] = 0\cdot(p^2+q^2) + 4\cdot(2pq) = 8pq$$
$$\mathbb{P}[F\neq F'] = 2pq$$

Therefore,

$$\frac{1}{2}\mathbb{E}[(F-F')^2] = 2\mathbb{P}[F\neq F']$$

(4)
$$\begin{aligned}
\mathbb{E}[|F-\mu|] &= |1-\mu|p + |\mu+1|q \\
&= (1-\mu)p + (\mu+1)q \quad (-1\leq\mu\leq 1 \implies |1-\mu|, |1+\mu|\geq 0) \\
&= (1-p+q)p + (p-q+1)q && (\mu = p-q) \\
&= p - p^2 + pq + pq - q^2 + q \\
&= 1 - (p^2+q^2) + 2pq && (p+q=1) \\
&= 4pq && ((p+q)^2 = 1)
\end{aligned}$$

Thus,

$$\mathbb{E}[|F-\mu|] = 2\mathbb{P}[F\neq F']$$

□

Problem 1.18 We perform a simple computation:

$$\begin{aligned}
\langle f^{=k}, f^{=\ell}\rangle &= \langle \sum_{S\colon |S|=k} \widehat{f}(S)\chi_S, \sum_{T\colon |T|=\ell} \widehat{f}(T)\chi_T\rangle \\
&= \sum_{S\colon |S|=k}\sum_{T\colon |T|=\ell} \widehat{f}(S)\widehat{f}(T)\langle\chi_S,\chi_T\rangle && \text{(bilinearity)} \\
&= \begin{cases} \sum_{S\colon |S|=k}\widehat{f}(S)^2 & \text{if } k=\ell \\ 0 & \text{if } k\neq\ell \end{cases} \\
&= \begin{cases} \mathbf{W}^k[f] & \text{if } k=\ell \\ 0 & \text{if } k\neq\ell \end{cases}.
\end{aligned}$$

This is what we wanted to show. □

Problem 1.19

(a) Suppose, for the sake of contradiction, that there exists an $f:\{-1,1\}^n\to\{-1,1\}$ such that $\mathbf{W}^1[f]=1$ but is not of the form $f=\pm\chi_S$. This implies that $\widehat{f}(S_{i_1}),...,\widehat{f}(S_{i_k})\neq 0$ for some $1<k\leq n$, where $1\leq i_j\leq n$ and $|S_{i_j}|=1$. Then, without loss of generality

$$f(\mathbf{1}) = \widehat{f}(S_{i_1}) + ... + \widehat{f}(S_{i_k}) = 1$$

Now denote $\mathbf{1_p} = (1,1,1,....,-1,...,1,1)$ where the pth index is set to -1. Then,

$$f(\mathbf{1}_{i_k}) = \widehat{f}(S_{i_1}) + ... - \widehat{f}(S_{i_k}) = \pm 1$$

If $f(\mathbf{1}_{i_k}) = 1$, then $\widehat{f}(S_{i_k}) = 0$ which contradicts our assumption that it is nonzero. If $f(\mathbf{1}_{i_k}) = -1$, then $\widehat{f}(S_{i_k}) = 1$ which by Parseval's means that $\widehat{f}(S_{i_1}), ..., \widehat{f}(S_{i_{k-1}}) = 0$ – contradiction.

(b) Suppose, for the sake of contradiction, that there exists an $f : \{-1,1\}^n \to \{-1,1\}$ such that $\mathbf{W}^{\leq 1}[f] = 1$ but depends on more than 1 input coordinate. If $\widehat{f}(\phi) = 0$ and $\widehat{f}(S_{i_1}), ..., \widehat{f}(S_{i_k}) \neq 0$ for some $1 < k \leq n$, where $1 \leq i_j \leq n$ and $|S_{i_j}| = 1$, then part (a) argues a contradiction. Otherwise, $\widehat{f}(\phi), \widehat{f}(S_{i_1}), ..., \widehat{f}(S_{i_k}) \neq 0$ for some $1 < k \leq n$, where $1 \leq i_j \leq n$ and $|S_{i_j}| = 1$. Then, without loss of generality

$$f(\mathbf{1}) = \widehat{f}(\phi) + \widehat{f}(S_{i_1}) + ... + \widehat{f}(S_{i_k}) = 1$$

Then,

$$f(\mathbf{1}_{i_k}) = \widehat{f}(\phi) + \widehat{f}(S_{i_1}) + ... - \widehat{f}(S_{i_k}) = \pm 1$$

If $f(\mathbf{1}_{i_k}) = 1$, then $\widehat{f}(S_{i_k}) = 0$ which contradicts our assumption that it is nonzero. If $f(\mathbf{1}_{i_k}) = -1$, then $\widehat{f}(S_{i_k}) = 1$ which by Parseval's means that $\widehat{f}(\phi), \widehat{f}(S_{i_1}), ..., \widehat{f}(S_{i_{k-1}}) = 0$ – contradiction.

(c) need to figure out

$\square$

Problem 1.21

(a) Suppose, for the sake of contradiction, that there exists an $f : \{-1,1\}^n \to \{-1,1\}$ that has exactly 2 non-zero Fourier coefficients. By Parseval's theorem,

$$\widehat{f}(S_1)^2 + \widehat{f}(S_2)^2 = 1$$

Consider input $x \in \{0,1\}^n$ given by indices $S_1 \cup S_2$ set to 1 and all else set to -1 (we denote this as $x_{S_1 \cup S_2}$). Then,

$$\begin{aligned} |f(x)| &= |\widehat{f}(S_1)\chi_{S_1}(x_{S_1 \cup S_2}) + \widehat{f}(S_2)\chi_{S_2}(x_{S_1 \cup S_2})| \\ &= |\widehat{f}(S_1) + \widehat{f}(S_2)| = 1 \end{aligned}$$

Similarly, there exist inputs such that $f(x) = \pm\widehat{f}(S_1) \pm \widehat{f}(S_2)$ all four combinations. Thus, any Fourier coefficients must in essence satisfy $\|f\|_1 = \|f\|_2$. The only solutions to this system are $(\pm 1, 0)$ and $(0, \pm 1)$. Thus, one of $\widehat{f}(S_1)$ or $\widehat{f}(S_2)$ is 0 and we have reached a contradiction.

(b) Suppose, for the sake of contradiction, that there exists an $f : \{-1,1\}^n \to \{-1,1\}$ that has exactly 3 non-zero Fourier coefficients. By Parseval's theorem,

$$\widehat{f}(S_1)^2 + \widehat{f}(S_2)^2 + \widehat{f}(S_3)^2 = 1$$

Consider input $x_{\cup_{i=1}^3 S_i}$

$$|f(x)| = |\sum_{i=1}^{3} \widehat{f}(S_i)\chi_{S_i}(x_{\cup_{i=1}^3 S_i})| = |\sum_{i=1}^{3} \widehat{f}(S_i)| = 1$$

Without loss of generality, assume $|S_1| \leq |S_2| \leq |S_3|$. We now claim that there exists an input x^* such that $\chi_{S_1}(x^*) = -\chi_{S_2}(x^*) = -\chi_{S_3}(x^*) = 1$. To see this, note that we can set $x_i^* = 1$ for all $i \in S_1$. Then, consider the following three cases

(1) $S_3 \backslash (S_1 \cup S_2) \neq \phi$ – $|S_2| \geq |S_1| \implies S_2 \backslash S_1 \neq \phi$. Thus, we can trivially make $\chi_{S_2}(x^*)$ odd parity by choosing $i \in S_2 \backslash S_1$ and setting $x_i^* = -1$. Similarly, since $S_3 \backslash (S_1 \cup S_2) \neq \phi$ we can appropriately set $j \in S_3 \backslash (S_1 \cup S_2)$ so as to give χ_{S_3} odd parity.

(2) $S_3 = (S_1 \cup S_2)$ – We set S_2 exactly as in case (1). Then, $\chi_{S_3}(x^*) = \chi_{S_1}(x^*) \cdot \chi_{S_2}(x^*) = -1$.

(3) $S_3 \subset (S_1 \cup S_2)$ – $|S_3| \geq |S_1| \implies S_3 \backslash S_1 \neq \phi \implies S_3 \cap S_2 \neq \phi$. Choose some $i \in S_3 \cap S_2$ and set $x_i^* = -1$ and set the remaining indices to 1. Then, $\chi_{S_2}(x^*) = \chi_{S_3}(x^*) = -1$.

Then,

$$|f(x^*)| = |\widehat{f}(S_1) - \widehat{f}(S_2) - \widehat{f}(S_3)| = 1$$

From above, however,

$$= |2\widehat{f}(S_1) - [\widehat{f}(S_1) + \widehat{f}(S_2) + \widehat{f}(S_3)]| = |2\widehat{f}(S_1) - 1| = 1$$

where we have assumed without loss of generality that $\widehat{f}(S_1) + \widehat{f}(S_2) + \widehat{f}(S_3) = 1$. $\widehat{f}(S_1) = 0$ and $\widehat{f}(S_1) = -1$ are the only two solutions. If $\widehat{f}(S_1) = 0$, f no longer has three non-zero coefficients and we have reached a contradiction. If $\widehat{f}(S_1) = -1$, then $\widehat{f}(S_2) = \widehat{f}(S_3) = 0$ by Parseval's theorem and we have reached another contradiction.

$\square$

Problem 1.24 We can compute $||\varphi||_2^2$ directly:

$$\begin{aligned}
||\varphi||_2^2 &= \mathbb{E}_x[|\varphi(x)|^2] \\
&= \mathbb{E}_x[\varphi(x)^2] && \text{(nonnegativity of density)} \\
&= \frac{1}{2^n} \sum_{x \in \{-1,1\}^n} \varphi(x)^2 \\
&= \frac{1}{2^n} \sum_{x \in A} \varphi(x)^2 \\
&\geq \frac{1}{2^n} \sum_{x \in A} \left(\frac{2^n}{2^n \delta}\right)^2 && \text{(convexity of } f(z) = z^2) \\
&= \frac{1}{2^n} \frac{\delta 2^n}{\delta^2} \\
&= \frac{1}{\delta}.
\end{aligned}$$

□

Problem 1.27 (Necessary) Suppose, for the sake of contradiction that one could determine which linear function f is using only k queries, for $k < n$. However, there are at least 2^{n-k} linear functions that have the same mapping at these k inputs. There is no way to distinguish between these candidates given this information, so we have reached a contradiction and can conclude that n queries are necessary.

(Sufficient) Consider querying f on $\mathbf{1_k}$ where $\mathbf{1_k} = (1, 1, 1, ..., 0, ..., 1, 1)$ where the kth index is set to 0. Notice that $k \in S \iff f(\mathbf{1_k}) = 0$ and that there is $2^{n-n} = 1$ unique f with the given answers to these queries. □

Problem 1.28

(a) We have $\widehat{f}(S) \leq 2\delta \iff \text{dist}(f, \chi_S) \geq \frac{1}{2} - \delta$ and $\widehat{f}(S) \geq -2\delta \iff \text{dist}(f, \chi_S) \leq \frac{1}{2} + \delta$ in general. Now, take $S \neq S^*$. By the triangle inequality (equivalently a union bound), we have:

$$\text{dist}(f, \chi_S) \leq \text{dist}(\chi_S, \chi_{S^*}) + \text{dist}(f, \chi_{S^*}) = \frac{1}{2} + \delta$$

And:

$$\text{dist}(\chi_S, f) \geq \text{dist}(\chi_S, \chi_{S^*}) - \text{dist}(f, \chi_{S^*}) = \frac{1}{2} - \delta.$$

We conclude that $-2\delta \leq \widehat{f}(S) \leq 2\delta$ so $|\widehat{f}(S)| \leq 2\delta$, which is what we wanted to show.

(b) As in the chapter, we have $\Pr[\text{BLR accepts } f] = \frac{1}{2} + \frac{1}{2}\sum_{S\subseteq[n]} \widehat{f}(S)^3$. But now we have tighter control on $\widehat{f}(S)$ for all $S \neq S^*$. In particular, we can write:

$$\begin{aligned}
\frac{1}{2} + \frac{1}{2}\sum_{S\subseteq[n]} \widehat{f}(S)^3 &= \frac{1}{2} + \frac{1}{2}\widehat{f}(S^*)^3 + \frac{1}{2}\sum_{S\neq S^*} \widehat{f}(S)^3 \\
&\leq \frac{1}{2} + \frac{1}{2}(1-2\delta)^3 + \frac{1}{2}\max_{S\neq S^*}\{\widehat{f}(S)\} \cdot \sum_{S\neq S^*} \widehat{f}(S)^2 \\
&\leq \frac{1}{2} + \frac{1}{2}(1-2\delta)^3 + \delta(1-(1-2\delta)^2) \quad (\text{as } \widehat{f}(S) \leq |\widehat{f}(S)| \leq 2\delta \text{ for } S \neq S^*) \\
&= 1 - 3\delta + 10\delta^2 - 8\delta^3.
\end{aligned}$$

Therefore, $\Pr[\text{BLR rejects } f] \geq 3\delta - 10\delta^2 + 8\delta^3$ which is what we wanted to show. On a practical note, this implies that if the BLR Test rejects some function f with probability at most ε, then $\min_S \text{dist}(f, \chi_S) = \delta \leq \varepsilon/3$ approximately (ignoring higher-order terms). In other words, f is essentially $\varepsilon/3$-close to some linear function (stronger than the ε-closeness shown in the chapter).

(c) need to figure out

□

Problem 1.29

(a) ($\Leftarrow$) Suppose $f(x) = a \cdot x$. Then,

$$f(x) + f(y) + f(z) = a \cdot x + a \cdot y + a \cdot z = a \cdot (x + y + z) = f(x + y + z)$$

($\Rightarrow$) Suppose $f(x) + f(y) + f(z) = f(x + y + z)$ and define $g(x) = f(x) + f(0)$. Then,

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

By the above, $g(x) = a \cdot x$. Thus, $f(x) = g(x) + f(0) = a \cdot x + f(0)$.

(b)

$$\begin{aligned}
\mathbb{E}_{x,y,z}[f(x)f(y)f(z)f(x+y+z)] &= \mathbb{E}_{x,y}[f(x)f(y)\mathbb{E}_z[f(z)f(x+y+z)]] \\
&= \mathbb{E}_{x,y}[f(x)f(y)(f * f)(x+y)] && \text{(Convolution definition)} \\
&= \mathbb{E}_x[f(x)\mathbb{E}_y[f(y)(f * f)(x+y)]] \\
&= \mathbb{E}_x[f(x)(f * f * f)(x)] \\
&= \sum_{S \subseteq [n]} \widehat{f}(S)\widehat{f * f * f}(S) && \text{(Plancherel)} \\
&= \sum_{S \subseteq [n]} \widehat{f}(S)\widehat{f * (f * f)}(S) && \text{(Associativity of Convolution)} \\
&= \sum_{S \subseteq [n]} \widehat{f}(S)\widehat{f}(S)\widehat{f * f}(S) && \text{(Theorem 1.27)} \\
&= \sum_{S \subseteq [n]} \widehat{f}(S)\widehat{f}(S)\widehat{f}(S)\widehat{f}(S) && \text{(Theorem 1.27)} \\
&= \sum_{S \subseteq [n]} \widehat{f}(S)^4 && \text{(Theorem 1.27)}
\end{aligned}$$

(c) **Affine Test.** *Given query access to* $\mathbb{F}_2^n \to \mathbb{F}_2$:

- *Choose* $x, y, z \sim \mathbb{F}_2^n$ *independently*
- *Query* f *at* x, y, z, *and* $x + y + z$
- *"Accept" if* $f(x + y + z) = f(x) + f(y) + f(z)$

Claim. *Suppose Affine Test accepts* $f : \mathbb{F}_2^n \to \mathbb{F}_2$ *with probability* $1 - \epsilon$. *Then* f *is* ϵ*-close to being linear.*

We use the indicator function

$$\frac{1}{2} + \frac{1}{2}f(x)f(y)f(z)f(x+y+z) = \begin{cases} 1 \text{ if } f(x)f(y)f(z) = f(x+y+z) \\ 0 \text{ if } f(x)f(y)f(z) \neq f(x+y+z) \end{cases}$$

where the output of f is encoded as ± 1. Then,

$$\begin{aligned} 1-\epsilon = \mathbb{P}_{x,y,z}[\textbf{Affine} \text{ accepts } f] &= \mathbb{E}_{x,y,z}[\frac{1}{2}+\frac{1}{2}f(x)f(y)f(z)f(x+y+z)] \\ &= \frac{1}{2}+\frac{1}{2}\sum_{S\subseteq[n]}\widehat{f}(S)^4 && \text{(part (b))} \end{aligned}$$

Rearranging, we get

$$\begin{aligned} 1-2\epsilon &= \sum_{S\subseteq[n]}\widehat{f}(S)^4 \\ &\le \left(\sum_{S\subseteq[n]}\widehat{f}(S)^2\right)\left(\sum_{S\subseteq[n]}\widehat{f}(S)^2\right) \\ &= \sum_{S\subseteq[n]}\widehat{f}(S)^2 && \text{(Parseval)} \\ &\le \max_{S\subseteq[n]}\{\widehat{f}(S)\}\cdot\sum_{S\subseteq[n]}\widehat{f}(S) = \max_{S\subseteq[n]}\{\widehat{f}(S)\}\cdot f(\mathbf{1}) \\ &\le \max_{S\subseteq[n]}\{\widehat{f}(S)\} \end{aligned}$$

Notice that affine functions are of the form $\chi_S + b$ where χ_S is a linear function and $b \in \mathbb{F}_2^n$. Thus,

$$\begin{aligned} \langle f, \chi_S + b\rangle &= \langle \sum_{T\subseteq[n]}\widehat{f}(T)\chi_T, \chi_S + b\rangle \\ &= \sum_{T\subseteq[n]}\widehat{f}(T)\langle \chi_T, \chi_S + b\rangle \\ &= \sum_{T\subseteq[n]}\widehat{f}(T)\langle \chi_T, \chi_S\rangle + \sum_{T\subseteq[n]}\widehat{f}(T)\langle \chi_T, b\rangle \\ &= \sum_{T\subseteq[n]}\widehat{f}(T)\langle \chi_T, \chi_S\rangle \\ &= \langle f, \chi_S\rangle = \widehat{f}(S) \end{aligned}$$

Therefore, $1 - 2\text{dist}(f, \chi_S + b) = 1 - 2\text{dist}(f, \chi_S)$ meaning that $\text{dist}(f, \chi_S + b) = \text{dist}(f, \chi_S)$. Denote $S^* = \text{argmax}_{S\subseteq[n]}\{\widehat{f}(S)\}$. Then,

$$1-2\epsilon \le 1 - \text{dist}(f, \chi_{S^*}) = 1 - 2\text{dist}(f, \chi_{S^*} + b)$$

which means that $\text{dist}(f, \chi_{S^*} + b) \le \epsilon$ and thus that f is ϵ-close to the space of affine functions.

(d) ($\Rightarrow$) Suppose f is affine. Then,

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

($\Leftarrow$) Suppose $f(x+y) = f(x) + f(y) + f(0)$. Then,

$$\begin{aligned} f(x+y+z) &= f(x) + f(y+z) + f(0) \\ &= f(x) + f(y) + f(z) + f(0) + f(0) \\ &= f(x) + f(y) + f(z) \end{aligned}$$

Then, the following 3-random affine test gives the same guarantees as the affine test and BLR test from 1.29(c) and the chapter.

3-Random Affine Test. *Given query access to* $\mathbb{F}_2^n \to \mathbb{F}_2$:

- *Choose* $x, y \sim \mathbb{F}_2^n$ *independently*
- *Query* f *at* x, y, *and* $x + y$
- *"Accept" if* $f(x+y) = f(x) + f(y) + f(0)$

□

2 Basic Concepts and Social Choice

Problem 2.2 Notice that the functions can be written as follows

(i) Majority: $\mathrm{sgn}(x_1 + ... + x_n)$

(ii) AND: $\mathrm{sgn}((n - \frac{1}{2}) + x_1 + ... + x_n)$

(iii) OR: $\mathrm{sgn}((\frac{1}{2} - n) + x_1 + ... + x_n)$

(iv) $\pm\chi_i$: $\mathrm{sgn}(\pm x_i)$

(v) ± 1: $\mathrm{sgn}(\pm 1)$

□

Problem 2.3

(a) ($\Leftarrow$) Suppose f is a weighted majority function so that we can write $f(x) = \mathrm{sgn}(a_0 + x_1 + \ldots + x_n)$. f is clearly symmetric and monotone.

($\Rightarrow$) Suppose $f \colon \{-1,1\}^n \to \{-1,1\}$ is symmetric and monotone. By symmetry of f, we can write $f(x) = g(t(x))$ where $t(x)$ denotes the number of ones in x. Since f is monotone, we can write:

$$f(-1, \ldots, -1) \leq f(1, -1, \ldots, -1) \leq \cdots \leq f(1, \ldots, 1)$$

And thus

$$g(0) \leq \ldots \leq g(n).$$

It follows that $g(k) = \text{sgn}(a + k)$ for some $a \in \mathbb{R}$. But for any x we have $x_1 + \ldots + x_n = t(x) - (n - t(x)) = 2t(x) - n$ and therefore $t(x) = \frac{1}{2}(n + x_1 + \ldots + x_n)$. Hence:

$$f(x) = g(t(x)) = \text{sgn}\left(a + \frac{1}{2}(n + x_1 + \ldots + x_n)\right) = \text{sgn}(2a + n + x_1 + \ldots + x_n)$$

Therefore f is a weighted majority function.

(b) Suppose $f\colon \{-1,1\}^n \to \{-1,1\}$ is symmetric, monotone and odd. By part (a), we can write $f(x) = \text{sgn}(a_0 + x_1 + \ldots + x_n)$ for some $a_0 \in \mathbb{R}$. Since f is also odd, we have $\text{sgn}(a_0 + x_1 + \ldots + x_n) = -\text{sgn}(a_0 - x_1 - \ldots - x_n)$, implying that $|x_1 + \ldots + x_n| > |a_0|$ for all $x \in \{-1,1\}^n$. Note that the minimum value of $|x_1 + \ldots + x_n|$ is 0 is n is even and 1 is n is odd. Since $0 > |a_0|$ cannot hold, we must have n odd, in which case $1 > |a_0|$ implies that we can set $a_0 = 0$ without loss of generality. Hence $f = \text{sgn}(x_1 + \ldots + x_n) = \text{Maj}_n$.

$\square$

Problem 2.4 ($\Rightarrow$) Given a string $z \in \{-1,1\}^n$ and real r

$$\mathbf{1}_A = \text{sgn}((n - 2r + 1) + \sum_i a_i x_i)$$

where $a_i = -z_i$. To see that this is true, notice that $\Delta(x,z) < r \iff \sum_i x_i z_i > (n-r+1)-(r-1) = n - 2r + 2$. Equivalently, $\Delta(x,z) < r \iff \sum_i x_i(-z_i) < -(n - r + 1) + (r - 1) = -(n - 2r + 2)$.

($\Leftarrow$) Given a linear threshold function of the form $f(x) = \text{sgn}(a_0 + \sum_i a_i x_i)$ where $|a_1| = ... = |a_n|$, we show that it is an indicator function for some z and r. By the above, $z_i = -\frac{a_i}{|a_i|}$ and $r = \left\lceil \frac{n-a_0}{2} \right\rceil$. $\square$

Problem 2.11 For each coordinate $i \in S$

$$\mathbf{Inf}_i(f) = \sum_{T \ni i} \widehat{f}(T)^2 \geq \widehat{f}(S)^2 > 0$$

since $\widehat{f(S)} > 0$ by assumption. Therefore, all $i \in S$ have non-zero influence and are thus relevant. $\square$

Problem 2.12 Both computations are self-explanatory:

$$\begin{aligned}
\mathbb{E}_f[\mathbf{Inf}_i[f]] &= \mathbb{E}_f\left[\Pr_x[f(x) \neq f(x^{\oplus i})]\right] \\
&= \mathbb{E}_f\left[\mathbb{E}_x[\mathbf{1}_{\{f(x) \neq f(x^{\oplus i})\}}]\right] \\
&= \mathbb{E}_x\left[\mathbb{E}_f[\mathbf{1}_{\{f(x) \neq f(x^{\oplus i})\}}]\right] \\
&= \mathbb{E}_x\,[1/2] \\
&= \frac{1}{2}.
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}_f[\mathbf{I}[f]] &= \mathbb{E}_f\left[\sum_{i=1}^{n} \mathbf{Inf}_i[f]\right] \\
&= \sum_{i=1}^{n} \mathbb{E}_f[\mathbf{Inf}_i[f]] \\
&= \frac{n}{2}.
\end{aligned}$$

$\square$

Problem 2.13

(a) We compute the expected value and variance of f as follows

$$\begin{aligned}
\mathbb{E}[f] &= \mathbb{P}_x[f(x) = 1] - \mathbb{P}_x[f(x) = -1] \\
&= \left((1 - \frac{1}{2^w})^{2^w}\right) - \left(1 - (1 - \frac{1}{2^w})^{2^w}\right) \\
\lim_{w \to \infty} \mathbb{E}[f] &= \frac{1}{e} - (1 - \frac{1}{e}) = \frac{2}{e} - 1
\end{aligned}$$

and

$$\begin{aligned}
\mathbf{Var}[f] &= \mathbb{E}[f^2] - \mathbb{E}[f]^2 \\
&= 1 - \mathbb{E}[f]^2 = 1 - \left(2\left((1 - \frac{1}{2^w})^{2^w}\right) - 1\right)^2 \\
&= 1 - \left(4\left((1 - \frac{1}{2^w})^{2^w}\right)^2 - 4\left((1 - \frac{1}{2^w})^{2^w}\right) + 1\right) \\
\lim_{w \to \infty} \mathbf{Var}[f] &= 1 - (\frac{4}{e^2} - \frac{4}{e} + 1) = \frac{4(e-1)}{e^2}
\end{aligned}$$

(b) Observe that the derivative

$$D_1 f = \frac{f^{(i \to 1)} - f^{(i \to -1)}}{2}$$

can never take value -1 for $f = \text{Tribes}_{w,2^w}$ since the OR and AND functions are monotone (and the composition of monotone functions is monotone). This implies the following casework

$$D_1 f = \begin{cases} 0 \text{ if } \exists j \in [2, ..., 2^w].\ AND(x^j) = -1 \\ 0 \text{ if } \forall j \neq 1.\ AND(x^j) = 1 \text{ and } \exists k \in [2, ..., w] x^{1,k} = 1 \\ 1 \text{ otherwise} \end{cases}$$

(c) From part (b), $D_1 f = 1$ if $\forall j \neq 1.\ AND(x^j) = 1$ and $x^{1,k} = -1$ for $k = 2, ..., w$. Therefore,

$$\begin{aligned}
\mathbf{Inf}_1[f] &= \frac{2 \cdot (2^w - 1)^{2^w - 1}}{2^{w2^w}} \\
&= \frac{2 \cdot (2^w - 1)^{2^w}}{(2^w - 1)2^{w2^w}} \\
&= \left(\frac{2^w - 1}{2^w}\right)^{2^w} \cdot \frac{2}{2^w - 1} \\
\lim_{w\to\infty} \mathbf{Inf}_1[f] &= \frac{2}{e} \cdot \lim_{w\to\infty} \frac{1}{2^w - 1} = 0
\end{aligned}$$

Since each voter is equal, the total influence is given by

$$\begin{aligned}
\mathbf{I}[f] = w2^w \mathbf{Inf}_i[f] &= \left(\frac{2^w - 1}{2^w}\right)^{2^w} \cdot \frac{2w2^w}{2^w - 1} \\
\lim_{w\to\infty} \mathbf{I}[f] = \infty &
\end{aligned}$$

Interestingly, the influence of any given voter converges to 0 while the total influence diverges in the limit.

$\square$

Problem 2.20 Recall that the Laplacian operator L has the property $\mathrm{L}f = f(x) \cdot \mathbf{sens}_f(x)$ for $f\colon \{-1,1\}^n \to \{-1,1\}$. Therefore, we have:

$$\begin{aligned}
\mathbb{E}_x[\mathbf{sens}_f(x)^2] &= \mathbb{E}_x[\mathbf{sens}_f(x)f(x) \cdot \mathbf{sens}_f(x)f(x)] && (f^2 = 1) \\
&= \langle Lf, Lf \rangle \\
&= \langle \sum_{S\subseteq[n]} |S|\widehat{f}(S)\chi_S, \sum_{T\subseteq[n]} |T|\widehat{f}(T)\chi_T \rangle \\
&= \sum_{S,T\subseteq[n]} |S||T|\widehat{f}(S)\widehat{f}(T)\langle \chi_S, \chi_T \rangle \\
&= \sum_{S\subseteq[n]} |S|^2 \widehat{f}(S)^2 \\
&= \mathbb{E}_{S\sim\mathcal{S}_f}[|S|^2].
\end{aligned}$$

We remark that $\mathbb{E}_x[\mathbf{sens}_f(x)^3] \neq \mathbb{E}_{S\sim\mathcal{S}_f}[|S|^3]$ in general. For instance, take $f = \mathrm{Maj}_3$ and observe that $\mathbb{E}_x[\mathbf{sens}_f(x)^3] = \frac{1}{4} \cdot 0 + \frac{3}{4} \cdot 2^3 = 6$ whereas $\mathbb{E}_{S\sim\mathcal{S}_f}[|S|^3] = \frac{1}{4} \cdot (1 + 1 + 1 + 27) = \frac{15}{2}$. $\square$

Problem 2.22

(a)

$$\begin{aligned}
\mathbf{Inf}_i[\mathrm{Maj}_n] &= \Pr_x[\mathrm{Maj}_n(x) \neq \mathrm{Maj}_n(x^{\oplus i})] \\
&= \Pr_x\left[\sum_{j \neq i} x_j = 0\right] && (x_i \text{ must break tie}) \\
&= \binom{n-1}{(n-1)/2} \cdot 2^{1-n}.
\end{aligned}$$

(b) Let $n \in \mathbb{N}$ be odd. Then:

$$\frac{\mathbf{Inf}_1[\mathrm{Maj}_n]}{\mathbf{Inf}_1[\mathrm{Maj}_{n+2}]} = 4 \cdot \frac{\binom{n-1}{(n-1)/2}}{\binom{n+1}{(n+1)/2}} < 4 \cdot \frac{\binom{n-1}{(n-1)/2}}{4 \cdot \binom{n-1}{(n-1)/2}} = 1.$$

Here, the last inequality follows from:

$$\binom{n+1}{(n+1)/2} = \binom{n}{(n+1)/2} + \binom{n}{(n-1)/2} = \binom{n-1}{(n+1)/2} + 2 \cdot \binom{n-1}{(n-1)/2} + \binom{n-1}{(n-3)/2} < 4 \cdot \binom{n-1}{(n-1)/2}$$

due to maximality of central binomial coefficients. It follows that $\mathbf{Inf}_1[\mathrm{Maj}_n]$ is (strictly) decreasing for odd n.

(c)

$$\begin{aligned}
\mathbf{Inf}_1[\mathrm{Maj}_n] &= \binom{n-1}{(n-1)/2} \cdot 2^{1-n} \\
&= \frac{(n-1)!}{(((n-1)/2)!)^2} \cdot 2^{1-n} \\
&\approx \frac{n!}{((n/2)!)^2} \cdot 2^{-n} && \text{(asymptotically the same)} \\
&= \frac{(n/e)^n \cdot (\sqrt{2\pi n} + O(n^{-1/2})}{(n/2e)^n \cdot (\sqrt{\pi n} + O((n/2)^{-1/2}))^2} \cdot 2^{-n} \\
&= \frac{\sqrt{2\pi n} + O(n^{-1/2})}{\pi n + O(1/n) + O(1)} \\
&= \frac{\sqrt{2\pi n} + O(n^{-1/2})}{\pi n} \\
&= \frac{\sqrt{2/\pi}}{\sqrt{n}} + O(n^{-3/2})
\end{aligned}$$

(d) Since Maj_n is monotone, we have $\mathbf{Inf}_i[\mathrm{Maj}_n] = \widehat{f}(i)$. Thus:

$$\begin{aligned}
\mathbf{W}^1[\mathrm{Maj}_n] &= \sum_{i=1}^{n} \widehat{f}(i)^2 \\
&= \sum_{i=1}^{n} \mathbf{Inf}_i[\mathrm{Maj}_n]^2 \\
&= n \cdot \left(\frac{\sqrt{2/\pi}}{\sqrt{n}} + O(n^{-3/2}) \right)^2 \\
&= n \cdot \left(\frac{2/\pi}{n} + O(n^{-3}) + \frac{4}{\pi} \cdot O(n^{-2}) \right) \\
&= \frac{2}{\pi} + O(n^{-2}) + \frac{4}{\pi} \cdot O(n^{-1}) \\
&\in \left[\frac{2}{\pi}, \frac{2}{\pi} + \frac{5}{\pi} \cdot n^{-1} \right] && \text{(for large enough } n\text{)}
\end{aligned}$$

Hence $2/\pi \le \mathbf{W}^1[\mathrm{Maj}_n] \le 2/\pi + O(n^{-1})$, as desired.

(e) Since Maj_n is symmetric, we have $\mathbf{W}^1[\mathrm{Maj}_n] = n \cdot \widehat{f}(1)^2$. By part (d), this implies

$$\widehat{f}(1)^2 \in \left[\frac{2}{\pi n}, \frac{2}{\pi n} + O(n^{-2}) \right].$$

Since $(\sqrt{2/\pi n} + O(n^{-3/2}))^2 = \frac{2}{\pi n} + O(n^{-2})) + O(n^{-3}) \ge \frac{2}{\pi n} + O(n^{-2}))$, we have:

$$\widehat{f}(1) \in \left[\sqrt{\frac{2}{\pi n}}, \sqrt{\frac{2}{\pi n}} + O(n^{-3/2}) \right].$$

Finally, we have:

$$\begin{aligned}
\mathbf{I}[\mathrm{Maj}_n] &= n \cdot \widehat{f}(1) \\
&\in \left[\sqrt{2/\pi} \cdot \sqrt{n}, \sqrt{2/\pi} \cdot \sqrt{n} + O(n^{-1/2}) \right].
\end{aligned}$$

□

Problem 2.23 Notice that if f is monotone, then by proposition 2.31

$$\mathbf{I}[f] = \sum_i \widehat{f(i)}$$

By the Cauchy-Schwarz inequality

$$\begin{aligned}
(\sum_i \widehat{f(i)} \cdot 1)^2 &\le (\sum_i \widehat{f(i)}^2)(\sum_i 1^2) \\
&\le n && \text{(Parseval's theorem)}
\end{aligned}$$

Therefore,

$$\sum_i \widehat{f(i)} \le \sqrt{n}$$

for monotone f. $\square$

Problem 2.27 Let x_1, x_2, x_3 be the three inputs that evaluate to 1 under f. If every dimension i-edge is a boundary edge for x_1, x_2, x_3, then the influence of coordinate i is maximized – namely $\mathbf{Inf}_i[f] = \frac{6}{2^n}$. For every coordinate to attain its maximum influence, it must be true that the pairwise hamming distances between x_1, x_2, x_3 are greater than 1 (if not, then there exists some coordinate i for which $x_1^{\oplus i} = x_2$ but $f(x_1) = f(x_2)$, meaning this dimension i-edge is not a boundary edge). Therefore,

$$\max_f \mathbf{I}[f] = \frac{6n}{2^n}$$

which is attained by f such that the pairwise hamming distances between x_1, x_2, x_3 are all greater than 1. $\square$

Problem 2.28 Suppose f is even. Then $\widehat{f}(S) = 0$ whenever $|S|$ is odd. Thus $\mathbf{W}^1[f] = 0$ so:

$$\begin{aligned}
\mathbf{I}[f] &= \sum_{S \subseteq [n]} |S| \cdot \widehat{f}(S)^2 \\
&= \sum_{k=1}^{n} k \cdot \mathbf{W}^k[f] \\
&= \sum_{k=2}^{n} k \cdot \mathbf{W}^k[f] \\
&\ge 2 \sum_{k>0}^{n} \mathbf{W}^k[f] \\
&= 2 \cdot \mathbf{Var}[f].
\end{aligned}$$

Equivalently $\mathbf{Var}[f] \le \frac{1}{2} \cdot \mathbf{I}[f]$, strengthening the Poincaré inequality. $\square$

Problem 2.29

(a) Since $\mathbb{E}[f] = 0$, $f(\phi) = 0$. Note that

$$\mathbf{Var}[f] = \mathbb{E}[f^2] - \mathbb{E}[f]^2 = \mathbb{E}[f^2] = 1$$

$$\mathbf{Var}[f] = 1 = \sum_{k=1}^{n} \mathbf{W}^k[f] \le \sum_{k=1}^{n} k\mathbf{W}^k[f] = \mathbf{I}[f] = \sum_{i=1}^{n} \mathbf{Inf}_i[f] \le n \cdot \mathbf{MaxInf}[f]$$

Therefore,

$$1 \le n\mathbf{MaxInf}[f]$$

$$\mathbf{MaxInf}[f] \ge \frac{1}{n}$$

(b) Observe that

$$\begin{aligned}
\mathbf{I}[f] = \sum_{k=0}^{n} k\mathbf{W}^k[f] &\geq \mathbf{W}^1[f] + 2(1 - \mathbf{W}^1[f]) \\
&= 2 - \mathbf{W}^1[f] \\
&= 2 - \sum_i \widehat{f(i)}^2 \\
&\geq 2 - \sum_i \mathbf{Inf}_i[f]^2 && \text{(Exercise 2.5)} \\
&\geq 2 - n\mathbf{MaxInf}[f]^2 && \text{(Monotonicity of the quadratic)}
\end{aligned}$$

Finally, we deduce that

$$\begin{aligned}
n\mathbf{MaxInf}[f] \geq \mathbf{I}[f] &\geq 2 - n\mathbf{MaxInf}[f]^2 \\
n\mathbf{MaxInf}[f]^2 + n\mathbf{MaxInf}[f] - 2 &\geq 0 \\
\mathbf{MaxInf}[f] &\geq \frac{-n + \sqrt{n^2 - 4(n)(-2)}}{2n} && \text{(Quadratic Formula)}
\end{aligned}$$

Thus,

$$\mathbf{MaxInf}[f] \geq \frac{n}{2} - \frac{1}{2} \geq \frac{2}{n} - \frac{4}{n^2}$$

$\square$

Problem 2.31

(i) Since $f \geq 0$, $\mathbb{E}_{y \sim N_\rho(x)}[f(y)] \geq 0$

(ii) Notice that $f \geq 0, f \neq 0 \implies \exists z.\ f(z) > 0$. Since $p \in (-1, 1)$, $\mathbb{P}_{y \sim N_\rho(x)}[y = z] > 0$. Therefore, $\mathbb{E}_{y \sim N_\rho(x)}[f(y)] \geq \mathbb{P}_{y \sim N_\rho(x)}[y = z] \cdot f(z) > 0$.

$\square$

Problem 2.32 Let $\rho_1, \rho_2 \in [-1,1]$ and $f\colon \{-1,1\} \to \mathbb{R}$. Observe that if $y \sim N_{\rho_1}(x)$ and $z \sim N_{\rho_2}(y)$, then for each $i \in [n]$ independently, we have:

$$\begin{aligned}
z_i &= \begin{cases} y_i & \text{with probability } \rho_2 \\ \text{unif. random} & \text{with probability } 1-\rho_2 \end{cases} \\
&= \begin{cases} \begin{cases} x_i & \text{with probability } \rho_1\rho_2 \\ \text{unif. random} & \text{with probability } (1-\rho_1)\rho_2 \end{cases} \\ \\ \begin{cases} \text{unif. random} & \text{with probability } \rho_1(1-\rho_2) \\ \text{unif. random} & \text{with probability } (1-\rho_1)(1-\rho_2) \end{cases} \end{cases} \\
&= \begin{cases} x_i & \text{with probability } \rho_1\rho_2 \\ \text{unif. random} & \text{with probability } 1-\rho_1\rho_2 \end{cases}
\end{aligned}$$

Hence $z \sim N_{\rho_1\rho_2}(x)$, which gives us:

$$\begin{aligned}
T_{\rho_1}T_{\rho_2}f(x) &= \mathbb{E}_{y\sim N_{\rho_1}(x)}[T_{\rho_2}f(y)] \\
&= \mathbb{E}_{y\sim N_{\rho_1}(x)}[\mathbb{E}_{z\sim N_{\rho_2(y)}}[f(z)]] \\
&= \mathbb{E}_{z\sim N_{\rho_1\rho_2}(x)}[f(z)] \\
&= T_{\rho_1\rho_2}f(x).
\end{aligned}$$

It follows that $T_{\rho_1}T_{\rho_2} = T_{\rho_1\rho_2}$ in general.

□

Problem 2.33 We show that the linear operator is a contraction on L^p by proving

$$\sum_x \mathbb{E}_{y\sim N_\rho(x)}[f(y)]^p \leq \sum_x f(x)^p$$

Observe that

$$\begin{aligned}
\sum_x \mathbb{E}_{y\sim N_\rho(x)}[f(y)]^p &= \sum_x \left(\sum_z f(z)\mathbb{P}_{y\sim N_\rho(x)}[y=z]\right)^p \\
&\leq \sum_x \sum_z f(z)^p \; \mathbb{P}_{y\sim N_\rho(x)}[y=z] && \text{(Jensen's inequality)} \\
&= \sum_z f(z)^p \sum_x \mathbb{P}_{y\sim N_\rho(x)}[y=z]
\end{aligned}$$

However,

$$\begin{aligned}
\sum_x \mathbb{P}_{y\sim N_\rho(x)}[y=z] &= \binom{n}{0}(\frac{1}{2}+\frac{1}{2}\rho)^n + \binom{n}{1}(\frac{1}{2}+\frac{1}{2}\rho)^{n-1}(\frac{1}{2}-\frac{1}{2}\rho)^1 + \ldots + \binom{n}{n}(\frac{1}{2}-\frac{1}{2}\rho)^n \\
&= \sum_{i=0}^{n} \binom{n}{i}(\frac{1}{2}+\frac{1}{2}\rho)^{n-i}(\frac{1}{2}-\frac{1}{2}\rho)^i \\
&= ((\frac{1}{2}+\frac{1}{2}\rho)+(\frac{1}{2}-\frac{1}{2}\rho))^n = 1^n = 1 && \text{(Binomial Theorem)}
\end{aligned}$$

Where we think of the quantity $\sum_x \mathbb{P}_{y\sim N_\rho(x)}[y=z]$ as the partitioned sum over x distance k from z. Then, for y ρ-correlated with x, there must be k bit flips and $n-k$ bit preservations. Then,

$$\begin{aligned}\sum_x \mathbb{E}_{y\sim N_\rho(x)}[f(y)]^p &\le \sum_z f(z)^p \sum_x \mathbb{P}_{y\sim N_\rho(x)}[y=z] \\ &= \sum_z f(z)^p\end{aligned}$$

and we can conclude that $\|T_\rho f\|_p \le \|f\|_p$. $\square$

Problem 2.34 Let $f\colon \{-1,1\} \to \mathbb{R}$ and $\rho \in [-1,1]$. For any $x \in \{-1,1\}$ we have:

$$\begin{aligned}|T_\rho f(x)| &= \left|\mathbb{E}_{y\sim N_\rho(x)}[f(y)]\right| \\ &\le \mathbb{E}_{y\sim N_\rho(x)}[|f(y)|] && \text{(convexity of } |\cdot|) \\ &= T_\rho|f|(x).\end{aligned}$$

Suppose $\rho \in (-1,1)$ so that for any $z \in \{-1,1\}$ we have $\Pr_{y\sim N_\rho(x)}[y=z] > 0$. We observe that equality above holds if and only if $\mathrm{sgn}(f(y)) = \mathrm{sgn}(f(z))$ for all $y,z \in \{-1,1\}^n$ where $f(y), f(z) \neq 0$, meaning $f \ge 0$ everywhere or $f \le 0$ everywhere. Another way of seeing this is that equality in Jensen's holds only if $\varphi(t) = |t|$ is linear on the support of f, meaning the support of f is either contained in $\mathbb{R}^{\ge 0}$ or contained in $\mathbb{R}^{\le 0}$. $\square$

Problem 2.41 Observe that

$$\frac{d^2}{d\rho^2}\mathbf{Stab}_\rho[f] = \sum_{k=0}^{n} k(k-1)\rho^{k-2}\mathbf{W}^k[f] \ge 0$$

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

Problem 2.42 Let $f\colon \{-1,1\}^n \to \{-1,1\}$ and let $\delta \in [0,1]$. Let $x \sim \{-1,1\}^n$ uniformly, and for each $i \in [n]$ independently let $y_i = \begin{cases} -x_i & \text{w.p. } \delta \\ x_i & \text{w.p. } 1-\delta \end{cases}$. Then $\mathbf{NS}_\delta[f] := \Pr_{x,y}[f(x) \neq f(y)]$. Now, define $y^{(0)} = (x_1, \ldots, x_n) = x$ and for each $i \in [n]$ define $y^{(i)} = (y_1, \ldots, y_i, x_{i+1}, \ldots, x_n)$ so that $y^{(n)} = y$. We make the simple observation that $f(x) \neq f(y)$ implies that $f(y^{(i)}) \neq f(y^{(i-1)})$ for some $i \in [n]$. Moreover, since x is uniformly random, the $y^{(i)}$ are individually uniformly random as well. By construction, we have $y^{(i)} = \begin{cases} (y^{(i-1)})^{\oplus i} & \text{w.p. } \delta \\ y^{(i-1)} & \text{w.p. } 1-\delta \end{cases}$. Therefore, we have:

$$\begin{aligned}
\mathbf{NS}_\delta[f] &= \Pr[f(x) \neq f(y)] \\
&= \Pr\left[\bigcup_{i=1}^{n} \{f(y^{(i)}) \neq f(y^{(i-1)})\}\right] \\
&\leq \sum_{i=1}^{n} \Pr\Big[f(y^{(i-1)}) \neq f(y^{(i)})\Big] \\
&= \sum_{i=1}^{n} \Pr\Big[f(y^{(i-1)}) \neq f(y^{(i)}) \mid y^{(i-1)} \neq y^{(i)}\Big] \cdot \Pr\Big[y^{(i-1)} \neq y^{(i)}\Big] \\
&= \sum_{i=1}^{n} \Pr\Big[f(y^{(i-1)}) \neq f(y^{(i)}) \mid y^{(i-1)} \neq y^{(i)}\Big] \cdot \delta \\
&= \delta \cdot \sum_{i=1}^{n} \Pr_{z \sim \{-1,1\}^n}[f(z) \neq f(z^{\oplus i})] \\
&= \delta \cdot \sum_{i=1}^{n} \mathbf{Inf}_i[f] \\
&= \delta \cdot \mathbf{I}[f].
\end{aligned}$$

$\square$

Problem 2.45 Let $0 < \delta \leq 1$ and $k \in \mathbb{N}^+$. Observe the following inequalities:

$$\sum_{j=0}^{k-1} (1-\delta)^j \geq \sum_{j=0}^{k-1} (1-\delta)^{k-1} = (1-\delta)^{k-1} k$$

And

$$\sum_{j=0}^{k-1} (1-\delta)^j \leq \sum_{j=0}^{\infty} (1-\delta)^j = \frac{1}{1-(1-\delta)} = \frac{1}{\delta}.$$

Hence $(1-\delta)^{k-1} k \leq \frac{1}{\delta}$. $\square$

Problem 2.46 Let $f\colon \{-1,1\}^n \to \mathbb{R}$ and $0 \leq \rho - \varepsilon \leq \rho < 1$ For $t \in (0,1)$ let $g(t) = \dfrac{d}{dt}\mathbf{Stab}_t[f] = \sum_{k=1}^{n} k t^{k-1} \mathbf{W}^k[f]$. If $\varepsilon = 0$ then the inequality is trivial. Otherwise, by the mean value theorem, we can find $c \in (\rho - \varepsilon, \rho)$ for which $|\mathbf{Stab}_\rho[f] - \mathbf{Stab}_{\rho-\varepsilon}[f]| = \varepsilon |g(c)|$. Thus, it suffices to show that

$|g(c)| \leq \frac{1}{1-\rho} \cdot \mathbf{Var}[f]$, which we can do as follows:

$$\begin{aligned}
|g(c)| &= \left| \frac{d}{dt}\Big|_{t=c} \mathbf{Stab}_t[f] \right| \\
&= \sum_{k=1}^{n} kc^{k-1} \mathbf{W}^k[f] \\
&= \left| \frac{d}{dt}\Big|_{t=c} \mathbf{Stab}_t[f] \right| \\
&= \sum_{k=1}^{n} \frac{1}{1-c} \cdot \mathbf{W}^k[f] \\
&= \left| \frac{d}{dt}\Big|_{t=c} \mathbf{Stab}_t[f] \right| \\
&= \frac{1}{1-c} \cdot \mathbf{Var}[f] \\
&\leq \frac{1}{1-\rho} \cdot \mathbf{Var}[f]. && \text{(as } c \geq \rho\text{)}
\end{aligned}$$

$\square$

Problem 2.54 We start with the base case $n = 1$. We have:

$$\mathbf{Var}[f] = \left(\frac{f(1) - f(-1)}{2} \right)^2 = \mathbb{E}_x[\mathrm{D}_i f(x)^2] = \mathbf{I}[f].$$

Now, suppose assume the inequality holds for n, and let $f \colon \{-1,1\}^{n+1} \to \{-1,1\}$. Consider $g, h \colon \{-1,1\}^n \to \mathbb{R}$ given by $g(x_1, \ldots, x_n) = f(x_1, \ldots, x_n, -1)$ and $h(x_1, \ldots, x_n) = f(x_1, \ldots, x_n, 1)$. Then, we have:

$$\mathbb{E}[f^2] = \frac{1}{2} \cdot \mathbb{E}[g^2 + h^2] = \frac{1}{2} \cdot \mathbb{E}[g^2] + \frac{1}{2} \cdot \mathbb{E}[h^2]$$

And:

$$\mathbb{E}[f]^2 = \left(\frac{1}{2} \cdot \mathbb{E}[g] + \frac{1}{2} \cdot \mathbb{E}[h] \right)^2 = \frac{1}{4} \cdot \mathbb{E}[g]^2 + \frac{1}{4} \cdot \mathbb{E}[h]^2 + \frac{1}{2} \cdot \mathbb{E}[g] \cdot \mathbb{E}[h].$$

Moreover, we notice that for $1 \leq i \leq n$ we have $\mathbf{Inf}_i[f] = \frac{1}{2} \cdot \mathbf{Inf}_i[g] + \frac{1}{2} \cdot \mathbf{Inf}_i[h]$, whereas $\mathbf{Inf}_{n+1}[f] = \mathbb{E}[z^2]$ where $z := \mathrm{D}_{n+1} f = \frac{g-h}{2}$. Therefore, we can write:

$$\mathbf{I}[f] = \frac{1}{2} \cdot \mathbf{I}[g] + \frac{1}{2} \cdot \mathbf{I}[h] + \mathbb{E}[z^2]$$

We now prove the desired inequality:

$$
\begin{aligned}
\mathbf{Var}[f] &= \mathbb{E}[f^2] - \mathbb{E}[f]^2 \\
&= \left(\frac{1}{2}\cdot\mathbb{E}[g^2] + \frac{1}{2}\cdot\mathbb{E}[h^2]\right) - \left(\frac{1}{4}\cdot\mathbb{E}[g]^2 + \frac{1}{4}\cdot\mathbb{E}[h]^2 + \frac{1}{2}\cdot\mathbb{E}[g]\cdot\mathbb{E}[h]\right) \\
&= \frac{1}{2}\cdot\left(\mathbb{E}[g^2] - \mathbb{E}[g]^2\right) + \frac{1}{2}\cdot\left(\mathbb{E}[h^2] - \mathbb{E}[h]^2\right) + \frac{1}{4}\cdot\left(\mathbb{E}[g]^2 + \mathbb{E}[h]^2 - 2\cdot\mathbb{E}[g]\cdot\mathbb{E}[h]\right) \\
&= \frac{1}{2}\cdot\mathbf{Var}[g] + \frac{1}{2}\cdot\mathbf{Var}[h] + \mathbb{E}[z]^2 && (z = \tfrac{g-h}{2}) \\
&\le \frac{1}{2}\cdot\mathbf{I}[g] + \frac{1}{2}\cdot\mathbf{I}[h] + \mathbb{E}[z]^2 && \text{(induction hypothesis)} \\
&= \mathbf{I}[f] - \mathbb{E}[z^2] + \mathbb{E}[z]^2 \\
&\le \mathbf{I}[f] && \text{(Jensen's inequality)}
\end{aligned}
$$

By induction, we have proved the Poincaré inequality for all $n \in \mathbb{N}$. □

Problem 2.55

(a) By definition of the Laplacian operator

$$
\begin{aligned}
Lg(x) = \sum_{i=1}^{n} L_i g(x) &= \sum_{i=1}^{n} \frac{g(x) - g(x^{\oplus i})}{2} \\
&= \frac{n}{2}g(x) - \frac{1}{2}\sum_{i=1}^{n} \| - x_i w_i + \sum_{j\neq i}^{n} x_j w_j \| \\
&\le \frac{n}{2}g(x) - \frac{1}{2}\| \sum_{i=1}^{n} \left[-x_i w_i + \sum_{j\neq i}^{n} x_j w_j \right] \| && \text{(Triangle Inequality)} \\
&= \frac{n}{2}g(x) - \frac{1}{2}\|(n-2)\sum_{i=1}^{n} x_i w_i\| \\
&= \frac{n-(n-2)}{2}g(x) = g(x) && \text{(Property of } \|\cdot\| \text{ norm)}
\end{aligned}
$$

since $x \in \{-1,1\}^n$ was chosen arbitrarily, we can conclude that the Laplacian of g is pointwise smaller than g.

(b) We first note that g is an even function. This is because

$$
g(-x) = \| \sum_{i=1}^{n} -x_i w_i \| = \| \sum_{i=1}^{n} x_i w_i \| = g(x) \qquad \text{(Property of } \|\cdot\| \text{ norm)}
$$

Then, by exercise 2.28 we have

$$\begin{aligned}
2\mathbf{Var}[g] \leq \mathbf{I}[g] &= \sum_{i=1}^{n} \mathbf{Inf}_i[g] \\
&= \sum_{i=1}^{n} \langle g, L_i g \rangle && \text{(Proposition 2.26)} \\
&= \langle g, \sum_{i=1}^{n} L_i g \rangle = \langle g, Lg \rangle && \text{(Bilinearity of } \langle \cdot, \cdot \rangle\text{)} \\
&\leq \langle g, g \rangle = \mathbb{E}[g^2] && \text{(Part (a))}
\end{aligned}$$

We can now deduce the *Khintchine-Kahane* inequality

$$\begin{aligned}
2\mathbf{Var}[g] = 2\mathbb{E}[g^2] - 2\mathbb{E}[g]^2 &\leq \mathbb{E}[g^2] \\
\mathbb{E}[g^2] &\leq 2\mathbb{E}[g]^2 \\
\mathbb{E}[g] &\geq \frac{1}{\sqrt{2}} \mathbb{E}[g^2]^{\frac{1}{2}}
\end{aligned}$$

Expanding, we get

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

(c) To see that the bound is tight, we construct an example that achieves equality. Consider $V = \mathbb{R}$, $n = 2$, and $w_1 = w_2 = 1$. Here, we get that

$$\begin{aligned}
\mathbb{E}_x[|x_1 + x_2|] &= \frac{1}{2} \cdot 2 + \frac{1}{2} \cdot 0 = 1 \\
\mathbb{E}_x[|x_1 + x_2|^2]^{\frac{1}{2}} &= \sqrt{\frac{1}{2} \cdot 4 + \frac{1}{2} \cdot 0} = \sqrt{2} \\
\implies \mathbb{E}_x[|x_1 + x_2|] &= \frac{1}{\sqrt{2}} \mathbb{E}_x[|x_1 + x_2|^2]^{\frac{1}{2}}
\end{aligned}$$

$\square$

3 Spectral Structure and Learning

Problem 3.2

$\square$

Problem 3.3

$$\begin{aligned}
\mathbb{E}[f^2] - \mathbf{Stab}_{1-\delta}[f] &= \sum_S \left[1 - (1-\delta)^{|S|}\right] \widehat{f}(S)^2 \\
&= \sum_{k=1} n \left[1 - (1-\delta)^{k}\right] \mathbf{W}^k[f] \\
&\geq \left(1 - (1-\delta)^{\frac{1}{\delta}}\right) \mathbf{W}^k[f] \\
&\geq (1 - \frac{1}{e}) \Pr\left[|S| \geq \frac{1}{\delta}\right]
\end{aligned}$$

Therefore,

$$\Pr\left[|S| \geq \frac{1}{\delta}\right] \leq \frac{\mathbb{E}[f^2] - \mathbf{Stab}_{1-\delta}[f]}{1 - \frac{1}{e}}$$

$\square$

Problem 3.4 We prove the claim inductively. For, $n = k$ it is trivially true that $\mathbb{P}_x[f(x) \neq 0] \geq 2^{-n} = 2^{-k}$ since f is not identically 0. We assume that for $f : \{-1,1\}^n \to \mathbb{R}$ where $\deg(f) \leq k$ we have $\mathbb{P}_x[f(x) \neq 0] \geq 2^{-k}$. Then, consider $f : \{-1,1\}^{n+1} \to \mathbb{R}$ such that $\deg(f) \leq k$ and its corresponding sub-functions $f_1 = f(x_1, ..., x_n, 1)$ and $f_2 = f(x_1, ..., x_n, -1)$. There are then two cases

1. Sub-functions f_1 and f_2 are both not identically 0. Notice that because $\deg(f) \leq k$, it must be the case that $\deg(f_1), \deg(f_2) \leq k$. Then, by the inductive hypothesis,

$$\begin{aligned}
\mathbb{P}_{x\in\{-1,1\}^n\}}[f_1(x) \neq 0] \geq 2^{-k} \\
\mathbb{P}_{x\in\{-1,1\}^n\}}[f_2(x) \neq 0] \geq 2^{-k}
\end{aligned}$$

   Notice that

$$\mathbb{P}_{x\in\{-1,1\}^{n+1}}[f(x) \neq 0] = \frac{1}{2}\left(\mathbb{P}_{x\in\{-1,1\}^n\}}[f_1(x) \neq 0] + \mathbb{P}_{x\in\{-1,1\}^n\}}[f_2(x) \neq 0]\right) \geq 2^{-k}$$

2. Without loss of generality, sub-function f_2 is identically 0. By the definition of differentiation

$$D_{n+1}f = \frac{f(x^{(n+1)\mapsto 1}) - f(x^{(n+1)\mapsto -1})}{2} = \frac{f_1 - f_2}{2} = \frac{f_1}{2}$$

   By the differentiation formula in chapter 2, $\deg(g) \leq k \implies \deg(D_i g) \leq k-1$. Therefore, $\deg(\frac{f_1}{2}) = \deg(f_1) = \deg(D_{n+1}f) \leq k-1$. By the inductive hypothesis then, $\mathbb{P}_x[f_1(x) \neq 0] \geq$

2^{1-k}. We can then conclude that

$$\begin{aligned}\mathbb{P}_{x\in\{-1,1\}^{n+1}}[f(x)\neq 0] &= \frac{1}{2}\left(\mathbb{P}_{x\in\{-1,1\}^n\}}[f_1(x)\neq 0] + \mathbb{P}_{x\in\{-1,1\}^n\}}[f_2(x)\neq 0]\right)\\ &\geq \frac{1}{2}\cdot 2^{1-k} = 2^{-k}\end{aligned}$$

Note that both sub-functions cannot be identically 0 because this would contradict the assumption that f is not identically 0. The two above cases are then exhaustive and show the claim.

$\square$

Problem 3.7

(a)

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

(b) Let $\widehat{f}_{J|z}(S)\neq 0$. Then $\sum_{T\subseteq\overline{J}}\widehat{f}(S\cup T)z^T\neq 0$ meaning $\widehat{f}(S\cup T)\neq 0$ for some $T\subseteq\overline{J}$. Now, observe that if $S,S'\subseteq J$ and $T,T'\subseteq\overline{J}$ then $S\cup T=S'\cup T' \iff S=S', T=T'$. In particular, each unique S for which $\widehat{f}_{J|z}(S)\neq 0$ corresponds to at least one $R=S\cup T$ for which $\widehat{f}(R)\neq 0$, and the same R will not be found for multiple sets S. We conclude that $\text{sparsity}(\widehat{f_{J|z}})\leq\text{sparsity}(\widehat{f})$, as desired.

$\square$

Problem 3.10 Consider the Fourier expansion of $D_i f$:

$$D_i f(x) = \sum_{\substack{S\subseteq[n]\\ S\ni i}}.$$

Thus, we can write $\widehat{D_i f}(S) = \begin{cases} 0 & \text{if } i \in S \\ \widehat{f}(S \cup \{i\}) & \text{if } i \notin S \end{cases}$. From Exercise 3.9, we know that $\hat{\|}D_i f\hat{\|}_\infty \leq \|D_i f\|_1$, which gives us:

$$\begin{aligned} \max_{S \ni i} |\widehat{f}(S)| &\leq \mathbb{E}_x[D_i f(x)] \\ &= \mathbb{E}_x[D_i f(x)^2] && (D_i f \to \{0,1\} \text{ since } f \text{ monotone}) \\ &= \mathbf{Inf}_i[f] \\ &= \widehat{f}(i). \end{aligned}$$

It follows that $\hat{\|}f\hat{\|}_\infty$ is achieved by an S of cardinality 0 or 1. $\square$

Problem 3.13

(a) A must obviously be nonempty, so let $a \in A$ be arbitrary. Since A is not affine, $A + a$ is not a subspace of $\mathbb{F}_2^n$, meaning there exist $b, c \in A + a$ for which $b + c \notin A + a$. Hence, $a + b, a + c \in A$ but $a + b + c \notin A$. Then if $B = \{0, b, c, b + c\} + a = \{a, a + b, a + c, a + b + c\}$, we have that B is an affine subspace of dimension two, and it intersects with A on three points (all except $a + b + c$).

(b) For the B obtained in (a), write $B = H + a$ for some subspace $H \leq \mathbb{F}_2^n$ and let $B \backslash A = \{b\}$. As shown previously, we have $\widehat{\varphi_B}(S) = \begin{cases} \chi_S(a) & \text{if } S \in H^\perp \\ 0 & \text{otherwise} \end{cases}$, and $\widehat{\varphi_b}(S) = \chi_S(a)$. Thus, $|\widehat{\psi}(S)| = |\widehat{\varphi_B}(S) - \frac{1}{2}\widehat{\varphi_b}(S)| = | \pm \frac{1}{2}\chi_S(a)| = 1/2$, meaning $\hat{\|}\psi\hat{\|}_\infty = \frac{1}{2}$.

(c) We have:

$$\begin{aligned} \langle \psi, f \rangle &= \langle \varphi_B, f \rangle - \frac{1}{2} \cdot \langle \varphi_b, f \rangle \\ &= \mathbb{E}_{x \sim B}[f(x)] - \frac{1}{2} \cdot f(b) \\ &= \mathbb{E}_{x \sim B}[f(x)] && (b \notin A) \\ &= 3/4. \end{aligned}$$

Then, we conclude:

$$\begin{aligned} \hat{\|}f\hat{\|}_1 &= \sum_{S \subseteq [n]} |\widehat{f}(S)| \\ &\geq 2 \cdot \sum_{S \subseteq [n]} \frac{1}{2} \cdot \widehat{f}(S) \\ &\geq 2 \cdot \sum_{S \subseteq [n]} \widehat{\psi}(S) \cdot \widehat{f}(S) && \text{(part (b))} \\ &= 2 \cdot \langle \psi, f \rangle && \text{(Plancherel)} \\ &= 3/2. \end{aligned}$$

$\square$

Problem 3.14

$$\hat{||}f\hat{||}_1 = \sum_{S\subseteq[n]} |\widehat{f}(S)| \le \left(\sum_{S\subseteq[n]} \widehat{f}(S)^2\right)^{\frac{1}{2}} \left(\sum_{S\subseteq[n]} 1^2\right)^{\frac{1}{2}} \qquad \text{(Cauchy Schwartz)}$$
$$\le 1\cdot 2^{n/2} \qquad (\mathbb{E}[f^2]\le 1)$$

Now, consider the inner product function $IP^{2n} : \{-1,1\}^{2n} \to \{-1,1\}$. From exercise 1.1, we know that $|\widehat{IP}(S)| = \frac{1}{2^n}$ for all $S \subseteq [2n]$. Therefore,

$$\begin{aligned}\hat{||}IP\hat{||}_1 &= \sum_{S\subseteq[2n]} |\widehat{IP}(S)| \\ &= \frac{1}{2^n}\cdot 2^{2n} = 2^n = 2^{2n/2}\end{aligned}$$

Therefore, for any $n \in \mathbb{N}$, IP^{2n} achieves equality (which also implies the bound is tight). $\square$

Problem 3.16 Consider set $\mathcal{F} = \{S|\widehat{f}(S) \ge \frac{\epsilon}{\hat{||}f\hat{||}_1}\}$. We first show that $|\mathcal{F}| \le \frac{\hat{||}f\hat{||}_1^2}{\epsilon}$. Suppose, for the sake of contradiction that $|\mathcal{F}| > \frac{\hat{||}f\hat{||}_1^2}{\epsilon}$. Then,

$$\sum_{S\in\mathcal{F}} \widehat{f}(S)^2 \ge |\mathcal{F}|\cdot \min_{S\in\mathcal{F}} \widehat{f}(S)^2 > \frac{\hat{||}f\hat{||}_1^2}{\epsilon}\cdot\frac{\epsilon}{\hat{||}f\hat{||}_1} = \hat{||}f\hat{||}_1$$

Contradiction. Finally, note that

$$\begin{aligned}\sum_{S\notin\mathcal{F}} \widehat{f}(S)^2 &\le \max_{S\notin\mathcal{F}} \widehat{f}(S)\cdot\sum_{S\notin\mathcal{F}}\widehat{f}(S) \\ &\le \frac{\epsilon}{\hat{||}f\hat{||}_1}\cdot \hat{||}f\hat{||}_1 = \epsilon\end{aligned}$$

and we can conclude that f is ϵ-concentrated on $\mathcal{F}$. $\square$

Problem 3.17 For any S, we have $|\widehat{g}(S)| \le |\widehat{g-f}(S)| + |\widehat{f}(S)|$. Squaring both sides, and summing over $\overline{\mathcal{F}}$, we have:

$$
\begin{aligned}
\sum_{S \notin \mathcal{F}} \widehat{g}(S)^2 &\le \sum_{S \notin \mathcal{F}} (|\widehat{g-f}(S)| + |\widehat{f}(S)|)^2 \\
&\le \sum_{S \notin \mathcal{F}} \widehat{g-f}(S)^2 + \sum_{S \notin \mathcal{F}} \widehat{f}(S)^2 + 2 \cdot \sum_{S \notin \mathcal{F}} |\widehat{g-f}(S)| \cdot |\widehat{f}(S)| \\
&\le \hat{\|}g-f\hat{\|}_2^2 + \sum_{S \notin \mathcal{F}} \widehat{f}(S)^2 + 2 \cdot \sum_{S \notin \mathcal{F}} |\widehat{g-f}(S)| \cdot |\widehat{f}(S)| \\
&= \|g-f\|_2^2 + \sum_{S \notin \mathcal{F}} \widehat{f}(S)^2 + 2 \cdot \sum_{S \notin \mathcal{F}} |\widehat{g-f}(S)| \cdot |\widehat{f}(S)| && \text{(Parseval's)} \\
&\le \varepsilon_1 + \varepsilon_2 + 2 \cdot \sum_{S \notin \mathcal{F}} |\widehat{g-f}(S)| \cdot |\widehat{f}(S)| \\
&\le \varepsilon_1 + \varepsilon_2 + 2\sqrt{\left(\sum_{S \notin \mathcal{F}} \widehat{g-f}(S)^2\right) \cdot \left(\sum_{S \notin \mathcal{F}} \widehat{f}(S)^2\right)} && \text{(Cauchy-Schwarz)} \\
&\le \varepsilon_1 + \varepsilon_2 + 2\sqrt{\varepsilon_1 \varepsilon_2} && \text{(same bounds as before)} \\
&\le 2(\varepsilon_1 + \varepsilon_2). && \text{(AM-GM inequality)}
\end{aligned}
$$

Hence, g is $2(\varepsilon_1 + \varepsilon_2)$-concentrated on $\mathcal{F}$, as desired. □

Problem 3.19 We assume that f can be computed by decision tree T. Then, notice that $-f(x)$ can be computed by $-T$, where $-T$ flips the outputs at leaf nodes of tree T. Decision tree $-T$ retains size s and depth k. Furthermore, the dual of f is defined as $f^\dagger = -f(-x)$. It suffices to show that $f(-x)$ can be computed by decision tree of size s and depth k. Decision tree $T(-)$ that flips every value along an edge of tree T computes $f(-x)$. Decision tree $T(-)$ retains size s and depth k. Therefore, decision tree $-T(-)$ computes $f^\dagger$ with size s and depth k. □

Problem 3.22 We observe that $T'(x) \neq T(x)$ only if $|C_p(x)| > \log(s/\varepsilon)$. Since a computation path of length ℓ is followed by at most $2^{n-\ell}$ inputs, each truncated computation path is followed by at most $2^{n-\log(s/\varepsilon)} = 2^n \cdot \frac{\varepsilon}{s}$ inputs. Since the number of computation paths is the number of leaves, namely s, it follows that $T'(x) \neq T(x)$ can only occur on at most $s \cdot 2^n \cdot \frac{\varepsilon}{s} = \varepsilon 2^n$ inputs, meaning the function computed by T' is ε-close to the function computed by T. □

Problem 3.23 Given function f that is computable by a decision tree, we construct a linear threshold function and prove its equivalence with f. By definition, a decision list only contains one path of internal nodes, which we denote $x_{i_0}, ..., x_{i_k}$ where x_{i_l} is at level l. Notice that $x_{i_0}, ..., x_{i_{k-1}}$ each connect to exactly one leaf while x_{i_k} connects to two leaves. To see this, suppose that x_{i_l} is the first internal node in the decision list that connects to two leaves for some $l \in \{0, ..., k-1\}$. Then, internal nodes $x_{i_{l+1}}$ onwards cannot exist. Conversely, if x_{i_k} did not connect to two leaf nodes, then another internal node $x_{i_{k+1}}$ must exist.

Now consider x_{i_l} and define t_l corresponding to leaf edges as follows

$$t_l = \begin{cases} \frac{x_{i_l}+1}{2} \cdot 2^{n-l} \text{ if } x_{i_l} = 1 \implies f(x) = 1 \\ \frac{x_{i_l}+1}{2} \cdot -2^{n-l} \text{ if } x_{i_l} = 1 \implies f(x) = -1 \\ \frac{x_{i_l}-1}{2} \cdot -2^{n-l} \text{ if } x_{i_l} = -1 \implies f(x) = 1 \\ \frac{x_{i_l}-1}{2} \cdot 2^{n-l} \text{ if } x_{i_l} = -1 \implies f(x) = -1 \end{cases}$$

Then, let $g(x) = \text{sgn}(\sum_{l=0}^{k-1} t_l + t_k^{(1)} + t_k^{(2)})$ where $t_k^{(1)}$ and $t_k^{(2)}$ denote the two leaf edges for x_{i_k}. We claim that $\forall x.g(x) = f(x)$. To see this, consider $g(x)$ and $f(x)$ for arbitrary input x and let the traversal down the decision list terminate at a leaf on level l. This means $t_s = 0$ for all $s \in \{0, ..., l-2\}$. In particular, $|t_{l-1}| = 2^{n+1-l}$. Then, no matter the satisfiability of leaf edges level l and below, $\text{sgn}(\sum_{l=0}^{k-1} t_l + t_k^{(1)} + t_k^{(2)}) = \text{sgn}(t_l)$ since $\sum_{r=0}^{l} 2^r = 2^{r+1} - 1$. Therefore, $g(x)$ correctly reports the value of the first satisfied leaf edge in the decision list computing $f(x)$ and we can conclude that f is therefore a linear threshold function.

$\square$

Problem 3.24 Consider $\square$

Problem 3.25 Let f be computable by read-once decision tree T. Notice that T must be complete. For, if not, there exists an internal node v with fewer than two children. Suppose v lies at level l. For $l = k-1$, v lies at the deepest level and does not have two leaf children – a contradiction. For $l < k-1$, v has exactly one child which we denote w. But this means that any path reaching v must reach w, giving rise to an equivalent simplified tree T' with node v spliced out. Tree T', however, then has a leaf to node path via w of length less than k – a contradiction.

We show that for internal node v representing variable x_i at level $l \in \{0, ..., k-2\}$ we have $\mathbf{Inf}_i[f] = 2^{-l-1}$. To see this, first notice both the left and right subtrees of v contain equal amounts of -1 and 1 leaf nodes. We denote $X_v = \{x \in \{0,1\}^n | \text{The path of } x \text{ in T passes through } v\}$. Notice that $|X_v| = 2^{n-l}$. Now consider flipping coordinate i for an $x \in X_v$. Because T is read-once, the path x takes in the right subtree of v is independent of the path taken in the left subtree. Then, $\mathbb{P}_{x \in X_v}[f(x^{\oplus i}) = 1] = \mathbb{P}_{x \in X_v}[f(x^{\oplus i}) = -1] = \frac{1}{2}$. Therefore, $\mathbb{P}_{x \in X_v}[f(x^{\oplus i}) \neq f(x)] = \frac{1}{2}$. The influence of coordinate i is then $\frac{\frac{1}{2} \cdot 2^{n-l}}{2^n} = 2^{-l-1}$.

For internal node s representing coordinate j on level $l = k-1$ have influence $2^{n-l} = 2^{n-(k-1)}$. By a similar, argument, $2^{n-(k-1)}$ inputs x reach node s at the deepest level. Because each node at level k has a leaf of each -1 and 1, it follows that $\mathbb{P}_{x \in X_s}[f(x^{\oplus j}) \neq f(x)] = 1$. The influence of coordinate j is then $\frac{2^{n-k+1}}{2^n} = 2^{1-k}$.

The total influence is then

$$\begin{aligned}
\mathbf{I}[f] &= 2^{-n}\left(2^{n-k+1}\cdot 2^{k-1} + \sum_{l=0}^{k-2} 2^{n-l-1}\cdot 2^{l}\right) && \text{(Completeness of T)}\\
&= 2^{-n}\left(2^{n} + (k-2)\cdot 2^{n-1}\right)\\
&= 1 + \frac{k-2}{2} = \frac{k}{2}
\end{aligned}$$

This can also be seen via expected sensitivity. $\mathbf{I}[f] = \mathbb{E}_x[\mathbf{sens}(x)] = \frac{k}{2}$ since every input x traverses a path of length k in T, where the probability that any variable is pivotal is $\frac{1}{2}$ (due to the independence of subtrees and even leaf count). $\square$

Problem 3.28 By definition, we have:

$$\widehat{f^{\subseteq J}}(S) = \mathbb{E}_{x\sim\{-1,1\}^n}[f^{\subseteq J}(x)\cdot\chi_S(x)] = \mathbb{E}_{x\sim\{-1,1\}^n}[\mathbb{E}_{y\sim\{-1,1\}^{\overline{J}}}[f(x_J,y)]\cdot\chi_S(x)].$$

Now, if $S \subseteq J$ then $\chi_S(x)$ depends only on x_J, so the above expression is equal to:

$$\mathbb{E}_{x\sim\{-1,1\}^J}[\mathbb{E}_{y\sim\{-1,1\}^{\overline{J}}}[f(x_J,y)]\cdot\chi_S(x)] = \mathbb{E}_{z\in\{-1,1\}^n}[f(z)\cdot\chi_S(z)] = \widehat{f}(S).$$

If instead $S \nsubseteq J$, then we can find some $i \in S\backslash J$. Then, we can rewrite the previous expression as:

$$\mathbb{E}_{x_{\neq i}\in\{-1,1\}^{n-1}}\left[\mathbb{E}_{y\sim\{-1,1\}^{\overline{J}}}\left[\mathbb{E}_{x_i\sim\{-1,1\}}[f(x_J,y)\cdot\chi_S(x)]\right]\right]$$

Since $i \in S$, we have $\chi_S(x) = -\chi_S(x^{\oplus i})$, and since $i \notin J$, we have $f(x_J,y) = f((x^{\oplus i})_J,y)$. Therefore, flipping x_i from 1 to -1 will negate the random variable in the innermost expectation, meaning the innermost expectation (over x_i) evaluates to 0, and hence $\widehat{f^{\subseteq J}}(S) = 0$ in this case. The desired Fourier expansion follows. $\square$

Problem 3.30 Since the leaf b is a depth-k, the set of coordinates J encountered on the computation path to b satisfies $|J| \leq k$. Moreover, J has the property that $f(a,y) = b$ for any $a \in \{-1,1\}^J$, $y \in \{-1,1\}^{\overline{J}}$. It follows that $x_J = a \implies f^{\subseteq J}(x) = b$. Thus $f^{\subseteq J}(x) = \sum_{S\subseteq J}\widehat{f}(S)\chi_S(x) = b$. Let $x \in \{-1,1\}^n$ such that $x_J = a$. Then:

$$\begin{aligned}
|b| &= \left|\sum_{S\subseteq J}\widehat{f}(S)\chi_S(x)\right|\\
&\leq \sum_{S\subseteq J}|\widehat{f}(S)\chi_S(x)|\\
&= \sum_{S\subseteq J}|\widehat{f}(S)|.
\end{aligned}$$

Therefore, there exists some $S \subseteq J$ for which $|\widehat{f}(S)| \geq \frac{|b|}{2^{|J|}} \geq \frac{|b|}{2^k}$. Hence $\hat{||}f\hat{||}_\infty \geq \frac{|b|}{2^k}$, as desired. $\square$

Problem 3.33 To learn f with 0 error, we must randomly sample $(x, f(x))$ for all x. We denote random variable X_i to be the number of draws required to obtain the ith unique pair, given that we have obtained $i-1$ unique pairs. Notice then that $X = \sum_{i=1}^{2^n} X_i$ is the total number of random draws needed to learn f with 0 error. In expectation

$$\mathbb{E}[X] = \mathbb{E}[\sum_{i=1}^{2^n} X_i] = \sum_{i=1}^{2^n} \mathbb{E}[X_i]$$

Notice that X_i is a geometric random variable with parameter $p = \frac{2^n-i+1}{2^n}$. Therefore, $\mathbb{E}[X_i] = \frac{2^n}{2^n-i+1}$ and

$$\begin{aligned}\mathbb{E}[X] &= 2^n \sum_{i=1}^{2^n} \frac{1}{2^n - i + 1} \\ &= 2^n \sum_{i=1}^{2^n} \frac{1}{i} \\ &\leq 2^n \log(2^n)\end{aligned}$$

Therefore, we require $\tilde{O}(2^n)$ draws in expectation to learn f with zero error. Then, by the Markov bound

$$\mathbb{P}[\sum_{i=1}^{2^n} X_i \geq 10n2^n] \leq \frac{n2^n}{10n2^n} = 0.1$$

So f can be learned with 0 error with 0.9 probability using $\tilde{O}(2^n)$ random samples. □

Problem 3.40

(a) Our verification algorithm will work as follows: sample k (to be determined later) random examples $(x, f(x))$ and compute each $h(x)$. If the proportion of examples where $f(x) \neq h(x)$ is less than $\frac{3\varepsilon}{4}$ then output 'YES', and otherwise output 'NO'. Formally, for $1 \leq i \leq k$ let $X_i = 1$ if $f(x_i) \neq h(x_i)$ and 0 otherwise. Let $Y = \sum_{i=1}^{k} X_i$ and $\overline{X} := \mathbb{E}[Y] = \text{dist}(f, h) \cdot k$. Observe that in order for our algorithm to mistakenly reject h for which $\text{dist}(f, h) \leq \varepsilon/2$, or for it to mistakenly accept h for which $\text{dist}(f, h) > \varepsilon$, we would need $|Y - \overline{X}| \geq \frac{\varepsilon k}{4}$. But by the Chernoff bound, we have:

$$\begin{aligned}\Pr\big[|Y - \overline{X}| \geq \varepsilon k/4\big] &\leq e^{-ck\varepsilon^2} \\ &= e^{-\log(1/\delta)} && (\text{set } k = \tfrac{\log(1/\delta)}{c\varepsilon^2}) \\ &= \delta.\end{aligned}$$

Hence, our algorithm works with the desired error upper bound. Now, since we request $k = \log(1/\delta) \cdot \text{poly}(1/\varepsilon)$ examples, each of which taking $O(n + T)$ time (parsing x, computing h), in total our verification algorithm takes $\text{poly}(n, T, 1/\varepsilon) \cdot \log(1/\delta)$ time, as desired.

(b) need to finish

$\square$

Problem 3.43

(a) Observe that

$$\mathbb{P}_x[\gamma \cdot x = \gamma' \cdot x] = \mathbb{P}_x[(\gamma \oplus \gamma') \cdot x = 0]$$

where we denote $\gamma'' \in \{0,1\}^n$. Since $\gamma \neq \gamma'$, there must exist some index $i \in [n]$ for which $\gamma''_i = 1$. Now consider all inputs $x \in \{0,1\}^{n-1}$ that map to indices $[n]\backslash\{i\}$. Then,

$$\gamma'' \cdot x = \left[\sum_{j \in [n]\backslash\{i\}} \gamma''_j x_j\right] + x_i$$

We establish a bijection between the set of x such that $\gamma'' \cdot x = 0$ and x such that $\gamma'' \cdot x = 1$. Let $x \in \{0,1\}^{n-1}$ correspond to the indices $[n]\backslash\{i\}$. Then, $c = \left[\sum_{j\in[n]\backslash\{i\}} \gamma''_j x_j\right] \in \{0,1\}$. Then, $\gamma'' \cdot x = c + x_i$ and the two settings of x_i yield different results to the dot product. We then have a 1-1 correspondence between the two sets and can therefore conclude that

$$\mathbb{P}_x[\gamma \cdot x = \gamma' \cdot x] = \frac{1}{2}$$

(b) We begin by observing that if $x^{(1)}, ..., x^{(m)}$ are the n basis vectors of $\widehat{\mathbb{F}_2^n}$ then

$$\begin{aligned}\forall i \in [n].\ \gamma \cdot x^{(i)} = \gamma' \cdot x^{(i)} &\iff \forall i \in [n].\ (\gamma \oplus \gamma') \cdot x^{(i)} = 0 \\ \iff \gamma \oplus \gamma' \in \widehat{F_2^n}^{\perp} &\iff \gamma \oplus \gamma' = 0 \iff \gamma = \gamma'\end{aligned}$$

So drawing n linearly independent vectors will guarantee the equivalence of γ and γ'. Given that we have drawn k linearly independent vectors, its span has size 2^k. Then, the probability of drawing a new linearly independent vector is $\frac{2^n - 2^k}{2^n}$. Therefore, it takes $\frac{2^n}{2^n - 2^k}$ random draws in expectation to find the $k+1$th new linearly independent vector. We denote random variable X_i to count the number of draws to find the ith linearly independent vector given $i-1$ have been found. By the Markov bound,

$$\begin{aligned}\mathbb{P}[X = \sum_{i=1}^{n} X_i \geq 20n] &\leq \frac{\mathbb{E}[X]}{20n} = \frac{1}{20n}\sum_{i=1}^{n}\frac{2^n}{2^n - 2^{i-1}} \\ &\leq \frac{2n}{20n} = \frac{1}{10}\end{aligned}$$

Therefore, for $C \geq 20$, $x^{(1)}, ..., x^{(m)}$ will contain a basis for $\widehat{\mathbb{F}_2^n}$ with at least 0.9 probability. Then $\forall i \in [n].\ \gamma \cdot x^{(i)} \iff \gamma = \gamma'$.

(c) Linear functions $f : \{0,1\}^n \to \{0,1\}$ are of the form $f(x) = a \cdot x$. Consider drawing $m = Cn$ random $(x, f(x))$ pairs for C computed in part (b). With high probability, a basis will exist in this set. We construct matrix $M \in \{0,1\}^{n\times n}$ with row i equal to $x_b^{(i)}$, the ith basis vector, and vector v with $v_i = f(x_b^{(i)})$. We then solve the system

$$M\gamma = v$$

by computing $\gamma = vM^{-1}$. Notice that M has full rank since it comprises of a basis of $\widehat{\mathbb{F}_2^n}$ and is thus invertible. Matrix inversion and multiplication both take roughly $O(n^w)$ time.

□

4 DNF Formulas and Small Depth Circuits

Problem 4.1 For a function $f : \{0,1\}^n \to \{0,1\}$, we construct the following DNF ϕ:

$$\phi = \bigvee_{(a_1,\ldots,a_n)\in f^{-1}(1)} \bigwedge_{i=1}^{n} \begin{cases} x_i & \text{if } a_i = 1 \\ \overline{x_i} & \text{if } a_i = 0 \end{cases}$$

By construction, we have $f(a_1,\ldots,a_n) = 1 \iff \phi(a_1,\ldots,a_n) = 1$, meaning $f = \phi$. Moreover, ϕ consists of at most 2^n terms, each of which has width n, as desired. □

Problem 4.2 Consider CNF f given by

$$f = \bigwedge_{c\in C} \bigvee_{i\in c} x_i$$

where C is the set of clauses and each $c \in C$ is set of variable indices. Then notice that

$$\begin{aligned} f^\dagger(x) = \neg f(\neg x) &= \neg \bigwedge_{c\in C} \bigvee_{i\in c} \neg x_i \\ &= \bigvee_{c\in C} \bigwedge_{i\in c} x_i && \text{(DeMorgan's law)} \end{aligned}$$

which is exactly the DNF resulting from swapping ORs and ANDs. □

Problem 4.3 If ϕ is a monotone DNF, we observe that flipping any coordinate of the input from False to True cannot make any term False, since the AND function is monotone. Therefore, it cannot turn the entire DNF formula from True to False, meaning ϕ computes a monotone function.

If $f : \{0,1\}^n \to \{0,1\}$ is a monotone function, we slightly adapt the construction from Problem 4.1:

$$\phi = \bigvee_{(a_1,\ldots,a_n)\in f^{-1}(1)} \bigwedge_{i=1}^{n} \begin{cases} x_i & \text{if } a_i = 1 \\ 1 & \text{if } a_i = 0 \end{cases}.$$

Here, the 1's in each term can be omitted, leaving a disjunction of conjunctions of unnegated variables and hence a monotone DNF ϕ. To show that ϕ computes f, first observe that $f(a) = 1$

implies that $(a_1, \ldots, a_n) \in f^{-1}(1)$, meaning some term of ϕ is true and hence $\phi(a) = 1$. Therefore, it suffices to show that $\phi(a) = 0$ whenever $f(a) = 0$. Suppose for contradiction that $f(a) = 0$ and $\phi(a) = 1$ for some $a \in \{0,1\}^n$. Since $\phi(a) = 1$, there must exist some term in ϕ that is true at a. By construction, then, there is some $b = (b_1, \ldots, b_n) \in \{0,1\}^n$ for which $f(b) = 1$ and $b_i = 1 \implies a_i = 1$ for all i. But then $b \leq a$ coordinate-wise, and applying monotonicity of f yields $1 = f(b) \leq f(a) = 0$, a contradiction. Hence, $f = \phi$ meaning any monotone function can be computed by a monotone DNF.

The nonmonotone DNF $\phi = (x_1 \wedge \overline{x_2}) \vee (x_1 \wedge x_2)$ computes the monotone function $f(x_1, x_2) = x_1$. Hence, a nonmonotone DNF may compute a monotone function. □

Problem 4.4

(a) We first remark that the statement of Exercise 3.30 can easily be adapted to say: "Suppose $f : \{-1,1\}^n \to \mathbb{R}$ is computable by a DNF that has a term of width k. Then $\hat{\|} f \hat{\|}_\infty \geq 2^{-k}$". We justify this adaptation by referencing the proof supplied in this document, and considering the restriction $f^{\subseteq J}$ where J are the literals in a term of width k instead of the literals in a branch of depth k. Moreover, we observe that the proof actually gave us the existence of some S where $|S| \leq k$ and $|\widehat{f}(S)| \leq \frac{|b|}{2^k}$. Note that in the case of a DNF, $b = \pm 1$ so $|b| = 1$ necessarily.

□

Problem 4.5

□

Problem 4.7

□

Problem 4.9 Observe that

$$\begin{aligned}
\mathbb{E}_{J,z}[\mathbf{Inf}_i[f_{J|z}]] &= \mathbb{E}_J[\mathbb{E}_z[\mathbf{Inf}_i[f_{J|z}]]] + \mathbb{E}_J[\mathbb{E}_z[\mathbf{Inf}_i[f_{J|z}]]] \\
&= \mathbb{E}_J[\mathbb{E}_z[\mathbf{Inf}_i[f_{J|z}]|i \in J]] \cdot \mathbb{P}_J[i \in J] + \mathbb{E}_J[\mathbb{E}_z[\mathbf{Inf}_i[f_{J|z}]|i \notin J]] \cdot \mathbb{P}_J[i \notin J] \\
&= \delta \cdot \mathbb{E}_J[\mathbb{E}_z[\mathbf{Inf}_i[f_{J|z}]|i \in J]] \\
&= \delta \cdot \mathbb{E}_J\left[\frac{\sum_{z \in \{0,1\}^{|\bar{J}|}} \mathbf{Inf}_i[f_{J|z}]}{2^{|\bar{J}|}}\right]
\end{aligned}$$

For every $x \in \{x \in \{0,1\}^n | f(x) \neq f(x^{\oplus i})\}$, consider the mapping $g : \{0,1\}^n \to \{0,1\}^{|\bar{J}|}$ such that $g : x \mapsto x_{\bar{J}}$. Let $I_i = \{x \in \{0,1\}^n | f(x) \neq f(x^{\oplus i}\}$ and $I_{iz} = \{x \in \{0,1\}^{|J|} | f(x|z) \neq f(x^{\oplus i}|z)\}$. Then,

$$\mathbf{Inf}_i[f_{J|z}] = \frac{|I_{iz}|}{2^{|\bar{J}|}}$$

where $x|z$ denotes concatenation. Then it is true that $x \in I_i \wedge g(x) = z \iff x_J \in I_{iz}$. Therefore,

$$
\begin{aligned}
&= \delta \cdot \mathbb{E}_J \left[\sum_{z \in \{0,1\}^{|\bar{J}|}} \frac{|I_{iz}|}{2^{|J|} 2^{|\bar{J}|}} \right] \\
&= \frac{\delta}{2^n} \cdot \sum_{z \in \{0,1\}^{|\bar{J}|}} |I_{iz}| \\
&= \frac{\delta}{2^n} \cdot |I_i| = \delta \cdot \frac{|I_i|}{2^n} = \delta \cdot \mathbf{Inf}_i[f]
\end{aligned}
$$

□

Problem 4.11

□

Problem 4.16

□

5 Majority and Threshold functions