CMSC 754: Lecture 18
Coresets and Kernels

Reading: This material is based on a paper by Agarwal, Har-Peled, and Varadarjan (JACM 2004). The presentation follows that given in Chapt. 23 of Har-Peled's book, Geometric Approximation Algorithms (AMS, 2011).

Approximation by Sampling: One of the issues that arises when dealing with very large geometric data sets, especially in multi-dimensional spaces, is that the computational complexity of many geometric optimization problems grows so rapidly that it is not feasible to solve the problem exactly. In the previous lecture, we saw how the concept of a well-separated pair decomposition can be used to approximate a quadratic number of objects (all pairs) by a smaller linear number of objects (the well separated pairs). Another approach for simplifying large data sets is to apply some sort of sampling. The idea is as follows. Rather than solve an optimization problem on some (large) set $P \subset \mathbb{R}^d$, we will extract a relatively small subset $Q \subseteq P$, and then solve the problem exactly on Q.

The question arises, how should the set Q be selected and what properties should it have in order to guarantee a certain degree of accuracy? The simplest and most natural approach is to select a *random subset*. While this may work for many problems, there are some instances where it may fail miserably. For example, suppose that you wanted to approximate the minimum enclosing ball (MEB) for a point set P (see Fig. 1(a)). A random subset may result in a ball that is much smaller than the MEB. The problem is that the MEB is determined by points that are "peripheral", whereas random sampling may sample an excessive number of points near the "center" of the point set (see Fig. 1(b)). We would like a sampling method that is tailored to selecting points that will be most valuable to the MEB construction (see Fig. 1(c)).

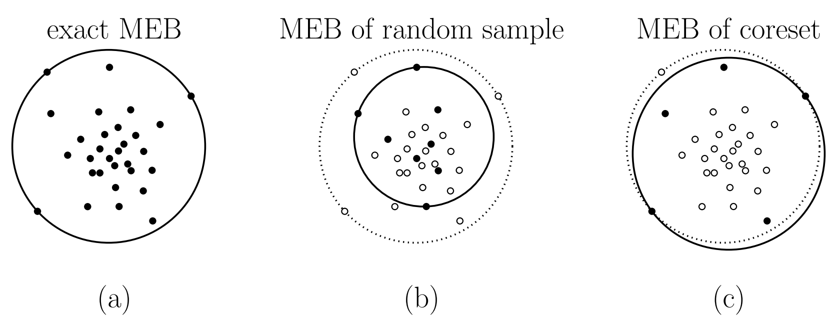

Fig. 1: (a) exact MEB, (b) MEB of a random sample, (c) MEB of a coreset.

Coresets: Abstractly, consider any optimization problem on point sets. For a point set P, let $f^*(P)$ denote the value of the objective function for the optimal solution. (For example, in the minimum enclosing ball problem, $f^*(P)$ is the radius of the MEB.) Given $\varepsilon > 0$, we say that subset $Q \subseteq P$ is an ε*-coreset* for this problem if, the relative error committed by solving the problem on Q is at most ε, that is:

$$1 - \varepsilon \ \leq\ \frac{f^*(Q)}{f^*(P)} \ \leq\ 1 + \varepsilon.$$

For a given optimization problem, the relevant questions are: (1) does a small coreset exist? (2) if so, how large must the coreset be to guarantee a given degree of accuracy? (3) how quickly can such a coreset be computed? Ideally, the coreset should be significantly smaller than n and it should be computable much faster than solving the original optimization problem. For many optimization problems, the coreset size is actually independent of n (but does depend on ε).

In this lecture, we will present algorithms for computing coresets for a problem called the *directional width*. This problem can be viewed as a way of approximating the convex hull of a point set.

Directional Width and Coresets: Consider a set P of points in real d-dimensional space $\mathbb{R}^d$. Given vectors $\vec{u}, \vec{v} \in \mathbb{R}^d$, let $(\vec{v} \cdot \vec{u})$ denote the standard inner (dot) product in $\mathbb{R}^d$. From basic linear algebra we know that, given any vector $\vec{u}$ of unit length, for any vector $\vec{v}$, $(\vec{v} \cdot \vec{u})$ is the length of $\vec{v}$'s orthogonal projection onto $\vec{u}$. The *directional width* of P in direction $\vec{u}$ is defined to be the minimum distance between two hyperplanes, both orthogonal to $\vec{u}$, that has P "sandwiched" between them. More formally, if we think of each point $p \in P$ as a vector $\vec{p} \in \mathbb{R}^d$, the directional width can be formally defined to be

$$W_u(P) \;=\; \max_{p \in P}(\vec{p} \cdot \vec{u}) - \min_{p \in P}(\vec{p} \cdot \vec{u})$$

(see Fig. 2(a)). Note that this is a signed quantity, but we are typically interested only in its magnitude.

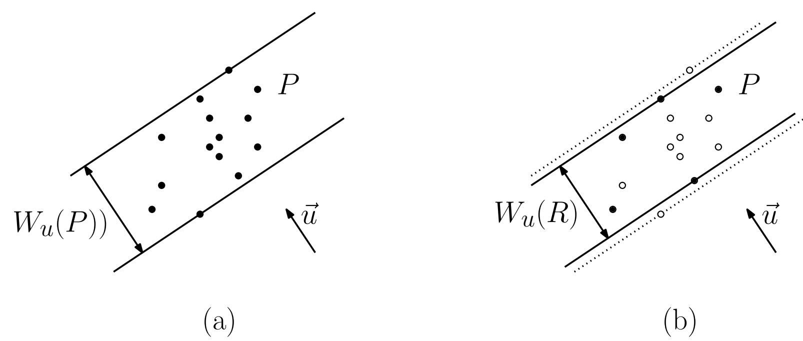

Fig. 2: Directional width and kernels. In (b) the points of R are shown as black points.

The directional width has a number of nice properties. For example, it is invariant under translation and it scales linearly if P is uniformly scaled.

Suppose we want to answer *width queries*, where we are given a vector $\vec{u}$ and we want to efficiently compute the width in this direction. We want a solution that is substantially faster than the $O(n)$ time brute force solution. We saw earlier in the semester that if P is a planar point set, then by dualizing the point set into a set P^* of lines, the vertical distance between two parallel lines that enclose P is the same as the vertical distance between two points, one on the upper hull of P^* and one on the lower hull. This observation holds in any dimension. Given the vertical width for any slope, it is possible to apply simple trigonometry to obtain the orthogonal width. The problem, however, with this approach is that the complexity of the envelopes grows as $O(n^{\lfloor d/2 \rfloor})$. Thus, a solution based on this approach would be quite inefficient (either with regard to space or query time).

Given $0 < \varepsilon < 1$, we say that a subset $R \subseteq P$ is an *ε-coreset for directional width*, also called an *ε-kernel*, if for any unit vector $\vec{u}$,

$$(1-\varepsilon)W_u(P) \leq W_u(R) \ \leq \ W_u(P).$$

Since $R \subseteq P$, the second inequality is trivially satisfied, so it is the first inequality that important. Intuitively, this says that for any unit vector $\vec{u}$, the directional width of R is smaller than that of P by a factor of only $(1-\varepsilon)$ (see Fig. 2(b)). We will show that, given an n-element point set P in $\mathbb{R}^d$, it is possible to compute an ε-kernel whose size depends only on ε and d (not on n).

Note that kernels combine nicely. In particular, it is easy to prove the following:

Chain Property: If X is an ε-kernel of Y and Y is an ε'-kernel of Z, then X is an $(\varepsilon+\varepsilon')$ kernel of Z.

Union Property: If X is an ε-kernel of P and X' is an ε-kernel of P', then $X \cup X'$ is an ε-kernel of $P \cup P'$.

Canonical Position: Let's begin by considering a very simple, but not very efficient, ε-kernel. Before giving the algorithm, it will be useful to precondition the point set by making it "fat". Let's start with some definitions. Given $\alpha \leq 1$, we say that a convex body K in $\mathbb{R}^d$ is *α-fat* if there exist two positive scalars $\lambda_- \leq \lambda_+$, such that K is sandwiched between two concentric Euclidean balls of radii λ_- and λ_+ such that $\lambda_-/\lambda_+ = \alpha$ (see Fig. 3(a)).

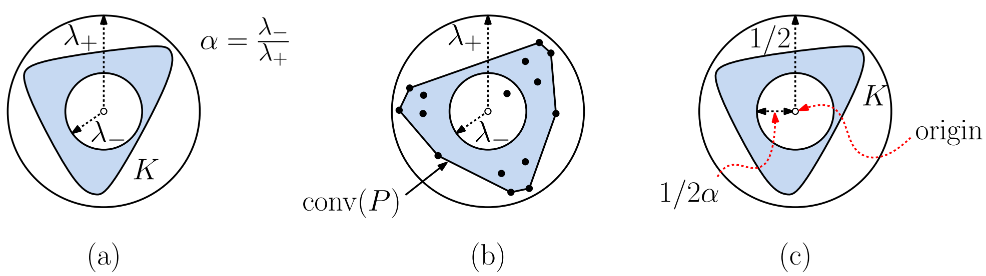

Fig. 3: The definition of α-fatness for: (a) a convex body K and (b) for a point set P.

Observe that any Euclidean ball is 1-fat. A line segment is 0-fat. It is easy to verify that a d-dimensional hypercube is $(1/\sqrt{d})$-fat. We say that a point set P is α-fat if its convex hull, conv(P), is α-fat (see Fig. 3(b)).

Once we have fattened the body, it will also be nice to scale it to a convenient size and center it about the origin. We say that a convex body K is in *α-canonical position* if it is α-fat and the two sandwiching balls are of radii $\lambda_+ = \frac{1}{2}$ and $\lambda_- = \frac{1}{2\alpha}$, respectively. The reason for chosing the outer radius to be $\frac{1}{2}$ is so that body has diameter at most 1. The following lemma makes reference to an "affine transformation". Such a transformation consists of a linear transformation (as you learned in linear algebra) combined with an arbitrary translation.

Lemma 1: Given an n-element point set $P \subset \mathbb{R}^d$, there exists an affine transformation T such that $T(P)$ is in $\frac{1}{d}$-canonical position. Further, a subset $R \subseteq P$ is an ε-kernel for P if and only if $T(R)$ is an ε-kernel for $T(P)$. The transformation T can be computed in $O(n)$ time.

Proof: (Sketch) Let $K = \text{conv}(P)$. If computation time is not an issue, it is possible to use a famous (and remarkable) fact from the theory of convexity. *John's Theorem*, states that if E is a maximum volume ellipsoid contained within K (called K's *John ellipsoid*), then K is contained within dE, where dE denotes a concentrically scaled copy of E by a factor of d (see Fig. 4(a)).

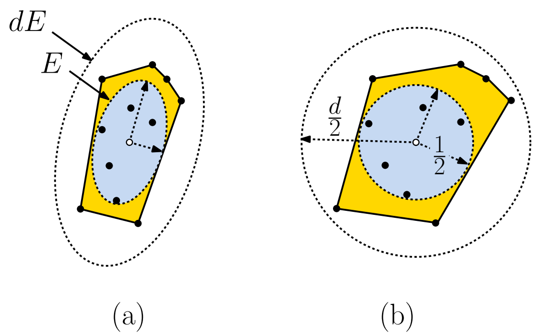

Fig. 4: The John ellipsoid and transformation to canonical position.

There is an $O(n)$ time algorithm for computing the maximum volume ellipsoid of a convex body given as the intersection of k halfspaces in any fixed dimensional space (by Chazelle and Matoušek), and there is an efficient algorithm which computes a good enough approximation to the John ellipsoid (by Barequet and Har-Peled).

To obtain the transformation T, we first translate E so that its center coincides with the origin. It is a well-known fact from linear algebra that for any origin-centered ellipsoid E, there exists an affine transformation that maps it to a unit ball. (Intuitively, this involves scaling each of the principal axes of the ellipsoid to unit length.) To complete the transformation to canonical position, we apply an additional uniform scaling by $1/2d$. Take T to be the resulting affine transformation (see Fig. 4(b)).

In the scaling process, the inner ball has radius $\frac{1}{2d}$ and the outer ball has radius $\frac{1}{2}$, and so the resulting body is in $\frac{1}{d}$-canonical position, as desired.

Generally, affine transformations do not preserve lengths, but the do preserve the ratio of lengths of vectors that the parallel to each other. The proof that ratios of directional widths are preserved will involve some technical details of linear algebra, which are beyond the scope of this course.

Quick and Dirty Kernel: Armed with the above lemma, we can compute an ε-kernel by the following "quick-and-dirty" approach. First, we assume that P has been mapped to $\frac{1}{d}$-canonical position by the above lemma. Since P is sandwiched between two balls of diameters 1 and $\frac{1}{d}$, it follows that the width of P along any direction is at most 1 and at least $\frac{1}{d}$. In order to achieve a relative error of ε, it suffices to approximate any width to an *absolute error* of at most $\frac{\varepsilon}{d}$, which we will denote by ε'.

Let $H = \left[-\frac{1}{2}, -\frac{1}{2}\right]^d$ be a hypercube of unit side length centered at the origin. Clearly, $P \subseteq H$. Subdivide H into a uniform square grid into cells of side length at most $\varepsilon'/2\sqrt{d}$, or equivalently of diameter at most $\varepsilon'/2$ (see Fig. 5(a)).

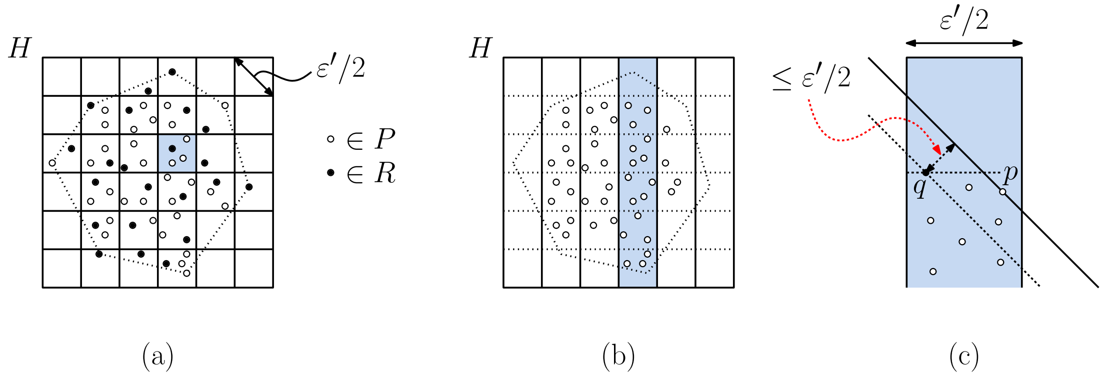

Fig. 5: (a) The quick-and-dirty kernel construction, (b) the improved construction, (c) correctness.

Each side of H is subdivided into $O(1/\varepsilon') = O(1/\varepsilon)$ intervals. Thus, the total number of cells in the grid is $O(1/\varepsilon^d)$. Next, we determine which cell of the grid each point of P belongs to. (This involves a straightforward application of integer division.) Once we know the indices of the grid square that contains a point, we can apply hashing to bucket all the points that lie within the same grid square. Since hashing takes $O(1)$ time in expectation, we can hash all the points into their grid squares in $O(n)$ expected time.

To obtain the final kernel, we select one *representative point* from each nonempty grid cell. (These are indicated using black dots in Fig. 5(a).) We output the resulting set R as the ε-kernel. Since we have at most one representative per grid square, we have $|R| = O(1/\varepsilon^d)$. The entire contstruction takes $O(n)$ time.

To establish the correctness consider any direction $\vec{u}$, and let $p_1, p_2 \in P$ be the two points of P that determine $W_u(P)$. Let $q_1, q_2 \in R$ be the corresponding representative points from these respective cells. (It may be that $q_1 = p_1$ and/or $q_2 = p_2$.) Since each cell is of diameter at most $\varepsilon'/2$, the directional width of the pair $\{q_1, q_2\}$ is smaller than the directional width of p_1 and p_2 by at most $\varepsilon'/2 + \varepsilon'/2 = \varepsilon'$. Since q_1 and q_2 are arbitrary elements of R, $W_u(R)$ is at least as large as the directional width of these two points. Because P is in canonical position, we know that $W_u(P) \geq \frac{1}{d}$. Putting this together, we have

$$\begin{aligned} W_u(P) &= W_u(\{p_1, p_2\}) \\ &\leq \varepsilon' + W_u(\{q_1, q_2\}) \;\leq\; \varepsilon' + W_u(R) \\ &= \frac{\varepsilon}{d} + W_u(R) \\ &\leq \varepsilon W_u(P) + w_u(R), \end{aligned}$$

which implies that $(1-\varepsilon)W_u(P) \leq w_u(R)$. Since $R \subseteq P$, we have $W_u(R) \leq w_u(P)$. Therefore, for any unit vector u, we have

$$(1-\varepsilon)W_u(P) \;\leq\; w_u(R) \;\leq\; w_u(P),$$

which implies that R is an ε-kernel of size $O(1/\varepsilon^d)$.

Small Improvement: It is possible make a small improvement in the size of the "quick-and-dirty" kernel. Observe from Fig. 5(a) that we may select many points from the interior of conv(P). Clearly, many of these play no useful role for the purposes of a kernel. The fix is that, rather than partition H into small hypercubes, we can instead partition it into thin columns of unit height. In particular, we create a square grid on just the upper $(d-1)$-dimensional facet of H into hypercubes of diameter at most $\varepsilon'/2$ (where ε' is the same as the previous construction). For each grid, we "extrude" it vertically into a square column of height 1. Finally, we select two representatives from each column, the topmost and bottommost points (see Fig. 5(b)). We output these representatives as the ε-kernel.

Since the top facet is of dimension $d-1$, the total number of columns in our construction is $O(1/\varepsilon^{d-1})$, and hence the total number of representatives R is also $O(1/\varepsilon^{d-1})$. (This is better than the previous construction by a factor of $1/\varepsilon$.)

To show correctness, as above we can show that for any given direction $\vec{u}$, if p is the extreme point of P along this direction, then we there exists a representative $q \in R$ (from p's column) that is within distance $\varepsilon'/2$ with respect to $\vec{u}$. We will not give a formal proof, but this follows because the diameter of the column is at most $\varepsilon'/2$. It is a relatively simple geometric exercise to prove that the highest (or lowest) point of the column has a directional distance with respect to u that is within $\varepsilon'/2$ from the true extreme point (see Fig. 5(c)). In summary, we have the following.

Theorem: Given a set P of points in $\mathbb{R}^d$ and $\varepsilon > 0$, in $O(n)$ time it is possible to construct an ε-kernel of size $O(1/\varepsilon^{d-1})$.

Optimal Kernel Construction: The above kernel construction has the advantage of simplicity, but, as shall see next, it is possible to construct much smaller kernels for directional widths. We will reduce the size from $O(1/\varepsilon^{d-1})$ to $O(1/\varepsilon^{(d-1)/2})$, thus reducing the exponential dependency by half. While we will not prove it here, it can be shown that this is asymptotically optimal (as a function of ε and d, assuming d is a constant).

Our general approach will be similar to the one taken above. First, we map P into canonical position, so it is sandwiched between balls of diameter 1 and $\frac{1}{d}$. After this, the width of P in any direction is between $\frac{1}{d}$ and 1. Thus, in order to achieve a relative error of ε, it suffices to compute a kernel whose absolute difference in width along any direction is at most $\varepsilon' = \frac{\varepsilon}{d}$.

A natural approach to solving this problem would involve uniformly sampling a large number (depending on ε) of different directions $\vec{u}$, computing the two extreme points that maximize and minimize the inner product with $\vec{u}$ and taking these to be the elements of R. It is noteworthy, that this construction does not result in the best solution. In particular, it can be shown that the angular distance between neighboring directions may need to be as small as ε, and this would lead to $O(1/\varepsilon^{d-1})$ sampled directions, which is asymptotically the same as the (small improvement to) the quick-and-dirty method. The approach that we will take is similar in spirit, but the sampling process will be based not on computing extreme points but instead on computing nearest neighbors.

We proceed as follows. Recall that P is contained within a unit ball B. Let S denote the sphere of radius 2 that is concentric with B. (The expansion factor 2 is not critical. Any constant factor expansion works, but the constants in the analysis will need to be adjusted.) Let $\delta = \sqrt{\varepsilon/4d}$. (The source of this "magic number" will become apparent later.) On the sphere S, construct a δ-dense set of points, denoted Q (see Fig. 6). This means that, for every point on S, there is a point of Q within distance δ. The surface area of S is constant, and since the sphere is a manifold of dimension $d-1$, it follows that $|Q| = O(1/\delta^{d-1}) = O(1/\varepsilon^{(d-1)/2})$. For each point of Q, compute its nearest neighbor in P.[1] Let R denote the resulting subset of P. We will show that R is the desired kernel.

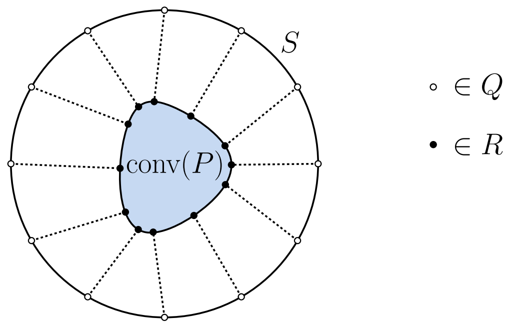

Fig. 6: Smarter kernel construction. (Technically, the points of Q are connected to the closest point of P, not conv(P).)

In the figure we have connected each point of Q to its closest point on conv(P). It is a bit easier to conceptualize the construction as sampling points from conv(P). (Recall that the kernel definition requires that the kernel is a subset of P.) There are a couple of aspects of the construction that are noteworthy. First, observe that the construction tends to sample points of P that lie close to regions where the curvature of P's convex hull is higher (see Fig. 6). This is useful, because areas of high curvature need more points to approximate them well. Also, because the points on S are chosen to be δ-dense on S, it can be shown that they will be at least this dense on P's convex hull. Before presenting the proof of correctness, we will prove a technical lemma.

Lemma 2: Let $0 < \delta \leq 1/2$, and let $q, q' \in \mathbb{R}^d$ such that $\|q\| \geq 1$ and $\|q' - q\| \leq \delta$ (see Fig. 7). Let $B(q')$ be a ball centered at q' of radius $\|q'\|$. Let $\vec{u}$ be a unit length vector from the origin to q. Then

$$\min_{p' \in B(q')} (p' \cdot \vec{u}) \;\geq\; -\delta^2.$$

Proof: (Sketch) We will prove the lemma in $\mathbb{R}^2$ and leave the generalization to $\mathbb{R}^d$ as an exercise. Let O denote the origin, and let $\ell = \|q\|$ be the distance from q to the origin. Let us assume (through a suitable rotation) that $\vec{u}$ is aligned with the x-coordinate axis. The quantity $(p' \cdot \vec{u})$ is the length of the projection of p' onto the x-axis, that is, it is just the x-coordinate of p'. We want to show that this coordinate cannot be smaller than $-\delta^2$.

[1]This clever construction was discovered in the context of polytope approximation independently by E. M. Bronstein and L. D. Ivanov, "The approximation of convex sets by polyedra," *Siber. Math J.*, 16, 1976, 852–853 and R. Dudley, "Metric entropy of some classes of sets with differentiable boundaries," *J. Appr. Th.*, 10, 1974, 227–236.

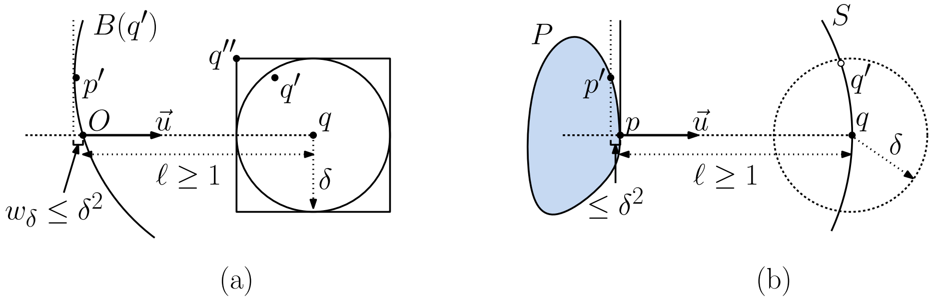

Fig. 7: Analysis of the kernel construction.

We will prove a slightly stronger version of the above. In particular, let us assume that q' is contained within a square of side length 2δ centered at q. This suffices because this square contains all points that lie within distance δ of q. Observe that the boundary of the ball $B(q')$ passes through the origin. We wish to bound how far such a ball might protrude over the $(-x)$-axis. Its easy to see that worst case arises when q' is placed in the upper left corner of the square (see Fig. 7(a)). Call this point q''.

The distance between q'' and the origin is $\ell'' = \sqrt{(\ell-\delta)^2+\delta^2}$. This is the same as the horizontal distance from q'' to the leftmost point of $B(q'')$. The horizontal distance between q'' and the origin is $\ell - \delta$. Therefore, the amount by which the ball of radius $\|q''\|$ centered at $\|q''\|$ may protrude over the $(-x)$-axis is at most w_q, which we define to be

$$w_\delta \;=\; \ell'' - (\ell-\delta) \;=\; \sqrt{(\ell-\delta)^2+\delta^2} - (\ell-\delta)$$

Since p lies in this ball, to complete the proof it suffices to show that $w_\delta \leq \delta^2$.

To simplify this, let us multiply it by a fraction whose numerator and denominator are both $\sqrt{(\ell-\delta)^2+\delta^2} + (\ell-\delta)$. Clearly, $\sqrt{(\ell-\delta)^2+\delta^2} \geq \ell-\delta$. Also, since $\ell \geq 1$ and $\delta \leq \frac{1}{2}$, we have $\ell - \delta \geq \frac{1}{2}$. Therefore,

$$\begin{aligned} w_\delta \;&=\; \frac{((\ell-\delta)^2+\delta^2)-(\ell-\delta)^2}{\sqrt{(\ell-\delta)^2+\delta^2}+(\ell-\delta)} \;=\; \frac{\delta^2}{\sqrt{(\ell-\delta)^2+\delta^2}+(\ell-\delta)} \\ &\leq\; \frac{\delta^2}{(\ell-\delta)+(\ell-\delta)} \;=\; \delta^2, \end{aligned}$$

as desired.

To establish the correctness of the construction, consider any direction $\vec{u}$. Let $p \in P$ be the point that maximizes $(p \cdot \vec{u})$. We will show that there is a point $p' \in R$ such that $(p \cdot \vec{u}) - (p' \cdot \vec{u}) \leq \varepsilon'/2$. In particular, let us translate the coordinate system so that p is at the origin, and let us rotate space so that $\vec{u}$ is horizontal (see Fig. 7(b)). Let q be the point at which the extension of $\vec{u}$ intersects the sphere S. By our construction, there exists a point $q' \in Q$ that lies within distance δ of q, that is $\|q' - q\| \leq \delta$. Let p' be the nearest neighbor of P to q'. Again, by our construction p' is in the kernel. Since q lies on a sphere of radius 2 and P is contained within the unit ball, it follows that $\|q\| \geq 1$. Thus, we satisfy the conditions of Lemma 2. Therefore,

$$(p' \cdot \vec{u}) \;\geq\; -\delta^2 \;=\; -\frac{\varepsilon}{4d} \;\geq\; -\frac{\varepsilon'}{2}.$$

Thus, the absolute error in the directional distance to p' compared to p is at most $\varepsilon'/2$. By combining the errors from both sides, the total absolute error is at most ε'. By the remarks made earlier (see also the analysis of the "quick and dirty" construction), this implies that the total relative error is ε, as desired.

Theorem: Given a set P of points in $\mathbb{R}^d$ and $\varepsilon > 0$, it is possible to construct an ε-kernel of size $O(1/\varepsilon^{(d-1)/2})$.