Computational Geometry: Lecture 18

Don Sheehy

March 17, 2010

1 Centerpoints as Robust Statistics

Recall that last time, I motivated the study of centerpoints by couching them as a kind of geometric median. Before we go forward, I want to emphasize what we mean by a geometric median and what types of properties we expect it to have. I also want to make it very clear in what sense a centerpoint is *not* a median

We started by defining the Tukey depth of a point. We saw that the median in $1D$ is just a point of maximum Tukey depth $n/2$. We then saw that in the case of a triangle in $\mathbb{R}^2$, there could be no point of Tukey depth $n/2$. This dashed our hopes that we could simply define a median by always picking a point of depth $n/2$. Instead, we said that we would settle for a point of depth *at least* $\frac{n}{d+1}$. We called such a point, a centerpoint.

It is worth pointing out that there are nice point sets that do admit points of depth $n/2$. In such cases, we might not be entirely happy calling a centerpoint of depth *at least* $\frac{n}{d+1}$. A *Tukey median* is a point of maximum Tukey depth.

To understand a median, it is helpful to contrast it with the mean. We have seen the mean of a set of geometric points before. You will recall that we used the term centroid or barycenter to describe it. It is the point that is the average of all the other points.

The mean is a kind of statistic, just like a median. That is to say, it is a summary–one point standing in for the whole set. The mean suffers from a major disadvantage over the mean in that it can be very sensitive to outliers. This is why the media is often more helpful than the mean.

Consider the case where among a set of numbers, one of them is an outlier. It is such an outlier that we have moved it all the way out to infinity. The mean will also be infinity. The median on the other hand is unchanged. In general, this is what we mean when we talk about medians being robust statistics.

2 Centerpoints via Tverberg's Theorem

I'm going to give an alternative proof of the Centerpoint Theorem using another classic theorem of convexity theory, Tverberg's Theorem.

Theorem 2.1 (Tverberg's Theorem)**.** *Given a set P of $n \geq (r-1)(d+1)+1$ points in $\mathbb{R}^d$, there exists a partition of P into sets $U_1 \ldots, U_r$ such that*

$$\bigcap_{i=1}^{r} CC(U_i) \neq \emptyset.$$

We will call the partition guaranteed by the theorem, a Tverberg partition, and a point in the common intersection, a Tverberg point.

The statement of this theorem should look at least vaguely familiar to you. If we set $r = 2$, it is exactly Radon's Theorem. It is not surprising that when Tverberg published this Theorem, he called it "A Generalization of Radon's Theorem". I would love to tell you that the proof is as easy as the proof of Radon's Theorem, but sadly, it is not. So in lieu of proving this for you in class, we'll all just marvel at it for a bit.

Okay, so what does Tverberg's Theorem have to do with centerpoints? Let's use Tverberg's theorem to prove the centerpoint theorem.

Theorem 2.2 (The Centerpiont Theorem (again))**.** *Given n points in $\mathbb{R}^d$, there exists a point $x \in \mathbb{R}^d$ such that $depth(x) \geq \frac{n}{d+1}$.*

Proof. Let $r = \lceil \frac{n}{d+1} \rceil$. If we take a Tverberg point x, then it is in the convex closure of $r = \lceil \frac{n}{d+1} \rceil$ disjoint subsets of P. If we take a halfspace containing x, then it must contain at least one point from each of the sets in the partition. Thus, every halfspace containing x also contains $\lceil \frac{n}{d+1} \rceil$ points of P, so $depth(x) \geq \frac{n}{d+1}$. □

That was easy. In fact it was easier than our proof last time. Somehow, we packed more of the hard work in to Tverberg's theorem than we did last time with Helly's Theorem.

Are the centerpoints implied by this proof, the same as the centerpoints implied by proof via Helly's theorem? In other words, if x is a centerpoint, is there necessarily a partition of P into $\lceil \frac{n}{d+1} \rceil$ sets such that x is in the convex closure of each? In general, the answer is no, but before we see those examples, let's look at a special case where the answer is yes.

Theorem 2.3. *Given a set of $n = 3k$ points P in $\mathbb{R}^2$, a point x is a centerpoint i and only if x is a Tverberg point for $r = k$.*

Proof. We have already seen one direction, namely that every such Tverberg point is a centerpoint. It will suffice to prove that for every centerpoint x, there exists a partition of P into k parts, each of which contains x in its convex closure.

Let x be a centerpoint. Sort the points of P radially around x. Define the set U_i to be $\{p_i, p_{i+k}, p_{i+2k}\}$. Suppose for contradiction that $x \ni CC(U_i)$. Without loss of generality, let x be the origin. Then there must be some vector $v \in \mathbb{R}^2$ such that $v \cdot p > 0$ for all p. Since we sorted the points radially around x, it must be that $v \cdot p_i > 0$ for all $i \in \{i, \ldots, i+2k\}$ modulo some shift in the indices by a multiple of k. This implies that there are $2k+1 = 2n/3+1$ such points in the open halfspace defined by v. Thus, the complement halfspace is closed and contains fewer than $k = n/3$ points. So $depth(x) < n/3$ and thus x is not a centerpoint. This is a contradiction.

$\square$

3 Preview of next time

The computational problems related to centerpoints and Tverberg points are quite interesting. As the dimension of the input points increases, the difficulty seems destined to go up. For general dimensional points, the problem of testing a centerpoint is coNP-complete. That is, given a point x and a point set P, it is coNP-complete to determine if the point x is a centerpoint. On the other hand, the corresponding problem for Tverberg points is NP-complete. That is, given a point x and a set P, determining if x is in the common intersection of the convex closure of $\frac{n}{d+1}$ disjoint subsets of P is NP-complete. These two facts should at least give you an inkling that our Theorem about centerpoints and Tverberg points in the plane should not hold in higher dimensions. If so, it would imply that the testing problem is both NP-complete and coNP-complete and thus $NP = coNP$, something widely believed to be false.