MASSACHUSETTS INSTITUTE OF TECHNOLOGY

6.265/15.070J Lecture 23 May 10, SP 2017

Lecturer: Yury Pulyanksiy Scribe notes by Jake Wellens

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They are posted to serve class purposes.*

Topics for upcoming three lectures on boolean analysis:

- Influences and isoperimetry (this lecture)
- Influences and Fourier analysis
- Influences and Noise Stability

The ultimate goal of this lecture is to prove the following theorem of Margulis:

Goal Theorem: (Margulis, '74): *Let G be a graph with edge connectivity $\lambda(G) \geq t$ (i.e. at least t edges must be deleted from G to disconnect it) and set*

$$\Psi(p) = \Pr_{X \sim (\{0,1\}^E, \pi_p^{\otimes E})}[G' = (V, E \setminus X) \text{ is disconnected}]$$

For $\epsilon \in (0, 1/2)$, let $p_\epsilon = \Psi^{-1}(\epsilon)$, $p_{1-\epsilon} = \Psi^{-1}(1-\epsilon)$. Then

$$p_{1-\epsilon} - p_\epsilon \leq \frac{c_\epsilon}{\sqrt{t}}$$

for a constant c_ϵ independent of G!

1 Preliminary definitions

- All functions considered today will be $\{0,1\}^n \to \mathbb{R}$. A **boolean function** has range $\{0,1\}$.
- The **influence** of coordinate i at level p on a boolean function f is

$$I_i^p(f) := \Pr_{x \sim \pi_p^{\otimes n}}[f(x) \neq f(x \oplus e_j)]$$

The **total influence** is the sum of all the coordinate influences, i.e.

$$I^p(f) := \sum_{i=1}^{n} I_i^p(f)$$

- We can define **discrete derivatives** for functions $f : \{0,1\}^n \to \mathbb{R}$:

$$\nabla_i f(x) := f(x) - f(x \oplus e_i)$$
$$\nabla_i^+ f(x) := (f(x) - f(x \oplus e_i))_+$$
$$|\nabla f(x)|^2 := \sum_{i=1}^n |\nabla_i f(x)|^2$$
$$|\nabla^+ f(x)|^2 := \sum_{i=1}^n |\nabla_i^+ f(x)|^2$$

2 An analytic approach to edge/vertex isoperimetry

Recall from Lecture 9 that for sets $A \subseteq \{0,1\}^n$ with $|A| \geq \delta 2^n$, the *internal boundary*

$$|\partial A| := \{x \in A : d_H(x, A^c) = 1\} = A \cap \Gamma^1(A^c)$$

is at least as large as for Hamming balls, i.e.

$$|\partial A| \geq \frac{c_\delta}{\sqrt{n}} 2^n \tag{1}$$

Indeed, for the Hamming ball $B = B(0, n/2 - c\sqrt{n})$ with $\delta = \Phi(-c) := \mathrm{NormalCDF}(-c)$, we have $|B| = \delta 2^n$ and

$$|\partial B| = 2^n \Pr[|x| = n/2 - c\sqrt{n}] \approx 2^n(\Phi(-c) - \Phi(-c - \frac{1}{\sqrt{n}})) \approx \frac{c_\delta}{\sqrt{n}} 2^n,$$

while for the subcube $C = \{x : x_1 = 0\}$, we have $|C| = \frac{1}{2}2^n$ and $|\partial C| = |C|$. However, when looking at the *edge boundary*

$$|E(A, A^c)| := \sum_{x \in A} |N(x) \cap A^c|$$

we see a different picture. Here, subcubes actually do the best: $|A| \geq 2^k \implies |E(A, A^c)| \geq (n-k)2^k$, and more generally, $|A| \geq \delta 2^n \implies |E(A, A^c)| \gtrsim \delta \log(1/\delta) 2^n$. Our goal in this section will be to prove (1) analytically (i.e. without horrible combinatorics), as well as characterize the *tension* between the internal and edge boundaries of a set. More specifically, we'll prove (a generalization of)

Theorem: *For $A \in \{0,1\}^n$, we have*

$$|\partial A|/2^n \geq \frac{1}{\sqrt{n}}(|A|/2^n)(|A^c|/2^n)$$

and if $|A| \geq \delta 2^n$, then

$$\frac{1}{2^n}|E(A, A^c)| \cdot \frac{1}{2^n}|\partial A| \geq \delta^2(1-\delta)^2$$

These results will follow from *Bobkov's Inequality:*

Theorem (Bobkov): *Let X_i be independent 0-1 valued random variables, and set $\phi(t) := t(1-t)$. Then for any $f : \{0,1\}^n \to [0,1]$, we have*

$$\phi(\mathbb{E}[f(X)]) \le \mathbb{E}[\sqrt{\phi^2(f(X)) + |\nabla^+ f(X)|^2}].$$

Proof: In the $n=1$ case, this amounts to checking that for any $a,b,p \in [0,1]$, we have

$$\phi((1-p)a + pb) \le (1-p)\sqrt{\phi^2(a) + (a-b)_+^2} + p\sqrt{\phi^2(b) + (b-a)_+^2}$$

Expanding this for any $0 \le a \le b \le 1$ (wlog) and subtracting the left from the right, we obtain a quadratic polynomial in p with leading coefficient a^2, one root at $p = 0$ and the other at

$$\frac{-\sqrt{a^2 - 2ab + b^4 - 2b^3 + 2b^2} + a^2 - 2ab + b}{a^2 - 2ab + b^2}$$

and hence it suffices to show this root is non-positive. After some mild algebraic manipulation, this is equivalent to proving

$$(b-1)^2 + 2ab - a^2 \ge 0$$

which is easy enough to verify. To extend to higher dimensions, we'll use the following generally useful tensorization lemma:

Tensorization lemma: *Fix $\alpha : [0,1] \to \mathbb{R}^+$, and suppose that for all $i \in [n]$ and all $f : \{0,1\}^n \to [0,1]$, we have $\alpha(\mathbb{E}_{x_i} f) \le \mathbb{E}_{x_i}\sqrt{\alpha^2(f) + |\nabla_i^+ f|^2}$. Then*

$$\alpha(\mathbb{E}_{x_1,\dots,x_n} f) \le \mathbb{E}_{x_1,\dots,x_n}\sqrt{\alpha^2(f) + |\nabla^+ f|^2}$$

Proof: We proceed by induction on n – suppose for any $g : \{0,1\}^{n-1} \to [0,1]$ that

$$\alpha(\mathbb{E}_{[n-1]} g) \le \mathbb{E}_{[n-1]}\sqrt{\alpha^2(g) + |\nabla^+ g|^2}$$

where $\mathbb{E}_{[n-1]}$ is shorthand for the operator $\mathbb{E}_{x_1,\dots,x_{n-1}}$. Then since for $f : \{0,1\}^n \to [0,1]$, $\mathbb{E}_{x_n} f$ maps $\{0,1\}^{n-1} \to [0,1]$, we have

$$\alpha(\mathbb{E} f) = \alpha(\mathbb{E}_{[n-1]}(\mathbb{E}_{x_n} f)) \le \mathbb{E}_{[n-1]}\sqrt{\alpha^2(\mathbb{E}_{x_n} f) + \sum_{i=1}^{n-1} |\nabla_i^+ \mathbb{E}_{x_n} f|^2}$$

First observe by convexity that $\nabla_i^+(\mathbb{E}_{x_n} f) \le \mathbb{E}_{x_n}(\nabla_i^+ f)$ and so

$$\alpha(\mathbb{E} f) \le \mathbb{E}_{[n-1]}\sqrt{\alpha^2(\mathbb{E}_{x_n} f) + \sum_{i=1}^{n-1} |\mathbb{E}_{x_n} \nabla_i^+ f|^2}$$

Now using our original assumption for $i = n$, this is at most

$$\mathbb{E}_{[n-1]}\sqrt{\left(\mathbb{E}_{x_n}\sqrt{\alpha^2(f) + |\nabla_n^+ f|^2}\right)^2 + \sum_{i=1}^{n-1} |\mathbb{E}_{x_n} \nabla_i^+ f|^2}$$

Ignoring the $\mathbb{E}_{[n-1]}$ on the outside, this looks like an L^2 norm (in the n-point counting measure) of an L^1 norm (in dx_n) – by Minkowski's inequality[1], we can switch the order of the norms and obtain the upper bound

$$\mathbb{E}_{[n-1]}\mathbb{E}_{x_n}\sqrt{\alpha^2(f)+|\nabla^+ f|^2}$$

as desired. □

By taking f to be the indicator of a set A, we obtain

Corollary: *For $A \subseteq \{0,1\}$, X_i independent Bernoullis,*

$$\mathbb{E}|\nabla^+ 1_A| \geq \Pr[X \in A] \cdot \Pr[X \notin A]$$

In other words, if we set

$$h_A(x) := |\nabla^+ 1_A(x)|^2 = \begin{cases} \#\{y \in A^c : d(x,y) = 1\} & \text{if } x \in \partial A \\ 0 & \text{if } x \notin \partial A \end{cases}$$

then $\mathbb{E}[\sqrt{h_A}] \geq \mathrm{Var}(1_A)$. Note that $\mathbb{E}[h_A] = \mu(\partial A)$, and since $h_A(x) \leq n \cdot 1_{\partial A}$, we have

$$\mu(\partial A) \geq \frac{1}{\sqrt{n}}\mu(A)\mu(A^c)$$

Also since $\frac{1}{2^n}E(A, A^c) = \frac{1}{2}\mathbb{E}|\nabla f|^2$, Cauchy-Schwarz gives us

$$\mathrm{Var}(1_A) \leq \mathbb{E}\sqrt{h_A} \leq \sqrt{(\mathbb{E}h_A)\mu(\partial A)},$$

that is,

$$\delta^2(1-\delta)^2 \leq \frac{1}{2^n}|E(A,A^c)| \cdot \frac{1}{2^n}|\partial A|$$

No combinatorics necessary!

3 Total influence and threshold width

Proposition: *For $f = 1_A$ boolean, we have*

$$I_i^p(f) = \mathbb{E}|\nabla_i f|^2$$

and so in particular, if $p = 1/2$, we have $I^{1/2}(f) = \frac{2}{2^n}|E(A,A^c)|$. Moreover, if f (i.e. A) is also monotone, then

$$I^p(f) = \begin{cases} \frac{1}{p}\mathbb{E}_{x\sim\pi_p^{\otimes n}} h_A(x), & \text{if } f \text{ increasing} \\ \frac{1}{1-p}\mathbb{E}_{x\sim\pi_p^{\otimes n}} h_A(x), & \text{if } f \text{ decreasing} \end{cases}$$

[1] Minkowski's inequality says that for $q \geq p \geq 1$ and σ-finite measure spaces X, Y, we have $\| \|F(x,y)\|_{p,Y}\|_{q,X} \leq \| \|F(x,y)\|_{q,X}\|_{p,Y}$.

For non-monotone f, we still have

$$\frac{1}{\max(p,1-p)}\mathbb{E}h_A \leq I^p(f) \leq \frac{1}{\min(p,1-p)}\mathbb{E}h_A,$$

with either equality iff f is monotone[2] *in the corresponding direction.*

Proof: The first part is obvious from the definitions. For the second, suppose f is increasing. Then each edge between A and A^c – say $(x, x^{\oplus i})$ – contributes $\Pr(x_{-i})$ to $I^p(f)$ and but only $p\Pr(x_{-i})$ to $\mathbb{E}h_A$. If f is decreasing, the contribution is $(1-p)\Pr(x_{-i})$, which settles the monotone case. Finally, for general boolean f, each edge contributes either $p\Pr(x_{-i})$ or $(1-p)\Pr(x_i)$ to $\mathbb{E}h_A$, depending on whether f is "going up" or "going down" at that point, and so we have the above inequalities, which are tight iff this coefficient is the same for all such edges. □

As we've seen in the case of Erdos-Renyi random graphs, it is natural to consider how $\Pr_{x\sim\pi_p^{\otimes n}}[f(x) = 1]$ varies with p. For monotone f, it turns out that the rate of change of this probability is exactly captured by the total influence:

Lemma (Margulis-Russo): *For any increasing set $A \subseteq \{0,1\}^n$,*

$$\frac{d}{dp}\mu_p(A) = I^p(1_A)$$

(and hence by taking complements, the same formula holds for decreasing sets with a minus sign.)

Proof: Set $g(p_1,\ldots,p_n) := \Pr[(X_1,\ldots,X_n) \in A]$, where X_i are independent Bernoulli(p_i) variables, so that $\mu_p(A) = g(p,\ldots,p)$. By the chain rule, $\frac{d}{dp}\mu_p(A) = \sum_{i=1}^n \frac{\partial}{\partial p_i} g(p_1,\ldots,p_n)|_{(p,\ldots,p)}$. We can compute each partial derivative explicitly:

$$\lim_{\epsilon\to 0}\frac{g(p+\epsilon,p,\ldots,p) - g(p,\ldots,p)}{\epsilon} = \frac{1}{\epsilon}\big(\Pr[(X_1',\ldots,X_n) \in A] - \Pr[(X_1,\ldots,X_n) \in A]\big)$$

where $X_1' \sim \text{Ber}(p+\epsilon)$ and $X_1 \sim \text{Ber}(p)$. We can couple X_1, X_1' as follows:

$$\begin{cases} \Pr((X_1,X_1') = (0,0)) = 1-p-\epsilon \\ \Pr((X_1,X_1') = (1,0)) = 0 \\ \Pr((X_1,X_1') = (0,1)) = \epsilon \\ \Pr((X_1,X_1') = (1,1)) = p \end{cases}$$

and so the difference quotient becomes

$$\frac{1}{\epsilon}\underbrace{\Pr[X_1' = 1, X_1 = 0]}_{\epsilon}\underbrace{\Pr[1 \text{ is pivotal for } X_2,\ldots,X_n]}_{I_1^p(f)}$$

which proves the lemma after summing over i. □

We are now ready to prove our main theorem, restated below:

[2]unless $p = 1/2$, in which case both equalities always hold.

Theorem: (Margulis, '74): *Let G be a graph with edge connectivity $\lambda(G) \geq t$ (i.e. at least t edges must be deleted from G to disconnect it) and set*

$$\Psi(p) = \Pr_{X \sim (\{0,1\}^E, \pi_p^{\otimes E})} [G' = (V, E \setminus X) \text{ is disconnected}]$$

For $\epsilon \in (0, 1/2)$, let $p_\epsilon = \Psi^{-1}(\epsilon)$, $p_{1-\epsilon} = \Psi^{-1}(1-\epsilon)$. Then

$$p_{1-\epsilon} - p_\epsilon \leq \frac{c_\epsilon}{\sqrt{t}}.$$

Proof: Let $\Omega = \{x : G' = (V, E(G) \backslash x) \text{ is disconnected}\}$, which is clearly a monotone increasing set. Then the Margulis-Russo lemma and the previous proposition imply

$$\Psi'(p) = I^p(1_\Omega) = \frac{1}{p}\mathbb{E}_p h_\Omega$$

A direct combinatorial argument[3] shows that $h_\Omega(x) \geq t$ for all $x \in \partial\Omega$, and hence

$$\mathrm{Var}_p(1_\Omega) \leq \mathbb{E}_p \sqrt{h_\Omega} = \sqrt{t} \cdot \mathbb{E}_p \underbrace{\sqrt{\frac{h_\Omega}{t}}}_{\geq 1} \leq \sqrt{t} \cdot \mathbb{E}_p \frac{h_\Omega}{t} = \frac{p}{\sqrt{t}} I^p(1_\Omega)$$

Thus,

$$\Psi'(p) = I^p(1_\Omega) \geq \sqrt{t}\mathrm{Var}_p(1_\Omega) \geq \sqrt{t}\epsilon(1-\epsilon) \text{ for } p \in [p_\epsilon, p_{1-\epsilon}]$$

and since $\int_{p_\epsilon}^{p_{1-\epsilon}} \Psi'(p)\, dp = 1 - 2\epsilon$, we conclude

$$p_{1-\epsilon} - p_\epsilon \leq \frac{1-2\epsilon}{\epsilon(1-\epsilon)\sqrt{t}}$$

as desired. □

[3]Let $x \in \partial\Omega$. Then removing the edges in x from G creates a graph G' with two connected components G_1' and G_2' (since there is a way to add back a single edge and connect the whole graph.) The number $h_\Omega(x)$ counts exactly $E(G_1', G_2')$ in G, since adding any one of these edges back in to G' yields a connected Hamming neighbor of x and vice versa. Note that starting from G and removing all of these edges yields a disconnected graph, so there must be at least t of them.