Spectral Graph Theory Lecture 20

Concentration of Measure from Eigenvalue Bounds

Daniel A. Spielman November 12, 2012

20.1 About these notes

These notes are not necessarily an accurate representation of what happened in class. The notes written before class say what I think I should say. The notes written after class way what I wish I said.

20.2 Overview

We are going to use spectral graph theory to establish concentration of measure theorems for the hypercube. I was very excited during lecture because I thought I was deriving the correct constant for the exponent. I now see that the constant I derive is off by a factor of approximately 3. Maybe one can improve the argument?

20.3 Concentration of Measure

Concentration of Measure refers to a techinque in probability for proving that certain random variables are unlikely to take values too much larger or smaller than their median. For example, consider the sum of d independent and uniformly chosen ± 1 variables, $x_1, \ldots, x_d$. The mean of the sum is obviously 0, and the symmetry of the distributions tells us that the median is 0 as well. The Chernoff bound tells us that for every $t > 0$,

$$\Pr\left[\sum_i x_i > t\right] \leq \exp\left(-t^2/2d\right).$$

The concentration of measure techniques apply to a much broader range of situations. To begin, they consider a function f on the probability space. In this case, we consider

$$f : \{\pm 1\}^d \to \mathbb{R}.$$

We require that the function f satisfy a *Lipschitz* condition. That is, there should be a constant L such that

$$|f(\boldsymbol{x}) - f(\boldsymbol{y})| \leq L \left\| \boldsymbol{x} - \boldsymbol{y} \right\|_1.$$

Such a function f is said to be L-Lipschitz. One could consider other norms on $\boldsymbol{x} - \boldsymbol{y}$, but this is the right norm for functions like the sum.

One can show that every Lipschitz function is concentrated. For example, if f is 1-Lipschitz and the median of f is μ, then for every t one can show [AM80, (1.3)]

$$\Pr_{\boldsymbol{x}}\left[f(\boldsymbol{x}) > \mu + t\right] \leq \exp\left(-t^2/2d\right). \tag{20.1}$$

That is, one can prove just as strong concentration bounds for every 1-Lipschitz funcation as one can prove for the summation. You can extend this bound to general L-Lipschitz functions by scaling.

In this lecture, we will prove a theorem that is almost this strong. The difference between our proof and the standard proofs is that ours will only exploit the eigenvalues of the Laplacian of the hypercube. I teach this for two reasons:

1. The technique we will use could possibly be applied to many other interesting probability spaces, and
2. I don't think that anyone has previously realized that tight analyses of the Conjugate Gradient yield such clean theorems in probability.

20.4 The Hypercube

For the rest of this lecture, we will consider the uniform probability on $\{0,1\}^d$, and we will associate $\{0,1\}^d$ with the vertices of the d-dimensional hypercube. The Lipschitz condition tells us that if $(\boldsymbol{x}, \boldsymbol{y})$ is an edge of the hypercube, then

$$|f(\boldsymbol{x}) - f(\boldsymbol{y})| \leq L.$$

For the rest of the lecture, we will assume $L = 1$. The translation from $\{\pm 1\}^d$ to $\{0,1\}^d$ changes the bound for a 1-Lipschitz function from that in (20.1) to

$$\Pr_{\boldsymbol{x}}\left[f(\boldsymbol{x}) > \mu + t\right] \leq \exp\left(-2t^2/d\right). \tag{20.2}$$

Let S be the set of vertices of the hypercube on which f is at most μ (which I defined to be the median). Then, S contains at least half the vertices. We will define $n = 2^d$, so $|S| \geq n/2$. Let T be the set of vertices on which f exceeds $\mu + t$. The Lipschitz condition then implies that for every $\boldsymbol{x} \in S$ and $\boldsymbol{y} \in T$,

$$d(\boldsymbol{x}, \boldsymbol{y}) \geq t.$$

We will use the fact to prove that T must be small if t is big. For our first proof, we will merely use the fact that the eigenvalues of the hypercube lie between 2 and $2d$.

20.5 Extending the Diameter Bound

We will begin by extending the diameter bound that we proved last lecture. I will state a slight generalization of the theorem we proved last class.

Theorem 20.5.1. *Let $G = (V, E)$ be a graph, and let S and T be disjoint subsets of the vertices. Let $|S| = \sigma n$ and $|T| = \tau n$. If there is a polynomial q of degree t such that $q(0) = 1$ and $q(\lambda_i) \leq \epsilon$ for all eigenvalues λ_i of $\boldsymbol{L}_G$, where*

$$\epsilon < \sqrt{\frac{\sigma\tau}{(1-\sigma)(1-\tau)}}, \tag{20.3}$$

then $d(S, T) \leq t$.

Proof. Let p be the polynomial such that

$$q(X) = 1 - Xp(X).$$

We then have

$$\|p(\boldsymbol{L})\boldsymbol{L} - \boldsymbol{\Pi}\| \leq \epsilon. \tag{20.4}$$

If we assume by way of contradiction that the distance between S and T is at least $t + 1$, then

$$\boldsymbol{\chi}_S^T (p(\boldsymbol{L})\boldsymbol{L})\boldsymbol{\chi}_T = 0.$$

On the other hand, as S and T are disjoint, we have

$$\boldsymbol{\chi}_S^T \boldsymbol{\Pi} \boldsymbol{\chi}_T = \boldsymbol{\chi}_S^T \frac{1}{n} \boldsymbol{L}_{K_n} \boldsymbol{\chi}_T = \frac{1}{n} |S| \, |T| \, .$$

Combining these facts with (20.4) gives

$$\frac{1}{n} |S| \, |T| = \boldsymbol{\chi}_S^T \left(p(\boldsymbol{L})\boldsymbol{L} - \boldsymbol{\Pi} \right) \boldsymbol{\chi}_T \leq \|\boldsymbol{\chi}_S\| \, \|(p(\boldsymbol{L})\boldsymbol{L} - \boldsymbol{\Pi})\| \, \|\boldsymbol{\chi}_T\| \leq \epsilon \, \|\boldsymbol{\chi}_S\| \, \|\boldsymbol{\chi}_T\| \, . \tag{20.5}$$

Before proceeding, I'd like to tighten this up slightly by observing that

$$\left(p(\boldsymbol{L})\boldsymbol{L} - \boldsymbol{\Pi} \right) = \boldsymbol{\Pi} \left(p(\boldsymbol{L})\boldsymbol{L} - \boldsymbol{\Pi} \right) \boldsymbol{\Pi},$$

so

$$\boldsymbol{\chi}_S^T \left(p(\boldsymbol{L})\boldsymbol{L} - \boldsymbol{\Pi} \right) \boldsymbol{\chi}_T \leq \epsilon \, \|\boldsymbol{\Pi}\boldsymbol{\chi}_S\| \, \|\boldsymbol{\Pi}\boldsymbol{\chi}_T\| = \epsilon n \sqrt{(\sigma - \sigma^2)(\tau - \tau^2)}$$

Combining these inequalities yields

$$\frac{1}{n}(\sigma n)(\tau n) \leq \epsilon n \sqrt{(\sigma - \sigma^2)(\tau - \tau^2)} \quad \Longleftrightarrow \quad \sqrt{\frac{\sigma\tau}{(1-\sigma)(1-\tau)}} \leq \epsilon.$$

As this inequality contradicts the assumption (20.3) , we may assume that the distance between S and T is at most t. ☐

Corollary 20.5.2. *If S is a subset of at least half the vertices, if every vertex in T has distance at least t from S, and if q is a polynomial of degree t satisfying $q(0) = 1$, then*

$$\tau \leq \max_{i \geq 2} q(\lambda_i)^2,$$

where $\lambda_2, \ldots, \lambda_n$ are the non-trivial eigenvalues of the Laplacian.

I remark that we can apply these statements to the Laplacian of the same graph with any edge weights. Some choices of edge weights might yield better bounds. In a later section, we will use one half the Laplacian. This doesn't change any bounds, but it simplifies the presentation to have eigenvalues between 1 and d.

For any $a > 0$ and $b > 1.5a$, we saw how to use Chebyshev polynomials to construct polynomials q of degree t such that $q(0) = 0$ and

$$|q(x)| \leq \frac{1}{2} \exp\left(-2t\sqrt{a/b}\right), \text{ for } a \leq x \leq b.$$

As the eigenvalues of $(1/2)\boldsymbol{L}_{H_d}$ lie between 1 and d, we obtain the bound

$$\tau \leq \frac{1}{4} \exp\left(-4t/\sqrt{d}\right).$$

This implies the following concentration bound for 1-Lipschitz functions on the $\{0,1\}^d$ hypercube:

$$\mathrm{Pr}_{\boldsymbol{x}}\left[f(\boldsymbol{x}) > \mu + t\right] = \mathrm{Pr}_{\boldsymbol{x}}\left[\boldsymbol{x} \in T\right] = \tau \leq \frac{1}{4} e^{-\frac{4t}{\sqrt{d}}}.$$

So, as soon as t exceeds $\sqrt{d}$, the probability that f exceeds the median by more than t becomes very small. This is a very reasonable concentration bound. I believe that the observation that such concentration bounds can be derived from the eigenvalues of the underlying graph is due to Alon and Milman [AM85a].

This bound is unfortunately far from the bound of (20.2). The difference is most pronounced when $t = \epsilon d$ for a constant ϵ: the tight bound decreases like the reciprocal of an exponential in d while the bound we just proved decreases with the reciprocal of an exponential in $\sqrt{d}$. In the next sections we will see how to prove a bound that is not too far from (20.2).

20.6 Improving the polynomials

I should begin by recalling how we used Chebyshev polynomials before. We defined the tth Chebyshev polynomial by

$$T_t(X) = \cos(t \,\mathrm{acos}(X)).$$

We observed that

$$|T_t(X)| \leq 1 \text{ for } |X| \leq 1, \text{ and}$$
$$T_t(1+\gamma) \geq \frac{1}{2}(1+\sqrt{2\gamma})^t, \text{ for } \gamma \geq 0.$$

We then defined a polynomial, which I will now denote $C_{a,b,t}$, by

$$C_{a,b,t}(X) \stackrel{\text{def}}{=} T_t\left(\frac{a+b-2X}{b-a}\right).$$

We observed that

$$|T_t(X)| \leq 1 \text{ for } a \leq X \leq b.$$

To put a lower bound on $T_t(0)$, note that

$$T_t(0) = T_t\left(\frac{b+a}{b-a}\right) = T_t\left(1 + \frac{2}{(b/a)-1}\right) \geq (1 + 2\sqrt{(1/(b/a)-1)})^t \geq \exp\left(2t\sqrt{a/b}\right),$$

where the last inequality relies on $b \geq 1.5a$. We then set

$$q(X) = C_{a,b,t}(X)/C_{a,b,t}(0).$$

This guarantees that $q(0) = 1$.

When applying these polynomials to $(1/2)\boldsymbol{L}_{H_d}$, we take $a = 1$ and $b = d$. But, if we are going to be applying polynomials of high degree then it seems we could get some improvement by handling small roots separately. For example, if we multiplied by $(X - 1)$ then 1 would be a root, and we could effectively take $a = 2$. We could push this further by considering a polynomial of the form

$$C_{a,d,t}(X) \prod_{i=1}^{a-1} (X - i)$$

for some a. This polynomial would be zero at the eigenvalues less than a. However, this constructions also has problems.

In Lecture 19, we eliminated large eigenvalues by forcing them to be roots of our polynomial in this fashion. This does not work as well here, because forcing eigenvalues to be roots increases the value of our polynomial at the larger eigenvalues. For example, consider $X = d$. We have

$$C_{a,d,t}(d) \prod_{i=1}^{a-1} (d - i) = \prod_{i=1}^{a-1} (d - i) = d!/a!,$$

which is pretty big.

We fix this problem by using a fancier polynomial than $(X - i)$. Following Jennings [Jen77] (and a similar analysis in [AL86]), we multiply by a shifted Chebyshev polynomial that has a root at i and that is at most 1 on the interval $[a, b]$. For $\lambda < a$, we define the polynomials that we will use by

$$g_\lambda(X) \stackrel{\text{def}}{=} T_r\left(\frac{(d - X)\cos(\pi/2r) + (\lambda - X)}{d - \lambda}\right). \tag{20.6}$$

We will set

$$r = \lceil d/a \rceil.$$

For X between λ and d, the ratio inside (20.6) is between -1 and $\cos(\pi/2r)$. So,

$$|g_\lambda(X)| \leq 1 \text{ for } \lambda \leq X \leq d.$$

We can also show that λ is a root:

$$g_\lambda(\lambda) = T_r\left(\frac{(d-\lambda)\cos(\pi/2r)}{d-\lambda}\right) = T_r\left(\cos(\pi/2r)\right) = \cos\left(r\,\mathrm{acos}\cos(\pi/2r)\right) = \cos(\pi/2) = 0.$$

So, we will use polynomials of the form

$$q(X) \stackrel{\text{def}}{=} \frac{C_{a,d,t}(X)\prod_{i=1}^{a-1} g_i(X)}{C_{a,d,t}(0)\prod_{i=1}^{a-1} g_i(0)}.$$

By definition, this polynomial satisfies $q(0) = 1$ and $q(i) = 0$ for every integer between 1 and $a-1$. To bound the values of the polynomial between a and d, we need a lower bound on $g_i(0)$. Here is a bound that we will prove in the next section.

Claim 20.6.1. *If $r \geq 6$ and $\lambda \leq a$, then*

$$g_\lambda(0) \geq (0.94)\lambda/a.$$

Given this claim, we can compute

$$\prod_{i=1}^{a-1} g_i(0) \geq \frac{(0.94)^a(a-1)!}{a^{a-1}} \geq \frac{(0.94)^a(a-1)^{a-1}}{(ea)^{a-1}} \geq \frac{(0.94)^a}{(e)^a} = \left(\frac{0.94}{e}\right)^a.$$

So,

$$C_{a,b,t}(0)\prod_{i=1}^{a-1} g_i(0) \geq \exp\left(2t\sqrt{a/b} - a\ln(e/0.94)\right) \geq \exp\left(2t\sqrt{a/b} - 1.06a\right).$$

We now choose $t = \left\lceil 2\sqrt{ab}\right\rceil$. So, the total degree of q is

$$\left\lceil 2\sqrt{ad}\right\rceil + (a-1)\left\lceil\sqrt{d/a}\right\rceil \approx 3\sqrt{ad} \tag{20.7}$$

and

$$C_{a,b,t}(0)\prod_{i=1}^{a-1} g_i(0) \geq \exp\left(4a - 1.06a\right) = \exp\left(2.94a\right).$$

To apply these bounds, we should state all the paramters in terms of the degree of the polynomial q. Let this degree be k. We then find

$$k \approx 3\sqrt{ad} \qquad \Longrightarrow \qquad a \approx \frac{k^2}{9d}$$

and

$$C_{a,b,t}(0)\prod_{i=1}^{a-1} g_i(0) \geq \exp\left((2.94)/9k^2/d\right).$$

As we will only consider $k \leq d/2$, this tells us that $a \leq d/36$, and so $r \geq 6$ and we can in fact apply Claim 20.6.1.

If we ignore the $\approx$ in (20.7) that comes from taking ceilings, Corollary 20.5.2 provides the bound

$$\tau \leq \exp\left(-2(2.94)/9k^2/d\right) = \exp\left(-0.653k^2/d\right).$$

If we restrict our attention to values of k on the order of ϵd, then the impact of the ceilings goes away and we can achieve this bound. As the correct constant in the exponent is 2, this exponent is off by a factor of approximately 3.

20.7 Analysis of $g_\lambda(0)$

Let

$$z = \frac{d\cos(\pi/2r) + \lambda}{d - \lambda}.$$

We need to prove a lower bound on

$$g_\lambda(0) = T_r(z).$$

If $z \geq 1$ then $T_r \geq 1$, and Claim 20.6.1 holds. If this is not the case, then we will define γ so that

$$\cos(\pi/2r - \gamma) = z.$$

We will then have

$$T_r(z) = \cos\left(r(\pi/2r - \gamma)\right) = \cos\left(\pi/2 - r\gamma\right) = \sin(r\gamma).$$

To prove a lower bound on γ, I will apply the following lemma.

Lemma 20.7.1. *Let $\theta \in [0, \pi/2]$ and $y \geq 0$ be numbers such that $\cos(\theta) + y \leq 1$. If we define $\gamma \in [0, \pi/2]$ so that*

$$\cos(\theta - \gamma) = \cos(\theta) + y,$$

then

$$\gamma \geq y/\theta.$$

Proof. We have

$$\cos(\theta - \gamma) = \cos\theta\cos\gamma + \sin\theta\sin\gamma,$$

which implies

$$\sin\theta\sin\gamma = y + (1 - \cos\gamma)\cos\theta \geq y,$$

and so

$$\theta\gamma \geq \sin\theta\sin\gamma \geq y.$$

□

Proof of Claim 20.6.1. We have

$$z \geq (\cos(\pi/2r) + \lambda/d)\,(1 + \lambda/d).$$

Assuming $r \geq 6$, $\cos(\pi/2r) \geq 0.96$. So,

$$z \geq \cos(\pi/2r) + (1.96)(\lambda/d).$$

As long as this is less than 1, Lemma 20.7.1 says

$$\gamma \geq \frac{3.92 r\lambda}{d\pi}.$$

As $T_r(X)$ is increasing for $X \geq \cos(\pi/2r)$,

$$T_r(z) = \sin(r\gamma) \geq \sin\left(\frac{3.92 r^2 \lambda}{d\pi}\right).$$

Substituting our choice of $r = \lceil d/a \rceil$ then yields

$$T_r(z) \geq \sin\left(\frac{3.92\lambda}{a\pi}\right) \geq \sin\left(\frac{3.92}{\pi}\right)\frac{\lambda}{a},$$

as sin is convex in this region. The claim now follows from

$$\sin\left(\frac{3.92}{\pi}\right) > 0.94.$$

□

20.8 Open Questions

First, it is clear that we can improve the bounds in this lecture. Can they be made tight?

Second, can this technique be used to improve the bounds for any other probability spaces. For example, one can consider the space of k-subsets of a set of d elements. The corresponding graph is a Johnson graph, and it is known that these graphs also have integer eigenvalues (see [GR01, KS07]). For examples of other probability spaces where isoperimetric theorems have been used to prove concentration of measure theorems, I recommend the paper by Amir and Milman [AM85b].

References

[AL86] Owe Axelsson and Gunhild Lindskog. On the rate of convergence of the preconditioned conjugate gradient method. *Numerische Mathematik*, 48(5):499–523, 1986.

[AM80] D. Amir and V. Milman. Unconditional and symmetric sets in n-dimensional normed spaces. *Israel Journal of Mathematics*, 37:3–20, 1980. 10.1007/BF02762864.

[AM85a] Noga Alon and V. D. Milman. λ_1, isoperimetric inequalities for graphs, and superconcentrators. *J. Comb. Theory, Ser. B*, 38(1):73–88, 1985.

[AM85b] D. Amir and V. Milman. A quantitative finite-dimensional Krivine theorem. *Israel Journal of Mathematics*, 50:1–12, 1985. 10.1007/BF02761116.

[GR01] Chris Godsil and Gordon Royle. *Algebraic Graph Theory*. Graduate Texts in Mathematics. Springer, 2001.

[Jen77] A. Jennings. Influence of the eigenvalue spectrum on the convergence rate of the conjugate gradient method. *IMA Journal of Applied Mathematics*, 20(1):61–72, 1977.

[KS07] M. Krebs and A. Shaheen. On the spectra of johnson graphs, 2007.