18.409 An Algorithmist's Toolkit Nov. 5, 2009

Lecture 16

Lecturer: Jonathan Kelner *Scribe: Paul Christiano*

The Chernoff Bound as Concentration of Measure

We have already seen some ways in which convex bodies are related to probability. For example, we can think of the Chernoff bound as the statement that for any unit vector a and real t, if x is chosen uniformly at random from the cube then

$$\Pr\left[|a \cdot x| > t\right] \leq 2e^{-6t^2}.$$

Since $|a \cdot x|$ is the distance of x from the hyperplane orthagonal to a, this says that all but $2e^{-6t^2}$ of the volume of the cube lies at distance at most t from this hyperplane. Since a was arbitrary, we can conclude that $1 - 2e^{-6t^2}$ of the volume of the cube lies within t of any hyperplane through the origin.

On the sphere we also observed that almost all of the volume lies very close to any hyperplane through the origin. In light of the probabalistic implications of this assertion for the cube we are motivated to consider them for the sphere. First we will need to derive a stronger statement for the sphere.

The Isoperimetric Inequality on the Sphere

We will consider the analogue of the isoperimetric question for subsets of the surface of the sphere. This requires analogues of the notions of distance, volume, and surface area.

We define the distance $d(x, y)$ between points $x, y \in S^{n-1}$ to be their distance in the usual Euclidean metric in $\mathbb{R}^n$.

For volumes, we use the unique rotationally invariant measure on the surface of the sphere. The volume $\mathrm{Vol}(A)$ of a region A on the surface of the sphere is the volume of the union in $\mathbb{R}^n$ of all segments connecting the origin to a point of A, normalized so that the volume of the whole sphere is 1. Alternatively, this is Haar measure when the sphere is given the natural Lie group structure. (You can do anything reasonable and get the same measure.)

For surface areas, we use the same definition as in $\mathbb{R}^n$. Namely, for a set $A \subset S^{n-1}$, define A_ϵ to be the set of points in S^{n-1} at a distance of less than ϵ from some point of A. The surface area of A may be defined as $\partial_\epsilon A_\epsilon$. We won't work with this quantity–instead we will derive bounds on $\mathrm{Vol}(A_\epsilon)$ itself for $\epsilon > 0$.

Now the isoperimetric question is: among sets with a fixed $\mathrm{Vol}(A)$, what is the minimal possible value of $\mathrm{Vol}(A_\epsilon)$?

The answer is the analogue of a ball: a spherical cap. More precisely, define

$$C(r, v) = \left\{x \in S^{n-1} : d(x, v) \leq r\right\}.$$

This is the ball of radius r centered at v in the metric we have defined on the sphere. This result is precisely analogous to the isoperimetric inequality in $\mathbb{R}^n$. (The statement itself will be slightly more complicated because the optimal ratio A_ϵ/A depends on the volume of A: a small cap is basically a ball in $\mathbb{R}^{n-1}$, while a very large cap has a very small surface area)

For convenience, we will also define the "cap at height t":

$$c_t = c(t, v) = \left\{x \in S^{n-1} : x \cdot v \geq t\right\}.$$

We have seen previously that the volume of a section of the sphere at height t is exponentiall small. From this it follows that the volume of $c(t, v)$ is exponentially small in t. In fact, $\mathrm{Vol}(c(t, v)) \approx e^{-nt^2/2}$.

We will prove an approximation to this result soon, but first we consider some consequences.

Theorem 1 *For any A with $\mathrm{Vol}(A) = 1/2$, $\mathrm{Vol}(A_\epsilon) \geq 1 - e^{-n\epsilon^2/2}$.*

Proof If A were a spherical cap, then A_ϵ would be the complement of the spherical cap at height ϵ, which has volume $1 - e^{-n\epsilon^2/2}$. But by the isoperimetric inequality this is the minimum possible value of $\mathrm{Vol}(A_\epsilon)$. ■

This theorem shows that for spheres in high enough dimension almost all of the volume of the sphere lies within ϵ of any set containing at least half the volume of the sphere. In fact almost all of the volume of the sphere lies within ϵ of any set containing any constant fraction of the volume of the sphere (although the constants in the theorem would change).

We will now go on to use this result to conclude that Lipschitz functions are almost always close to their median.

Lipschitz Functions and Concentration of Measure

Definition 2 (1-Lipschitz) *A function $f : S^{n-1}\mathbb{R}$ is 1-Lipschitz if $|f(a)-f(b)| \leq |a-b|$ for all $a, b \in S^{n-1}$.*

It turns out that many reasonable functions are Lipschitz. For example, distance from a fixed set is Lipschitz.

Define a median M of a Lipschitz function to be a value M such that $\mathrm{Vol}(\{x : f(x) \leq M\}) \geq \mathrm{Vol}(\{x : f(x) \geq M\}) = 1/2$.

If we take f were one of the coordinate functions (which are Lipschitz), then the statement that most of the volume of a sphere lies near any hyperplane through the origin becomes the statement that the value of f is almost always near its median. We will see that in fact all Lipschitz functions are almost always near their median.

Theorem 3 *If f is Lipschitz, M is its median, and $\epsilon > 0$, then*

$$\mathrm{Vol}(\{x : |f(x) - M| > \epsilon\}) \leq 2e^{-n\epsilon^2/2}.$$

Proof The set $A = f(x) \leq M$ has volume at least $1/2$. The set $f(x) \leq M + \epsilon$ contains A_ϵ. Therefore by the isoperimetric inequality, $f(x) \leq M + \epsilon$ holds for at least $1 - e^{-n\epsilon^2/2}$ of the volume of the sphere. Similarly, $f(x) \geq M - \epsilon$ holds for $1 - e^{-n\epsilon^2/2}$ of the volume of the sphere. Therefore in total $|f(x) - M| > \epsilon$ for at most $2e^{-n\epsilon^2/2}$ of the volume of the sphere (since at every point where this inequality holds at least one of the previous two must fail). ■

Although the range of a 1-Lipschitz function may have diameter 2, this result shows that 1-Lipschitz functions are almost constant over most of their domain. We call this result "concentration of measure."

Note that this result doesn't rely on the exact form of the isoperimetric inequality; it would be fine if the bound on the ratio $\mathrm{Vol}(A_\epsilon)/\mathrm{Vol}(A)$ was somewhat weaker.

The Isoperimetric Inequality

We will prove a weaker statement than the full isoperimetric inequality because it is somewhat easier. Normally we would have to use a symmetrization argument, but after weakening the constantants we will be able to apply Brunn-Minkowski.

Theorem 4 *For any $A \subset S^{n-1}$ and any $\epsilon > 0$*

$$\mathrm{Vol}(A_\epsilon) > 1 - \frac{2e^{-n\epsilon^2/16}}{\mathrm{Vol}(A)}.$$

Proof

We will need the following definition.

Definition 5 (Modulus of Convexity) *The modulus of convexity δ for a sphere is*

$$\delta(\epsilon) = \inf \left\{ 1 - |\frac{x+y}{2}| : x, y \in S^{n-1}, |x-y| \geq \epsilon \right\}.$$

It is a matter of two dimensional geometry to compute

$$\delta(\epsilon) = 1 - \sqrt{1 - \frac{\epsilon^2}{4}} \geq \epsilon^2/8.$$

(where the inequality comes from the Taylor series).

This quantity measures how much more curved the sphere is than required by convexity. Namely, by convexity we are guaranteed that $\delta(\epsilon) \leq 1$ (which we would obtain in the L_1 or L_∞ norm). If $\delta(\epsilon)$ is smaller, it means that longer segments lie well inside the convex body.

We would like to apply Brunn-Minkowski, but we don't have any result of that sort for the surface of the sphere. We will pass to a spherical shell, for which we can apply Brunn-Minkowski. Namely, if $A \subset S^{n-1}$ consider $B = [\frac{1}{2}, 1]A$– the union of the sets xA for $\frac{1}{2} \leq x \leq 1$. Note that $\text{Vol}(B) \geq \text{Vol}(A)/2$, where the volume of B is taken in $\mathbb{R}^{n-1}$ normalized so that B^n has volume 1 and the volume of A is taken in S^{n-1} normzlied so that S^{n-1} has volume 1. The choice of 1/2 in particular is not important. All that matters is that neighborhoods of $[\frac{1}{2}, 1]A$ centrally project to reasonable neighborhoods of S^{n-1}; if we took $(0, 1]A$, neighborhoods near the origin could project to almost all of S^{n-1}.

To go from a set $B \subset B^n$ to an $A \subset S^{n-1}$ we take $\left\{ \frac{x}{|x|} : x \in B \right\}$. Note that if we define $B = [\frac{1}{2}, 1]A$, take B_ϵ, and then convert this back to a subset of S^{n-1}, we do not necessarily obtain A_ϵ. A point within ϵ of $\frac{1}{2}A$ may project back to a point on S^{n-1} as far as 2ϵ from A. In fact this is the worst that can happen, so that B_ϵ is carried back into $A_{2\epsilon}$. We would like to say that the volume of $B_\epsilon \cap B^n$ is at least the volume of $A_{2\epsilon}$, so that we can convert a bound on the size of B_ϵ from Brunn-Minkowski into a bound on the size of $A_{2\epsilon} \supset A_\epsilon$. This isn't quite true–$B_\epsilon$ may contain points of norm $< 1/2$. However, all points in B_ϵ have norm at least $1/2 - \epsilon$, so it turns out this does not have a significant effect ($\text{Vol}\left([\frac{1}{2} - \epsilon, \frac{1}{2}]A\right)$ is very small).

We will show that $\text{Vol}(B_\epsilon \cap B^n) \geq 1 - e^{-2n\delta}/\text{Vol}(B)$. This will give us the desired result, since then

$$\text{Vol}(A_{2\epsilon}) > (1+\epsilon)\text{Vol}(B_\epsilon \cap B^n) \geq 1 - \frac{e^{-2n\delta(2\epsilon)}}{\text{Vol}(B)} \geq 1 - 2\frac{e^{n\epsilon^2/2}}{\text{Vol}(A)}$$

which is what we wanted.

To bound the volume of $B_\epsilon \cap B^n$, let C be the set of points of B^n at least ϵ away from every point of B. For any $x \in B$ and any $y \in C$, by the definition of modulus of convexity $\frac{|x+y|}{2} \leq 1 - \delta(\epsilon)$ (the worst case is that both lie in S^{n-1}). This implies that $B \oplus C \subset (1-\delta)B^n$, so that $\text{Vol}(B \oplus C)^{1/n} \leq (1-\delta)$. Now by Brunn-Minkowski,

$$(1-\delta) \geq \text{Vol}(B \oplus C)^{1/n} \geq \text{Vol}(B)^{1/n} + \text{Vol}(C)^{1/n}.$$

By easy calculus or the power-mean inequality, and the inequality $e^{-x} \geq 1 - x$, we conclude

$$\text{Vol}(B)^{1/2}\text{Vol}(C)^{1/2} \leq (1-\delta)^n$$

$$\text{Vol}(C) \leq (1-\delta)^2 n/\text{Vol}(B) \leq e^{-2n\delta}/\text{Vol}(B)$$

Taking complements in B^n,

$$\text{Vol}(B_\epsilon \cap B^n) = 1 - \text{Vol}(C) \geq 1 - \frac{e^{-2n\delta}}{\text{Vol}(B)}$$

as desired. ■

Johnson-Lindenstrauss

Johnson-Lindenstrauss can be proved by manipulating Gaussians, but it is quite easy with concentration of measure. For now we will just give the setup and outline some applications.

This is the first example we have seen of the notion of metric embeddings, which turn out to be generally algorithmically useful. Given a metric d on a finite set of points X, we would like to find a map $f : X \to \mathbb{R}^n$ such that $d(x, y) \approx d(f(x), f(y))$ for the normal Euclidean metric d on $\mathbb{R}^n$. More precisely, for any map $f : X \to \mathbb{R}^n$ we define the distortion D to be the ratio between the largest and smallest values of $\frac{d(x,y)}{d(f(x),f(y))}$ as x and y vary. We would like to find an embedding with $1 + \epsilon$ distortion.

The Johnson-Lindenstrauss Theorem states that if the metric on X arises from an embedding of X into any Euclidean space, then X can be embedded with distortion at most $1 + \epsilon$ in R^k for $k = O(\epsilon^2 \log |X|)$. More concretely, this embedding is given by projection onto a random k-dimensional subspace, and the ratio $d(x, y)/d(f(x), f(y))$ is very nearly $O(\sqrt{k/n})$.

This result is extremely useful in a number of situations. If I wish to answer some question about a fixed set of points which depends only on their pairwise distances, then Johnson-Lindenstrauss allows us to reduce the problem to one in logarithmic dimension (for fixed ϵ) by randomly projecting. If our algorithm has bad dependence on the dimension, this may reduce the runtime considerably (for example, exponential dependence becomes polynomial). Similarly, if I am dealing with a stream of very high-dimensional data and I do not have storage space to record it all, Johnson-Lindenstrauss allows us to retain a very small fraction of this data while preserving the answer to any question which depends only on distances.

We will prove this result in the next lecture.

18.409 Topics in Theoretical Computer Science: An Algorithmist's Toolkit
Fall 2009