Analysis of Boolean Functions **(CMU 18-859S, Spring 2007)**

Lecture 23: Majority Is Stablest cont., p-biased Fourier analysis

March 10, 2007

Lecturer: Ryan O'Donnell *Scribe: Ryan Williams*

1 Applications of Majority Is Stablest

We recall the statement of the Majority Is Stablest Theorem, and a corollary of it.

Theorem 1.1 (Majority Is Stablest) *Fix $0 < \rho < 1$. Let $f : \{-1,1\}^n \to [-1,1]$ have $E[f] = 0$ and be $(\epsilon, \frac{1}{\log(1/\epsilon)})$-quasirandom. Then*

$$\mathbb{S}_\rho(f) \leq \frac{2}{\pi}\arcsin\rho + O\left(\frac{\log\log 1/\epsilon}{\log 1/\epsilon}\right).$$

That is, $\mathbb{S}_\rho(f) \leq \mathbb{S}_\rho(\mathsf{Maj}_n) + O(1)$ as $n \to \infty$.

For a different range of ρ, a reverse inequality holds.

Corollary 1.2 ("Reverse" Majority Is Stablest) *Fix $-1 < \rho < 0$. Let $f : \{-1,1\}^n \to [-1,1]$ be $(\epsilon, \frac{1}{\log(1/\epsilon)})$-quasirandom. Then*

$$\mathbb{S}_\rho(f) \geq \frac{2}{\pi}\arcsin\rho - O\left(\frac{\log\log 1/\epsilon}{\log 1/\epsilon}\right).$$

That is, $\mathbb{S}_\rho(f) \geq \mathbb{S}_\rho(\mathsf{Maj}_n) - O(1)$ as $n \to \infty$.

Why can we omit the condition that $E[f] = 0$ in the corollary? A partial answer is that in the theorem, we had to rule out the constant functions, since when f is constant,

$$\mathbb{S}_\rho(f) = \sum_S \rho^{|S|}\hat{f}(S)^2 = \hat{f}(\varnothing)^2 = 1 > \frac{2}{\pi}\arcsin\rho + O\left(\frac{\log\log 1/\epsilon}{\log 1/\epsilon}\right)$$

for all $\rho < 1$. As the corollary proves a *lower bound* on stability, the constant functions are no longer a problem.

1.1 Condorcet's Paradox

Recall our discussion of Condorcet's paradox in Lecture 4. We looked at three-party elections that are determined by a social choice function $f : \{-1,1\}^n \to \{-1,1\}$ applied to each pair of candidates. We noted that when f is majority, an irrational outcome was possible, in that candidates A, B, and C could be ranked $A > B > C > A$ in a cycle. Further, recall we defined $\mathrm{Rationality}(f)$ to be the probability that no cycles are produced with f, when the "voters" decide between pairs of candidates uniformly at random (i.e. to decide between each pair of candidates, we draw random $x \in \{-1,1\}^n$ and evaluate $f(x)$). It was shown that

$$\mathrm{Rationality}(f) = \frac{3}{4} - \frac{3}{4}\mathbb{S}_{-1/3}(f).$$

Therefore, another corollary of the Majority Is Stablest theorem applies to voting paradoxes:

Corollary 1.3 *There is a function α such that for all $\epsilon > 0$, if a Boolean f is such that $\mathrm{Inf}_i(f) \leq \epsilon$, then*

$$\mathrm{Rationality}(f) \leq \frac{3}{4} - \frac{3}{4} \cdot \frac{2}{\pi} \arcsin\left(-\frac{1}{3}\right) + \alpha(\epsilon) \approx \mathrm{Rationality}(\mathsf{Maj}_n),$$

where $\alpha(\epsilon) \to 0$ as $\epsilon \to 0$.

Therefore, functions with small variable influences cannot be significantly "more rational" than Majority. The corollary follows from the definition of quasirandomness and the fact that attenuated variable influences are smaller than non-attenuated influences.

1.2 Testing For Stability and Hardness of Approximating Max Cut

The very definition of noise stability naturally suggests a function test. Namely, the

Noise Stability Test: For a Boolean function f with parameter $-1 < \rho < 0$:

1. Pick $x \in \{-1,1\}^n$ uniformly at random.
2. Pick $y \sim_\rho x$, i.e. y is a ρ-correlated copy of x.
3. Query $f(x)$ and $f(y)$.
4. ACCEPT if and only if $f(x) \neq f(y)$.

Observation 1.4 *The probability that a function f passes the above test is $\frac{1}{2} - \frac{1}{2}\mathbb{S}_\rho(f)$.*

Proof: Recall that $\mathbb{S}_\rho(f) = E_{x,y\sim_\rho x}[f(x)f(y)]$, so $\mathbf{Pr}_{x,y\sim_\rho y}[f(x) \neq f(y)] = \frac{(1-\mathbb{S}_\rho(f))}{2}$.□

Using the observation and the Reverse Majority Is Stablest Theorem, we deduce that

- $\mathbf{Pr}[\text{Dictator passes}] = \frac{1}{2} - \frac{1}{2}\rho$,
- $\mathbf{Pr}[\text{Constant function passes}] = 0$,
- If f is (ϵ, δ)-quasirandom, then

$$\begin{aligned}\mathbf{Pr}[f \text{ passes}] &\leq \frac{1}{2} - \frac{1}{2}\left(\frac{2}{\pi}\arcsin\rho\right) + \alpha(\delta, \epsilon), \\ &= \frac{1}{\pi}\arccos\rho + \alpha(\delta, \epsilon),\end{aligned}$$

  where $\alpha(\delta, \epsilon) \to 0$ as $\delta, \epsilon \to 0$.

Picking a ρ that maximizes the difference between the dictator and quasirandom probabilities, we find that when $\rho_0 \approx -0.69$,

- a dictator passes with probability exactly 0.845, and
- a quasirandom function (with negligible ϵ and δ) passes with probability at most 0.7424.

So the Noise Stability Test can be used to distinguish between dictators and quasirandom functions, a problem that came up in Lecture 6 concerning the unique games conjecture. In that lecture, we used the Hast-Odd Test to prove a hardness of approximation result for 3-Lin. Here, we'll use the Noise Stability Test to sketch a proof of a hardness of approximation result for the Max Cut problem.

Recall that every function test has associated with it a constraint satisfaction problem, where each constraint represents one of the possible choices of queries, and the constraint is satisfied if and only if the test accepts on those queries. If we translate the Noise Stability Test over to constraint satisfaction, observe that the resulting constraints are all of the form "$x \neq y$", for Boolean variables x, y. As usual, we consider the decision problem of finding a setting to the variables that maximizes the number of constraints satisfied, given a set of inequality constraints.

First, notice that the problem of inequality constraints is analogous to the well-known Max Cut problem: construe the variables as nodes, and each inequality constraint as an edge between two nodes.

Using the framework of Lecture 6, one can prove the following.

Theorem 1.5 *Assume the Unique Games Conjecture. Then $\forall\eta > 0$ and $-1 < \rho < 0$, it is* NP*-hard given a graph G to distinguish between the following two cases:*

- $\text{maxcut}(G) \geq \frac{1}{2} - \frac{1}{2}\rho$, *and*
- $\text{maxcut}(G) < \frac{1}{\pi}\arccos\rho + \eta$.

In particular, for $\rho_0 = 0.69$, it's hard to approximate Max Cut to within the ratio $\frac{\frac{1}{\pi}\arccos\rho_0 + \eta}{1/2 - 1/2\rho_0} > \frac{0.742}{0.845} \approx 0.878$.

What's even more interesting is that there *is* a polynomial time algorithm for Max Cut that approximates within a ratio that's arbitrarily close to the above hardness result.

Theorem 1.6 (Goemans-Williamson'95) *For all $\eta > 0$ and $-1 < \rho < 0$, there is a polynomial time algorithm which given G with $\mathrm{maxcut}(G) \geq \frac{1}{2} - \frac{1}{2}\rho$, finds a cut of fractional size $\geq \frac{1}{\pi}\arccos\rho - \eta$. Furthermore, the algorithm is a 0.878-approximation for Max Cut.*

Thus when the max cut has at least a $(1-\epsilon)$ fraction of edges crossing it, the Goemans-Williamson algorithm finds a cut that has at least a $1 - O(\sqrt{\epsilon})$ fraction of edges crossing it.

2 Introduction to p-Biased Fourier Analysis

Much of the Fourier analysis we have covered in this course can be generalized to functions $f : X^n \to \mathbb{R}$, where X^n is just assumed to be a product probability space. (Confer with question 1 on Homeworks 4 and 5.) For example, we could have

- $X = \{1, 2, 3\}$ where each is chosen with probability $1/3$.
- $X = \{T, F\}_p$, where $\mathbf{Pr}[X = T] = p$ and $\mathbf{Pr}[X = F] = 1 - p$. This is called the p-biased distribution.
- $X = (\mathbb{R}, \mathrm{Gaussian})$.

Of course, we have been studying $\{T, F\}_{1/2}$ all along. The main result we proved that was actually specific to $\{T, F\}_{1/2}$ was the hypercontractivity theorem (Lecture 16). Over the next few lectures, we'll look at Fourier analysis of Boolean functions on p-biased distributions.

3 Basic Definitions

Let's start laying the groundwork with definitions. Let $p \in (0, 1)$, write $q = 1 - p$, denote $\{T, F\}_p$ for the product distribution $\mathbf{Pr}[T] = p$, $\mathbf{Pr}[F] = q$. We begin with a re-definition of a character.

Definition 3.1 *The* p-biased character $\phi : \{T, F\}_p \to \mathbb{R}$ *is given by* $\phi(T) := -\sqrt{\frac{q}{p}}$, $\phi(F) := \sqrt{\frac{p}{q}}$.

Observe that in the $p = 1/2$ case, we have $\phi(T) = -1$ and $\phi(F) = 1$, as expected. Each character has

Observation 3.2 *For $x \sim \{T, F\}_p$,*

- $E[\phi(x)] = p \cdot \left(-\sqrt{\frac{q}{p}}\right) + q\left(\sqrt{\frac{p}{q}}\right) = -\sqrt{pq} + \sqrt{pq} = 0.$
- $E[\phi^2(x)] = p\left(\frac{q}{p}\right) + q\left(\frac{p}{q}\right) = p + q = 1.$

Definition 3.3 *For $S \subseteq [n]$, define $\phi_S : \{T, F\}_p^n \to \mathbb{R}$ by*

$$\phi_S(x) = \prod_{i \in S} \phi(x_i),$$

and define $\phi_\varnothing \equiv 1$.

Notation: We write $E_p[X(x)]$ to mean $E_{x \sim \{T,F\}^n}[X(x)]$, for a random variable X.

Definition 3.4 *For $f, g : \{T, F\}_p^n \to \mathbb{R}$, define the p*-biased inner product *as*

$$\langle f, g \rangle_p = E_p[f(x)g(x)].$$

The characters behave under this inner product as one would expect:

Proposition 3.5

$$\langle \phi_S, \phi_T \rangle = \begin{cases} 1 & \textit{if } S = T, \\ 0 & \textit{otherwise.} \end{cases}$$

Proof: If $S = T$, then

$$\begin{aligned} \langle \phi_S, \phi_T \rangle &= E_p[\prod_{i \in S} \phi(x_i)^2] \\ &= \prod_{i \in S} E_p[\phi(x)^2] \text{ by independence of } x_i\text{'s} \\ &= \prod_{i \in S} 1 = 1, \text{ by the observation.} \end{aligned}$$

If $S \neq T$, then

$$\langle \phi_S, \phi_T \rangle = E_p[\prod_{i \in S} \phi(x_i) \prod_{i \in T} \phi(x_i)].$$

Let $j \in S \Delta T$. Since $\phi(x_j)$ appears only once among the terms in $E_p[\cdots]$, this term is independent of the other terms being multiplied. Therefore

$$E_p[\prod_{i \in S} \phi(x_i) \prod_{i \in T} \phi(x_i)] = E[\phi(x_j)] \cdot E_p[\cdots] = 0,$$

since $E[\phi(x_j)] = 0$. □

At this stage, we must give a note of WARNING to the reader: there are $p \in (0, 1)$ for which their p-biased characters have $\phi_i^2(x) \neq 1$, and $\phi_S(x) \cdot \phi_T(x) \neq \phi_{S\Delta T}(x)$. That is, certain inequalities in the original case do NOT hold in general!

Corollary 3.6 *$\{\phi_S\}_{S\subseteq[n]}$ forms an orthonormal basis for the 2^n dimensional inner product space of functions $\{T, F\}_p^n \to \mathbb{R}$, with respect to $\langle\cdot, \cdot\rangle_p$.*

Having established the character functions on sets, there is a natural definition of Fourier coefficients in the p-biased setting.

Definition 3.7 *For a fixed $p \in (0, 1)$, $f : \{T, F\}_p^n \to \mathbb{R}$, and $S \subseteq [n]$, we write $\widetilde{f}(S) = \langle f, \phi_S\rangle_p$.*

Finally, we have the p-biased Fourier expansion of a function f:

Proposition 3.8 *$f = \sum_S \widetilde{f}(S)\phi_S$, as functions $\{T, F\}_p^n \to \mathbb{R}$.*

3.1 Comparison With Our Former Setting

Many propositions that held under our old Fourier analysis still hold in the p-biased world. For example:

Proposition 3.9 (Parseval's Identity) $E_p[f(x)^2] = \sum_S \widetilde{f}(S)^2$.

Proposition 3.10 (Plancherel's Identity) $\langle f, g\rangle_p = \sum_S \widetilde{f}(S) \cdot \widetilde{g}(S)$.

We leave the proofs as an exercise. Essentially, these results still hold because they only relied on the orthonormality of the characters. The expectation and variance have analogous formulas for functions $f : \{T, F\}_p^n \to \mathbb{R}$; these formulas follow readily from the definitions.

Definition 3.11 $\mathbf{Var}_p(f) = E_p[f^2] - E_p[f]^2$.

Proposition 3.12 *$E_p[f] = \widetilde{f}(\varnothing)$ and $\mathbf{Var}_p(f) = \sum_{S\neq\varnothing} \widetilde{f}(S)^2$.*

While much stays the same, a few of the notions from our old setting require some modification. In particular, the definition of the differential operator changes a bit.

Definition 3.13 *For $i \in [n]$, define the operator D_i on functions $\{T, F\}_p^n \to \mathbb{R}$ by*

$$(D_i f)(x) = \sqrt{pq} \cdot \left[f(x^{(i=F)}) - f(x^{(i=T)})\right].$$

The motivation for our definition of D_i is to preserve the following property, that now holds for all p-biased Fourier transforms.

Proposition 3.14 $D_i f = \sum_{S \ni i} \phi_{S \setminus \{i\}}$.

Proof: By definition, D_i is a linear operator ($D_i(f + cg) = D_i f + cD_i g$). Since every function f can be written in the form $f = \sum_S \widetilde{f}(S)\phi_S$, by linearity of D_i it suffices for us to check that

$$D_i \phi_S = \begin{cases} \phi_{S\setminus\{i\}} & \text{if } x \in S \\ 0 & \text{otherwise,} \end{cases}$$

for all $S \subseteq [n]$.

- If $i \notin S$, then ϕ_S doesn't depend on the ith coordinate at all, so $\phi_S(x^{(i=F)}) = \phi_S(x^{(i=T)})$ and $(D_i f)(x) = 0$.
- If $i \in S$, then

$$\begin{aligned} D_i \phi_S &= \sqrt{pq}\left[\phi_S(x^{(i=F)}) - \phi_S(x^{(i=T)})\right] \\ &= \sqrt{pq}\left[\phi_{S\setminus\{i\}} \cdot \phi(F) - \phi_{S\setminus\{i\}} \cdot \phi(T)\right] \quad \text{by definition} \\ &= \phi_{S\setminus\{i\}}\sqrt{pq}\left[\sqrt{\frac{p}{q}} - \left(-\sqrt{\frac{q}{p}}\right)\right] \\ &= \phi_{S\setminus\{i\}}\sqrt{pq}\left[\frac{p+q}{\sqrt{pq}}\right] = \phi_{S\setminus\{i\}}. \end{aligned}$$

□

Corollary 3.15 $E_p[(D_i f)^2] = \sum_{S \ni i} \widetilde{f}(S)^2$.

The proof of the corollary is similar to that in the original setting. Naturally, as another analogue we make the definition

$$\mathrm{Inf}_i(f) := E_p[(D_i f)^2].$$

Exercise: If $f : \{T, F\}_p^n \to \{-1, 1\}$ then $\mathrm{Inf}_i(f) = 4pq\mathbf{Pr}_p[f(x) \neq f(x^{(i)})] = 2\mathbf{Pr}_p[f(x) \neq f(x^{R(i)})]$, recalling that $x^{(i)}$ is our notation for x with the ith coordinate flipped, and $x^{R(i)}$ is notation for x with the ith coordinate assigned T with probability p, F with probability q.

Proposition 3.16 *If $f : \{T, F\}_p^n \to \{-1, 1\}$ and is monotone (flipping F to T can only flip f's value from 1 to -1), then* $\mathrm{Inf}_i(f) = 2\sqrt{pq}\widehat{f}(i)$.

Proof:

$$\begin{aligned}\widehat{f}(i) &= E_p[D_i f] \\ &= E_p[\sqrt{pq}\left[f(x^{(i=F)}) - f(x^{(i=T)})\right].\end{aligned}$$

Since f is monotone, the quantity $\left[f(x^{(i=F)}) - f(x^{(i=T)})\right]$ is either 2 or 0 on all x. (In particular, it can't be -2.) Therefore the above is equal to

$$= \frac{E_p[(D_i f)^2]}{2\sqrt{pq}} = \frac{\mathrm{Inf}_i(f)}{2\sqrt{pq}}.$$

□