Exercises 5 & 6

Analysis of Boolean Functions, Autumn 2025, University of Chicago
Instructor: William Hoza (williamhoza@uchicago.edu)

Submission. Solutions are due **Friday, October 24** at 11:59pm Central time. Submit your solutions through Gradescope. You are encouraged, but not required, to typeset your solutions using a LaTeX editor such as Overleaf.

The policies below can also be found on the course webpage.

Collaboration. You are encouraged to collaborate with your classmates on exercises, but you must adhere to the following rules.

- Work on each exercise on your own for at least five minutes before discussing it with classmates.
- Feel free to explain your ideas to your classmates in person, and feel free to use whiteboards/chalkboards/etc. However, do not share any written/typeset solutions with your classmates for them to study on their own time. This includes partial solutions.
- Write your solutions on your own. While you are writing your solutions, do not consult any notes that you might have taken during discussions with classmates.
- In your write-up, list any classmates who helped you figure out the solution. The fact that student A contributed to student B's solution does not necessarily mean that student B contributed to student A's solution.

Permitted Resources for Full Credit. In addition to discussions with me and discussions with classmates as discussed above, you may also use the course textbook, any slides or notes posted in the "Course Timeline" section of the course webpage, and Wikipedia. If you wish to receive full credit on an exercise, you may not use any other resources.

Outside Resources for Partial Credit. If you wish, you may use outside resources (ChatGPT, Stack Exchange, etc.) to solve an exercise for partial credit. If you decide to go this route, you must make a note of which outside resources you used when you were working on each exercise. You must disclose using a resource even if it was ultimately unhelpful for solving the exercise. Furthermore, if you consult an outside resource while working on an exercise, then you must not discuss that exercise with your classmates.

In this exercise (based on Exercise 3.45 in the textbook), you will study an application of Fourier analysis to cryptography. Informally, a *one-way function* is a function that is easy to compute on all inputs, but hard to invert on even a small fraction of inputs. It isn't terribly difficult to cook up good candidate one-way functions. For example, the function $f(x, y) = 1x \cdot 1y$ is a solid candidate, where $1x$ and $1y$ are equal-length binary representations of integers and $\cdot$ is integer multiplication.

Meanwhile, informally, a cryptographic *pseudorandom generator* (PRG) is an efficient algorithm that uses a truly random "seed" to sample a longer sequence of bits that "appear random" from the perspective of any efficient observer. This definition seems much stronger than the definition of a one-way function. However, the celebrated Håstad-Impagliazzo-Levin-Luby (HILL) theorem says that if there exists a one-way function, then there exists a cryptographic PRG.

In this exercise, you will prove a special case of the HILL theorem in which we assume the existence of a one-way *permutation*, i.e., a one-way function $f\colon \{0,1\}^n \to \{0,1\}^n$ that is also a bijection. The PRG $G\colon \{0,1\}^{2n} \to \{0,1\}^{2n+1}$ is given by

$$G(r, s) = \left(r, f(s), \bigoplus_{i=1}^{n} r_i s_i\right).$$

Clearly, the first $2n$ bits of the output are uniform random. To prove that G is a good PRG, the main thing is to prove that the last bit, $\bigoplus_{i=1}^{n} r_i s_i$, is "unpredictable" for any efficient algorithm that is given $(r, f(s))$. Let $A\colon \{0,1\}^{2n} \to \{0,1\}$, and assume that A does a decent job of predicting $\bigoplus_{i=1}^{n} r_i s_i$:

$$\Pr_{r,s}\left[A(r, f(s)) = \bigoplus_{i=1}^{n} r_i s_i\right] \geq \frac{1}{2} + \frac{1}{n^{100}}.$$

We will prove that A is not efficiently computable (under the assumption that f is a one-way permutation).

Exercise 5 (10 points)**.**

(a) Prove that there exists $B \subseteq \{0,1\}^n$ such that $|B| \geq 2^n/\operatorname{poly}(n)$ and for every $s \in B$, we have

$$\Pr_{r}\left[A(r, f(s)) = \bigoplus_{i=1}^{n} r_i s_i\right] \geq \frac{1}{2} + \frac{1}{\operatorname{poly}(n)}.$$

(b) For each $s \in \{0,1\}^n$, define $A_s\colon \{0,1\}^n \to \{\pm 1\}$ by $A_s(r) = (-1)^{A(r,s)}$. Prove that for every $s \in B$, we have $\widehat{A_{f(s)}}(s) \geq 1/\operatorname{poly}(n)$. Here we are thinking of s as the indicator vector of a set $s \subseteq [n]$.

(c) Prove that there is a $\operatorname{poly}(n)$-time randomized algorithm that uses query access to f and A and outputs $f^{-1}(y)$ with probability $1/\operatorname{poly}(n)$, given a uniform random $y \in \{0,1\}^n$. (If A were efficiently computable, this would violate the assumption that f is a one-way permutation.)

Exercise 6 (10 points)**.**

(a) Let $f\colon \{\pm 1\}^n \to \{\pm 1\}$. Prove that

$$f(x) \cdot \mathrm{sens}_f(x) = \sum_{S \subseteq [n]} |S| \cdot \widehat{f}(S) \cdot \chi_S(x).$$

(b) Let $f\colon \{\pm 1\}^n \to \{\pm 1\}$. In class, we showed that $\mathbb{E}_x[\mathrm{sens}_f(x)] = \mathbb{E}_{S \sim \mathcal{S}_f}[|S|]$. Prove that

$$\mathop{\mathbb{E}}_{x}[\mathrm{sens}_f(x)^2] = \mathop{\mathbb{E}}_{S \sim \mathcal{S}_f}[|S|^2].$$

(c) Let $f\colon \{\pm 1\}^n \to \{\pm 1\}$ and suppose $\mathrm{I}[f] = O(1)$. In class, we showed that f is concentrated on Fourier coefficients of degree $O(1)$, hence f is concentrated on $\mathrm{poly}(n)$ Fourier coefficients. Later in class, we will prove that f is in fact concentrated on just $O(1)$ Fourier coefficients! In this exercise, you will show that this improvement crucially relies on the assumption that f is $\{\pm 1\}$-valued.

For every constant $k \in \mathbb{N}$ and for all large enough n, prove that there exists $f\colon \{\pm 1\}^n \to \mathbb{R}$ such that $\|f\|_2 = 1$, $\mathbb{E}_{S \sim \mathcal{S}_f}[|S|] = k$, and every collection on which f is 0.1-concentrated has $\Omega(n^k)$ sets.