Analysis of Boolean Functions 27 Sep, 2017

Problem Set 2

- Due Date: **16 Oct, 2017**
- Turn in your problem sets electronically (LaTeX, pdf or text file) by email. If you submit handwritten solutions, start each problem on a fresh page.
- Collaboration is encouraged, but all writeups must be done individually and must include names of all collaborators.
- Refering sources other than the text book and class notes is strongly discouraged. But if you do use an external source (eg., other text books, lecture notes, or any material available online), ACKNOWLEDGE all your sources (including collaborators) in your writeup. This will not affect your grades. However, not acknowledging will be treated as a serious case of academic dishonesty.
- The points for each problem are indicated on the side.
- Problems 2–6 are from O'Donnell's book/course.
- Be clear in your writing.

---

1. **[Dictatorship Test for "not-one"]** (3+4+3+4+4=18)

   In this problem, we will design a dictatorship test for the following "not-one" predicate $P : \{1,-1\}^n \to \{0,1\}$ defined as follows:

$$P(x,y,z) = \begin{cases} 0 & \text{if exactly one of } x, y \text{ and } z \text{ is } -1, \\ 1 & \text{otherwise} \end{cases}.$$

   Consider the following dictatorship test DICT_P which checks if a given (folded) function $f : \{1,-1\}^m \to \{1,-1\}$ is a dictator. Recall that f is folded if $f(x) = -f(-x)$

   ε-DICT_P^f : Oracle access $f : \{1,-1\}^m \to \{1,-1\}$

   1. Choose $x, y \in_R \{1,-1\}^m$
   2. Pick $\eta \in \{1,-1\}^L$ (noise vector) as follows:
$$\eta_i \longleftarrow \begin{cases} 1 & \text{with probability } 1-\varepsilon \\ -1 & \text{with probability } \varepsilon \end{cases}$$
   3. Set $z \leftarrow x \odot y \odot \eta$, (i.e., $z_i = x_i y_i \eta_i, \forall i \in [m]$)
   4. Accept if $P(f(x), f(y), f(z)) = 1$.

   (a) (Completeness) Prove that if f is a dictator, then $\Pr[\varepsilon\text{-}\mathsf{DICT}_P^f \text{ accepts }] = 1 - \varepsilon$.

   (b) Show that the Fourier expansion of P is given by

$$P(a,b,c) = \frac{5 - a - b - c + ab + bc + ca + 3abc}{8}.$$

Observe that this implies that a random assignment satisfies P with probability $5/8$.

In the rest of the problem, we will prove the soundness of the test ε-DICT_P: if $\Pr[\varepsilon\text{-}\mathsf{DICT}_P^f \text{ accepts }] \geq \frac{5}{8} + \delta$, then then f "resembles" a dictator in the following sense:

$$\exists S \subseteq [m], \text{ such that } |S| \leq O\left(\frac{1}{\varepsilon}\log\frac{1}{\delta}\right) \text{ and } |\widehat{f}(S)| \geq \Omega(\delta).$$

(c) Show that $\Pr[\varepsilon\text{-}\mathsf{DICT}_P^f \text{ accepts }]$ is equal to

$$\frac{5 - \mathbb{E}[f(x)] - \mathbb{E}[f(y)] - \mathbb{E}[f(z)] + \mathbb{E}[f(x)f(y)] + \mathbb{E}[f(y)f(z)] + \mathbb{E}[f(z)f(x)] + 3\mathbb{E}[f(x)f(y)f(z)]}{8}.$$

(d) Argue that if $\Pr[\varepsilon\text{-}\mathsf{DICT}_P^f \text{ accepts }] \geq \frac{5+\delta}{8}$, then $\mathbb{E}[f(x)f(y)f(z)] \geq \delta/3$.

(e) Conclude that there exists $\exists S \subseteq [m]$, such that $|S| \leq O\left(\frac{1}{\varepsilon}\log\frac{1}{\delta}\right)$ and $|\widehat{f}(S)| \geq \Omega(\delta)$.

2. **[Fourier concentration of decision trees]** (15)

Let $f : \mathbb{F}_2^n \to \{1, -1\}$ be computed by a decision tree T of size s and let $\varepsilon \in (0, 1]$. Suppose each path in T is truncated (if necessary) so that its length does not exceed $\log(s/\varepsilon)$; new leaves with labels 1 and 1 may be created in an arbitrary way as necessary. Show that the resulting decision tree T' computes a function which is ε-close to f. Conclude that the Fourier spectrum of f is ε-concentrated on degree up to $\log(s/\varepsilon)$.

3. **[Max Influence pre-KKL Theorem]** (3+4+6+5=18)

Let $f : \{1, -1\}^n \to \{1, -1\}$ be unbiased (i.e., $\mathbb{E}[f] = 0$), and let $\mathbf{MaxInf}[f]$ denote $\max_{i \in [n]}\{\mathbf{Inf}_i[f]\}$. In the second half of the course, we will prove the KKL Theorem that shows that $\mathbf{MaxInf}[f] \geq \Omega(\frac{\log n}{n})$. In 1987, this was still a conjecture; all that was known were the following results, independently observed by Alon and by Chor and Geréb-Graus…

(a) Use the Poincaré Inequality (i.e., $\mathbf{Var}[f] \leq \mathbb{I}[f]$) to show $\mathbf{MaxInf}[f] \geq 1/n$.

(b) Prove $|\widehat{f}(i)| \leq \mathbf{Inf}_i[f]$ for all $i \in [n]$. (Hint: consider $\mathbb{E}[|D_i f|]$.)

(c) Prove that $\mathbb{I}[f] \geq 2 - n\,\mathbf{MaxInf}[f]^2$. (Hint: first prove $\mathbb{I}[f] \geq \mathbf{W}^1[f] + 2(1 - \mathbf{W}^1[f])$ and then use the previous exercise.)

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

(Later in 1987, Chor and Geréb-Graus managed to improve the lower bound to $\frac{3}{n} - o(1/n)$.)

4. **[Embedding ℓ_1 into ℓ_2 (Enflo, 1970)]** (4+6+6=16)

The Hamming distance $\Delta(x, y) = \#\{i : x_i \neq y_i\}$ on the discrete cube $\{1, -1\}^n$ is an example of an ℓ_1 *metric space*. 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:

$$\|F(x) - F(y)\|_2 \geq \Delta(x, y) \text{ for all } x, y; \qquad \text{("no contraction")}$$

$$\|F(x) - F(y)\|_2 \leq D \cdot \Delta(x, y) \text{ for all } x, y. \qquad \text{("expansion at most } D\text{")}$$

In this problem you will show that the least distortion possible is $D = \sqrt{n}$.

(a) Recalling the definition of f^{odd} from Homework 1, show that for any $f : \{1,-1\}^n \to \mathbb{R}$ we have $\|f^{\text{odd}}\|_2^2 \le \mathbb{I}[f]$ and hence

$$\mathbb{E}_{\boldsymbol{x}}[(f(\boldsymbol{x}) - f(-\boldsymbol{x}))^2] \le \sum_{i=1}^{n} \mathbb{E}_{\boldsymbol{x}}\left[\left(f(\boldsymbol{x}) - f(\boldsymbol{x}^{\oplus i})\right)^2\right].$$

(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 $\sqrt{n}$.

5. **[Khintchine–Kahane inequality (Latała–Oleszkiewicz, 1994)]** (6+6+6=18)

Let V be a vector space with norm $\|\cdot\|$ and fix $w_1, \ldots, w_n \in V$. Define $g : \{1,-1\}^n \to \mathbb{R}$ by $g(x) = \|\sum_{i=1}^n x_i w_i\|$.

(a) Let $\mathrm{L} := \mathrm{L}_1 + \mathrm{L}_2 + \cdots + \mathrm{L}_n$ be the Laplacian operator where L_i is as defined in Definition 2.25. Show that $\mathrm{L}g \le g$ pointwise. (Hint: triangle inequality.)

(b) Deduce $2\,\mathbf{Var}[g] \le \mathbb{E}[g^2]$ and thus the *Khintchine–Kahane inequality*:

$$\mathbb{E}_{\boldsymbol{x}}\left[\left\|\sum_{i=1}^{n} \boldsymbol{x}_i w_i\right\|\right] \ge \frac{1}{\sqrt{2}} \cdot \mathbb{E}_{\boldsymbol{x}}\left[\left\|\sum_{i=1}^{n} \boldsymbol{x}_i w_i\right\|^2\right]^{1/2}.$$

(Hint: first, show that the improved Poincaré inequality $\mathbf{Var}[f] \le \frac{1}{2}\mathbb{I}[f]$ holds whenever $f : \{1,-1\}^n \to \mathbb{R}$ is even, as defined in Homework 1.)

(c) Show that the constant $\frac{1}{\sqrt{2}}$ above is optimal (Hint: take $V = \mathbb{R}$ and $n = 2$.)

6. **[Low-Degree algorithm's hypothesis]** (8+7=15)

(a) The 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. Show that one can instead simply draw a single batch $\mathcal{E}$ of random examples $(x, f(x))$ of size $\mathrm{poly}(n^k, 1/\varepsilon)$ and then use $\mathcal{E}$ to estimate each of the low-degree coefficients; ie., with probability at least $1-\delta$ we have that $(\widetilde{f}(S) - \widehat{f}(S))^2 \le \varepsilon/n^k$ for every $|S| \le k$ where

$$\widetilde{f}(S) := \underset{(x,f(x))\in\mathcal{E}}{avg}\left[f(x)\chi_S(x)\right].$$

(b) Show that when using the above form of the Low-Degree Algorithm, the final hypothesis $h : \{-1,1\}^n \to \{-1,1\}$ is of the form

$$h(y) = \mathrm{sgn}\left(\sum_{(x,f(x))\in\mathcal{E}} w(\Delta(y,x)) \cdot f(x)\right),$$

for some function $w : \{0, 1, \ldots, n\} \to \mathbb{R}$. In other words, the hypothesis on a given y is equal to a weighted vote over all examples seen, where an example's weight depends only on its Hamming distance to y. Simplify your expression for w as much as you can.