18.409: An Algorithmist’s Toolkit
Lecture 16

Jonathan Kelner
Today

- Concentration of Measure and Geometric Probability Theory

- **Administrivia:** Posted L-S paper that proves rel. isop. inequality on course website
  - Note: I got the constant right last time
CONCENTRATION OF MEASURE AND GEOMETRIC PROBABILITY THEORY
Our First Concentration of Measure Theorem

- Last time, started looking at relationship between convex bodies and probability.
- Provided geometric interpretation of Chernoff bound:
  \[
  \Pr \left[ \left| \sum_{i=1}^{n} a_i x_i \right| > t \right] \leq 2e^{-6t^2} \quad x \in [-1/2, 1/2]^n \quad \sum_{i} a_i^2 = 1
  \]
  \[
  \sum a_i x_i = a \cdot x = \text{distance of } x \text{ from hyperplane } H_a = \{x \mid a \cdot x=0\}
  \]
  - Since \( \sum a_i^2 = 1 \)
- So theorem says “most of the mass of the cube lies near any hyperplane through the origin”
- **Can also phrase as:** Let \( S = \{\text{all points within distance } t \text{ of } H_a\} \)
  \[
  \frac{\text{Vol}(S)}{\text{Vol}([-1/2, 1/2]^n)} \geq 1 - 2e^{-6t^2}
  \]
- Feels like our statement that most of the volume of the n-sphere lies near the equator
The Sphere

- Let’s ask the isoperimetric question on the sphere $S = S^{n-1} \subseteq \mathbb{R}^n$
- For distance, will use distance in $\mathbb{R}^n$ (not on sphere)
- Volumes w.r.t. unique rotationally invariant measure on sphere normalized to have total $= 1$.
- For any $A \subseteq S$, let $A_\varepsilon$ be the set of points within distance $\varepsilon$ of $A$
- **Question:** Out of all sets with $\text{vol}(A) = q$, which ones gives min possible $\text{vol}(A_\varepsilon)$?
- **The answer:** Spherical caps
- **Definition:** Let the spherical cap $C(r,v)$ be given by

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

- Just a ball in the metric on the sphere
  - So like Euclidean isoperimetric inequality
- Need a slightly more complicated formulation because volume ratio depends on radius
Spherical Caps

![Diagram of Euclidean metric on the sphere and a spherical cap.]

**Figure 26.** The Euclidean metric on the sphere. A spherical cap (right) is a ball for this metric.

- Sometimes easier to talk about “cap at height $t$”:
  $$c_t = c(t, v) = \{ x \in S^{n-1} | x \cdot v \geq t \}$$

- $c_0 = \text{hemisphere}$
- $\text{vol}(c_t) \approx e^{-nt^2}/2$
What Does This Imply?

• Will prove approximate version of this later. Let’s first see why it’s interesting.
  ◦ Actual proof is sort of tricky, but can get the asymptotics right very easily with what we already know
• **Corollary 1:** For any A with $\text{vol}(A) = 1/2$, 
  $\text{vol}(A_{\varepsilon}) \geq 1 - e^{-n\varepsilon^2 / 2}$
  ◦ Why?
• So almost all of the sphere is near *any* set of volume 1/2, even though there may be many points far away from it.
• Could have used 1/4 or 1/8 too and gotten the same thing (but with different constants)!

*Figure 27.* An $\varepsilon$-neighbourhood of a hemisphere.
Lipschitz Functions

- **Definition:** A function $f: S^{n-1} \rightarrow \mathbb{R}$ is **1-Lipschitz** if for any $a, b \in S^{n-1}$,
  \[ |f(a) - f(b)| \leq ||a - b||_2 \]
- Let $M$ be the median of a 1-Lipschitz function $f$, so $\text{vol}\{f \leq M\} = \text{vol}\{f \geq M\} \geq 1/2$
- How many points have $f(x) \leq M + \varepsilon$?
  - See blackboard
- **Theorem:** $\text{vol}\{x : |f(x) - M| > \varepsilon\} \leq 2e^{-n\varepsilon^2/2}$
- So even though $f$ can vary by up to 2 on the sphere, it is almost equal to $M$ almost everywhere
- So “1-Lipschitz functions are almost constant”
  - A little weird...
- This is what we’ll call concentration of measure
- Note we just need the right asymptotics for this, not the exact isoperimetric inequality.
We have two types of properties:

**Isoperimetric inequalities:** Every set $A$ with $\text{vol}(A) \geq 1/2$ has $\text{vol}(A_{\epsilon} \geq 1 - e^{-ct^2})$

**Concentration theorems:** For every 1-Lipschitz function $f$, $\text{vol}\{x : |f(x) - M| > t\} \leq 2e^{-ct^2/2}$

We showed that isoperimetric inequalities imply concentration theorems

Can go the other way as well: apply the concentration theorem to the function $f(x) = \text{dist}(x,A)$ [exercise]

Works whenever we have a notion of volume and distance
Proof of Isoperimetric Inequality for the Sphere

- We’ll prove the following theorem, which is qualitatively the same as the isoperimetric inequality but has different constants:

  **Theorem:** For any \( A \subseteq S^{n-1} \), and any \( \varepsilon > 0 \),
  \[
  \text{vol}(A_{\varepsilon}) > 1 - \frac{2e^{-n\varepsilon^2/16}}{\text{vol}(A)}
  \]

- Note the dependence on \( \text{vol}(A) \)
- Proof will use following definition:

  **Definition:** The modulus of convexity of the sphere is given by
  \[
  \delta(\varepsilon) = \inf \left\{ 1 - \left| \frac{x+y}{2} \right| \middle| |x| = |y| = 1, |x-y| \geq \varepsilon \right\}
  \]

- Easy to check that
  \[
  \delta(\varepsilon) \geq 1 - \sqrt{1 - \frac{\varepsilon^2}{4}} \geq \frac{\varepsilon^2}{8}
  \]

- **Proof of Theorem:** See blackboard.
Metric Embeddings

- We’ll get a very nice theorem (Johnson-Lindenstrauss) as a simple corollary of measure concentration
  - It’s a first example of a very powerful algorithmic tool: *metric embeddings*
- Let $d$ be any metric $d(x_i, x_j)$ on a finite set $X=\{x_1, ..., x_n\}$
- Let $f: X \rightarrow \mathbb{R}^n$ be a such that
  $$||f(x_i) - f(x_j)|| \leq d(x_i, x_j)$$
- The distortion of $f$ will be the maximum $D$ for which $d(x_i, x_j) \leq D ||f(x_i) - f(x_j)||$
- **Claim [J-L]:** The Euclidean metric on any finite set $X$ can be embedded with distortion $(1+\varepsilon)$ in $\mathbb{R}^k$ for $k=O(\varepsilon^2 \log n)$
- **Note:** If we lose the $\varepsilon$, this becomes impossible for any $k<n$
- The embedding will just be:
  - Let $y$ = projection of $x$ onto random $k$-dim. subspace
  - Return $cy$ for some (fixed) constant $c$
The Johnson-Lindenstrauss Theorem

- Now the precise statement:
  - **Theorem (Johnson-Lindenstrauss):** Let:
    - $X = \{x_1, \ldots, x_n\} \subseteq \mathbb{R}^m$ for any $m$
    - $k = O(\varepsilon^{-2} \log n)$
    - $L \subseteq \mathbb{R}^m$ be a uniformly random $k$-dimensional subspace
    - $y_1, \ldots, y_n$ = projections of the $x_i$ onto $L$
    - Let $y'_i = cy_i$ for some fixed constant $c$
    - $c = \Theta \left( \sqrt{\frac{k}{n}} \right)$

Then, with high probability,

$$||x_i - x_j|| \leq ||y'_i - y'_j|| \leq (1+\varepsilon)||x_i - x_j||$$
Proof Idea

- Let $f(x) = \sqrt{x_1^2 + \ldots + x_k^2}$
  - Length of proj. of $x$ onto fixed $k$-dim subspace
- Let $M = \text{median}(f)$ over the sphere
- $f$ is Lipschitz, so tightly concentrated
  - That is, a random $x$ on sphere has $f(x)$ very close to $M$
- Instead of choosing random $x$ and fixed $k$-dim subspace, take fixed $x$ and random subspace
  - Same thing
- So projection onto random subspace is very likely to send a unit vector to something of length $\approx M$
  - So vector of length $k \rightarrow$ vector of length $\approx kM$
- Apply this to vectors $(x_i - x_j)$ for all $i,j$
- See board for missing details
Applications of J-L

- Tons of algorithmic applications of this theorem
- It says you can approximately answer questions about m-dimensional objects in a much lower-dimensional space
- High-dimensional geometry questions often have complexity growing very quickly in the dimension
- Can be used to speed up almost any such algorithm
- We’ll see a few examples
Where Do We Get A Random Subspace?

- The algorithms for these problems all say that we should embed our high-dimensional points into a low-dim space.
- This means that we actually need the embedding, not just the theorem that one exists.
- Proof was constructive(-ish): said a random subspace works.
- How do we get a random subspace?
- First, how do we get a random point on the sphere
  - Gaussians
Quick Comments on Gaussians

- Gaussians are really nice
- A multivariable Gaussian has pdf
  \[ f_X(x_1, \ldots, x_N) = \frac{1}{(2\pi)^{N/2}|\Sigma|^{1/2}} \exp \left( -\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right) \]
  for some nonsingular matrix \( \Sigma \) (covariance matrix) and vector \( \mu \) (mean)
- The following operations on Gaussian variables yield other Gaussian variables
  - Project onto a lower-dim subspace
  - Restrict to a lower-dim subspace (conditional probs)
  - Apply any linear map
- In particular, coordinates of Gaussians with \( \Sigma = \text{Id} \) are 1-d Gaussians
- So can get a vector from it by choosing each coordinate independently from 1-d Gaussians
- Can get other Gaussians by applying linear maps
More on Gaussians

- Lets you get a random vector on sphere
  - How?
- When each coord from Gaussian with mean 0 and var 1, most n-vectors have norm $n^{1/2}$
- So Gaussian variable looks a lot like random vector on sphere
  - Other Gaussians give ellipses
- Central limit theorem just says product distributions of (nice enough) 1-d distribs look like Gaussians if have enough products
  - So cubes look like spheres too
  - Chernoff is a concentration of measure theorem about discrete cubes
Generating a Random Subspace

• So can get random vector on sphere by generating n 1-dim Gaussians

• Can get a random k-dim subspace by generating kn 1-dim Gaussians
  ◦ How and why?
  ◦ Projections of Gaussians are Gaussians, so get random projection map
Applications of J-L

- Approximately solve a bunch of problems
- Proximity problems:
  - Input: a set $P$ of points in $\mathbb{R}^d$
  - Want to compute any property defined only in terms of distances between points in $P$
  - Closest pair, furthest pair, minimum spanning tree, minimum cost matching, many clustering problems, ...
  - Can solve in lower-dim space—much faster
Applications of J-L

- On-line problems:
  - Get points as you go
  - Have to create data structure that answers some query quickly, isn’t too hard to construct, and answers queries quickly
  - Tend to have exponential dependence on dimension: $(1/\varepsilon)^d$
  - So dimension reduction *really* helps
Applications of J-L

- Data structures with sublinear storage and streaming algorithms
- For example, get data in a stream and can’t store it all
- Want to be able to answer questions about it later
- E.g., are there a lot of repeated entries
  - How are the counts distributed?
- Idea is to use “sketches” based on J-L and other such probabilistic techniques
- Piotr Indyk taught a class on this
- See Indyk survey and its references for more on J-L applications
Let $C$ be an origin-symmetric convex body in $\mathbb{R}^n$
  ◦ And thus gives a norm
Let $S \subseteq \mathbb{R}^n$ be a vector subspace

**Question:** when does $Y = C \cap S$ look like a sphere?

Call these slices “sections”

Formally, when does $Y$ have small Banach-Mazur distance from sphere
  ◦ I.e., exists a linear transformation so that $Y$ contains the unit ball contained in the ball of radius $(1+\varepsilon)$

More interesting question: for how high a dimension does there exist an $S$ for which this occurs?
  ◦ Call this dimension $k$

Any guesses?
Examples

- An ellipse
  - $k = n$
- The cross-polytope (= unit ball in $l_1$)
  - Has $k = \Theta(n)$
- However, the cube (= unit ball in $l_{\infty}$) has $k = \log(n)$
- Turns out this is essentially the worst case
Dvoretzky’s Theorem

- **Theorem:** There is a positive constant $c > 0$ such that, for all $\varepsilon$ and $n$, every $n$-dimensional origin-symmetric $K$ has a section within distance $1 + \varepsilon$ of the unit ball of dimension

$$k \geq \frac{c\varepsilon^2}{\log(1 + \varepsilon^{-1}) \log n}$$

- We’ll sketch the proof
- Gives very interesting statements about norms
- Could also view this as a theorem about projections
  - How?
- How does this compare to Johnson-Lindenstrauss?