Lecture 10

Lecturer: Jonathan Kelner

In this lecture, we shall revisit the Spectral Sparsifiers and see a slightly different proof from last time. We will then begin a new topic: Convex Geometry.

1 Spectral Sparsification

Given a dense graph G, we would like want to create a sparse graph H where

$$L_h \preceq L_G \preceq (1+\epsilon)L_H$$

By "sparse," we mean that H has n.polylog(n) edges, where n is the number of nodes. More precisely, we show how to construct a spectral sparsifiers with $O(nlogn)$ edges in Polynomial time. This can actually be improved to a linear time construction, but will use geometric techniques that we will learn. It is possible to construct $O(n)$ edge sparsifiers in polynomial time. It is also a nice example of how generalizing can make things easier sometimes. The algorithm that we propose is very simple. It is similar in structure to the B-K algorithm, but we use different probabilities for sampling the edges.

- Compute probability p_e for each edge e.
- Sample each edge uniformly with probability p_e, and if an edge is selected, include it with weight $1/p_e$.

These probabilities are based on a linear algebra sense of importance, and have a nice interpretation in terms of effective resistance of circuits. To proceed with our analysis, however, we need to develop the ideas of pseudoinverses, calculating effective resistances, and a matrix version of the Chernoff Bound.

1.1 Laplacians and Electrical Flow

We mentioned earlier that Spectral Sparsification can be viewed as sampling edges with different probability. It turns out that the correct way to do this is to sample each edge with probability proportional to its "effective resistance." The basic idea is to treat each edge as a resistor with resistance 1. If the edge had a capacity of c, we give it a resistance of $1/c$. After calculating these values, we sample the edge (u, v) with probability proportional to the effective resistance between nodes u and v. For example, students may use a combination of Ohm's law and Kirchoff's law, as well as the rules for calculating effective resistances of resistors in series and parallel. To those who are comfortable with solving circuits, this may be a good way to think about the problem. However, the students who don't like solving circuits are in luck too: now that we have the tools of Spectral Graph Theory, we can solve circuits with only linear algebra! In fact, we will combine our frequent use of the graph Laplacian with the pseudoinverse defined above. We orient the edges arbitrarily and define U to be the edge-vertex adjacency matrix. That is, we define U as in:

$$U(e,v) = \begin{cases} 1 & \text{if } v \text{ is the head of } e \\ -1 & \text{if } v \text{ is the tail of } e \\ 0 & \text{otherwise} \end{cases}$$

We then let $L = U^T U$. From ohm's law, we have $iR_{eff} = Uv$ for $i \in \mathbb{R}^E$ and $v \in \mathbb{R}^v$. From the conservation of current, we have $i_{ext} = U^T i$, for $i_{ext} \in \mathbb{R}^V$. Finally, we have $i_{ext} = Lv$, and $v = L^+ i_{ext}$

Let u_e be the e^{th} row of U (as defined in the prequel), and $v = L^+ i_e xt$. We have

$$R_{eff}(e) = u_e L^+ u_e^T$$

and as a result,

$$R_{eff}(e) = (UL^{+}U^{T})_{e,e}$$

Thus, calculating the effective resistance of an edge is as simple as calculating the pseudoinverse of the Laplacian. Simple!

1.2 Towards Approximation

To show that H is a spectral sparsifier of G if suffices to show that

$$(1-\epsilon)x^T L_G x \leq x^T L_H x \leq (1+\epsilon)x^T L_G x, \quad \forall x$$

For this, it suffices to show that, $\forall y$,

$$(1-\epsilon) \leq \frac{y^T (L_G^+)^{\frac{1}{2}} L_H (L_G^+)^{\frac{1}{2}} y}{y^T (L_G^+)^{\frac{1}{2}} L_G (L_G^+)^{\frac{1}{2}}} \leq (1+\epsilon), \quad \text{(Just take } y = L_G^{\frac{1}{2}} x)$$

Equivalently, need to show that

$$\| \ (L_G^+)^{\frac{1}{2}} L_H (L_G^+)^{\frac{1}{2}}) - I_{im(L_G} \ \|_2 \ \leq \ \epsilon$$

We will use the following theorem (in which k is a universal constant):

Theorem 1 (RV Theorem) *For distributions on vectors y where $\| \ y \ \| \leq t$ and $\| \ Eyy^t \ \|_2 \leq 1$ (where we are using the l_2 norm) then:*

$$E \| \ Eyy^T - \frac{1}{q}\sum_{i=1}^{q} y_i y_i^T \ \|_2 \leq kt\sqrt{\frac{\log q}{q}}$$

This is a "concentration of measure theorem"(similar to Chernoff bounds).

$$\begin{aligned}
L_G &= \sum_{e\in E} L_e = \sum_{e\in E} u_e u_e^T \\
I_{im(L_G)} &= L_G^+)^{\frac{1}{2}} L_G (L_G^+)^{\frac{1}{2}} \\
&= \sum_{e\in E} L_G^+)^{\frac{1}{2}} L_e (L_G^+)^{\frac{1}{2}} \\
&= \sum_{e\in E} L_G^+)^{\frac{1}{2}} U_e U_e^T (L_G^+)^{\frac{1}{2}} \\
&= \sum_{e\in E} q_e q_e^T, \text{ where } q_e = (L_G^+)^{\frac{1}{2}} u_e \\
\| \ q_e \ \|^2 &= u_e^T L_G^+)^{\frac{1}{2}} L_G^+)^{\frac{1}{2}} u_e \\
&= u_e L^+) u_e = R_{eff}(e)
\end{aligned}$$

$$I_{im(L_G)} = \sum_{e\in E} q_e q_e^T \text{ and } \| \ q_e \ \|^2 = R_{eff}(e)$$

We would like all the vectors of same length, so set $\tau_e = \sqrt{\frac{n-1}{c_e R_{eff}(e)}}\pi_e$ with $\| \ \tau_e \ \| = \sqrt{n-1}$. Now make a distribution which picks τ_e with probability $p_e = \frac{c_2 R_{eff}(e)}{n-1}$. Recall that

$$\sum_e c_e R_{eff}(e) = \sum_e \Pi_{e,e} = n-1$$

Then, we find that

$$E[\tau_e \tau_e^T] = \sum_e p_e \tau_e \tau_e^T = \sum_e qeq_e^T = I_{im(L_G)XS}$$

Sample q times with replacement, and set $S(e,e) = \frac{1}{qc_e R_{eff}(e)} \times$ the number of times that e is chosen. Then, from the theorem above, we have

$$E[\| (L_G^+)^{\frac{1}{2}} L_H (L_G^+)^{\frac{1}{2}}) - I_{im(L_G} \|_2] \leq k\sqrt{n-1}\sqrt{\frac{logN}{N}} \leq \epsilon, \forall N = \theta(n \log n/\epsilon^2)$$

.

1.3 Algorithmics of the Construction

Thus, we see that our construction yields a spectral sparsifier as desired. From the algorithmics of the construction, it is easy to see that this is a poly-time procedure. The whole procedure is constructive, and uses the standard linear algebra operations. The bottleneck in this procedure comes from computing effective resistances, and in particular, the matrix inversions and multiplications. We claim that the procedure can be improved to nearly linear time. Doing so would involve two components:

- Close to linear algorithms for solving linear equations of the form $Lx = b$ for a laplacian L.
- A way to compute all the effective resistances by solving logarithmically many linear systems. This uses the Johnson-Lindenstrauss Lemma.

1.4 Spectral Sparsification is Easy

- Pick N τ_e vectors with replacement from this distribution.
- Take an edge e with weight:

$$\frac{1}{N.R_{eff}(e) \times (\text{ number of times chosen })}$$

- Note: Bigger q vectors get picked with higher probability, but are scaled down more!
- By R-V Theorem,

$$E[\| (L_G^+)^{\frac{1}{2}} L_H (L_G^+)^{\frac{1}{2}}) - I_{im(L_G} \|_2] \leq k\sqrt{n-1}\sqrt{\frac{logN}{N}} \leq \epsilon, \forall N = \theta(n \log n/\epsilon^2)$$

  .

2 Convex Geometry

This lecture we will just have many examples to build intuition. Next lecture we will start proving theorems.

Definition 2 *We say a set $C \subseteq \mathbb{R}^n$ is convex when for all $x, y \in C$ and $t \in [0,1]$, $tx + (1-t)y \in C$. A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex iff the region above its graph (in $\mathbb{R}^{n+1}$) is convex. A function $f : \mathbb{R}^n \to \mathbb{R}$ is* concave *iff $-f$ is convex. A* convex body *is a convex set which is both compact and has non-empty interior.*

Keith Ball can be quoted as saying "All convex bodies behave a lot like Euclidean balls". This claim is "almost true" if one adds a few extra shapes: the ball, ellipsoid, cube, regular simplex, cross-polytope, and spherical cone (and all linear transformations of these shapes). Of course, this is not a formal statements: one can easily construct theorems which are satisfied for these shapes but not some other convex bodies. The point here is that for "most" theorems one would want to prove about convex bodies, if there were a counter-example there is a good chance that one of these shapes would be it.

We now give formal definitions of the shapes mentioned.

1. The *Euclidean ball* B_2^n is the set $\{x \in \mathbb{R}^n \mid ||x||_2^2 \leq 1\}$.
2. The *ellipsoid* E is the set $\{x \in \mathbb{R}^n \mid x^T Ax \leq 1\}$ where A is a positive semidefinite $n \times n$ matrix. Note we get the sphere when A is the identity matrix.
3. The *cube* B_∞^n is the set $\{x \in \mathbb{R}^n \mid ||x||_\infty \leq 1\}$.
4. The *simplex* C is the set $\{x \in \mathbb{R}^n \mid x_i \geq 0, \sum_i x_i \leq 1\}$.
5. The *cross-polytope* B_1^n is the set $\{x \in \mathbb{R}^n \mid ||x||_1 \leq 1\}$, which is the convex hull of all points of the form $(0, 0, \ldots, 0, \pm 1, 0, \ldots, 0)$. In $\mathbb{R}^2$ the cross-polytope and square are equivalent up to rotation of $\pi/4$. In $\mathbb{R}^3$ the cross-polytope is the octahedron. In general the cross-polytope in $\mathbb{R}^n$ has 2^n faces and $2n$ vertices (compare with the cube which has exactly the reverse), and acts as the "opposite" of the cube.

2.1 Geometric Intuition in High Dimension

The first thing to notice in high dimensions is that the vast majority of volume lies near the boundary of a convex body. For example, in $\mathbb{R}^2$ to get 1% of the volume of the square $[-1, 1]^2$ we can take the square $[-.1, .1]^2$. In 100 dimensions to get 1% of the volume of $[-1, 1]^{100}$ we would need to take the cube $[-.955, .955]^{100}$!

Big differences between balls and cubes also appear in high dimensions. For any n, to get a cube with volume 1 in $\mathbb{R}^n$ we can take a cube with sidelength 1. The story for cubes is different. The volume of a radius-r sphere in $\mathbb{R}^n$ is

$$\frac{r^n \pi^{n/2}}{\Gamma(\frac{n}{2} + 1)} \approx \left(r\sqrt{\frac{2\pi e}{n}} \right)^n$$

implying that in $\mathbb{R}^n$ we need to take a sphere of radius roughly $\sqrt{n/2\pi e}$ to get a volume of 1. In other words, balls in high dimensions are much smaller than cubes! Intuitively this makes sense. As we said previously, much of the volume in high dimensions lies near the boundary. If one imagines a sphere inscribed in a cube with sidelength equal to the sphere's diameter, very little of the sphere is near the cube's boundary.

Another thing to notice about high-dimensional balls is that much of the volume is concentrated around the equator. More concretely, define $v(t)$ as the $(n - 1)$-dimensional volume of $B_2^n \cap \{x_0 = t\}$. It turns out that $v(t)$ drops off dramatically as t deviates from 0. Quantitatively, one can show that that

$$v(t) \approx \sqrt{e} \left(\frac{\sqrt{r^2 - t^2}}{r} \right)^{n-1}$$

Thus, if one wishes to know what distance from the equator one has to slice to get, say, 96% of the sphere's volume, one can solve for t in the equation $\int_{-t}^{t} v(t)dt = .96\text{vol}(B_2^n)$ to find that the required value of t is quite small as a function of n (we leave the computation to the interested reader).

2.2 Maximizing Volume with a Given Surface Area

One important question in convex geometry is the following: "What is the most volume that can be enclosed in a convex body with a given surface area?". In $\mathbb{R}^2$ we can view the problem as us being given a string of some finite length and must arrange the string in the plane so as to maximize the area it encloses. The shape achieving this maximum area is of course the circle, but the proof is not trivial. We show a false proof that stood for quite some time before its major flaw was uncovered:

1. Let C be the shape achieving the maximum area. We can assume C is convex since if the line segment between x and y for some $x, y \in C$ is not in C, we can reflect about the segment $\overline{xy}$ to increase area.

2. We can assume C is symmetric about both the x and y axes. If not, first reflect the smaller-perimeter half of C about a line parallel to the x axis that bisects C's area. Then, do the same for y. If the resulting shape has smaller perimeter than C we arrive at a contradiction, since that extra "piece of string" could be used to increase the area of C. Otherwise, shift C so that its center is the origin (implying $(x, y) \in C \Leftrightarrow (-x, -y) \in C$).

3. If C is not a circle, let p be the point on C's boundary that is farthest away such that there exists a p' equidistant from the origin with p such that p' is not on the boundary of C. Reflect about the line that bisects the angle between p and p' so that C contains both p and p'. The area of the new shape is the same as that of C.

The main problem with this proof is in Step 1. One cannot simply assume that there exists a shape C which maximizes the area. In particular, to perform this type of argument one would first have to show that some metric defined on the space of convex bodies is complete.

References

[1] "Randomized Approximation Schemes for Cuts and Flows in Capacitated Graphs, " A. Benczur, D. Karger, manuscript.

MIT OpenCourseWare
http://ocw.mit.edu

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

For information about citing these materials or our Terms of Use, visit: http://ocw.mit.edu/terms.