CMSC 754: Lecture 17
Applications of WSPDs

Reading: This material is not covered in our text. The WSPD utility lemma is from M. Smid, "The well-separated pair decomposition and its applications," (2005).

Review of WSPDs: Recall that given a parameter $s > 0$, we say that two sets of A and B are *s-well separated* if the sets can be enclosed within two Euclidean balls of radius r such that the closest distance between these spheres is at least sr (see Fig. 1(a)). Given an n-element point set P in $\mathbb{R}^d$ and *separation factor* $s > 0$, recall that an *s-well separated pair decomposition* (s-WSPD) is a collection of pairs of subsets of P $\{\{A_1, B_1\}, \{A_2, B_2\}, \ldots, \{A_m, B_m\}\}$ such that

(1) $A_i, B_i \subseteq P$, for $1 \leq i \leq m$
(2) $A_i \cap B_i = \emptyset$, for $1 \leq i \leq m$ (disjoint)
(3) $\bigcup_{i=1}^{n} A_i \otimes B_i = P \otimes P$ (cover all pairs)
(4) A_i and B_i are s-well separated, for $1 \leq i \leq m$, (well separated)

where $A \otimes B$ denotes the set of all unordered pairs from A and B. We refer to the pairs $\{A_i, B_i\}$ as *well-separated pairs* or *WSPs*.

In the previous lecture we showed that, given P and $s \geq 1$, there exists an s-WSPD of size $O(s^d n)$, which can be constructed in time $O(n \log n + s^d n)$. Because the construction concealed exponential factors depending on the dimension, we assumed that d is a constant, but we do not assume that s is a constant. (The WSPD definition allows for any $s > 0$, but we can simply apply this construction by with a modified separation factor $s' \leftarrow \max(s, 1)$.)

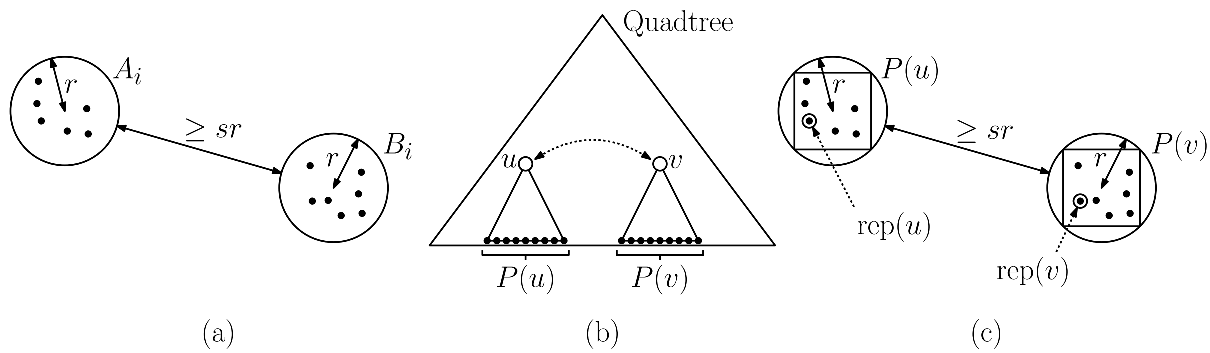

Fig. 1: WSPD and quadtree-based representation.

Quadtree-based Representation: Recall that our approach to constructing the WSPD was to first construct a *compressed quadtree* of size $O(n)$ storing the points of P in its leaves (see Fig. 1(b)). Each node u is implicitly associated with the set of points $P(u)$ stored within its subtree, and the WSPD is represented as a set of unordered pairs of nodes the quadtree. The pair $\{u, v\}$ implicitly refers to the WSP $\{P(u), P(v)\}$. Thus, the m pairs of the WSPD can be represented in $O(m)$ space. Also recall that each (nonempty) node u of the compressed quadtree is assumed to store a *representative point*, denoted rep(u). This can be chosen arbitrarily from among u's descendants (or its choice may be based on the application we have in mind).

Utility Lemma: If two sets are s-well separated, we can infer some bounds on the relative distances between point. This is encapsulated in the following useful lemma (see Fig. 2). (When reading this, it is useful to imagine that s is a large number, so quantities like $2/s$ are small.)

Lemma: (WSPD Utility Lemma) If the pair $\{P(u), P(v)\}$ is s-well separated and $x, x' \in P(u)$ and $y, y' \in P(v)$ then:

(i) $\|x - x'\| \leq \frac{2}{s} \cdot \|x - y\|$ (Pairs on the same side are close compared to opposite sides)
(ii) $\|x' - y'\| \leq \left(1 + \frac{4}{s}\right) \|x - y\|$ (Pairs on opposite sides have similar distances)

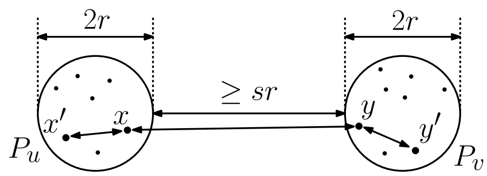

Fig. 2: WSPD Utility Lemma.

Proof: Since the pair is s-well separated, we can enclose each of $P(u)$ and $P(v)$ in a ball of radius r such that the minimum separation between these two balls is at least sr. It follows that $\max(\|x - x'\|, \|y - y'\|) \leq 2r$, and any pair from $\{x, x'\} \times \{y, y'\}$ is separated by a distance of at least sr. Thus, we have

$$\|x - x'\| \;\leq\; 2r \;=\; \frac{2r}{sr} sr \;\leq\; \frac{2r}{sr}\|x - y\| \;=\; \frac{2}{s}\|x - y\|,$$

which proves (i). Also, through an application of the triangle inequality ($\|a - c\| \leq \|a - b\| + \|b - c\|$) and the fact that $2r \leq \frac{2}{s}\|x - y\|$ we have

$$\begin{aligned} \|x' - y'\| &\leq \|x' - x\| + \|x - y\| + \|y - y'\| \;\leq\; 2r + \|x - y\| + 2r \\ &\leq \frac{2}{s}\|x - y\| + \|x - y\| + \frac{2}{s}\|x - y\| \;=\; \left(1 + \frac{4}{s}\right)\|x - y\|, \end{aligned}$$

which proves (ii).

Approximating the Diameter: The *diameter* of a point set P is defined to be the maximum distance between any pair of points of the set (see Fig. 3(a)). That is,

$$\operatorname{diam}(P) \;=\; \max_{x,y \in P} \|x - y\|.$$

The diameter can be computed exactly by brute force in $O(n^2)$ time. For points in the plane, it is possible to compute the diameter[1] in $O(n \log n)$ time. Unfortunately, generalizing this

[1]This is nontrivial, but is not much harder than a homework exercise. In particular, observe that the diameter points must lie on the convex hull. After computing the hull, it is possible to perform a rotating sweep that finds the diameter.

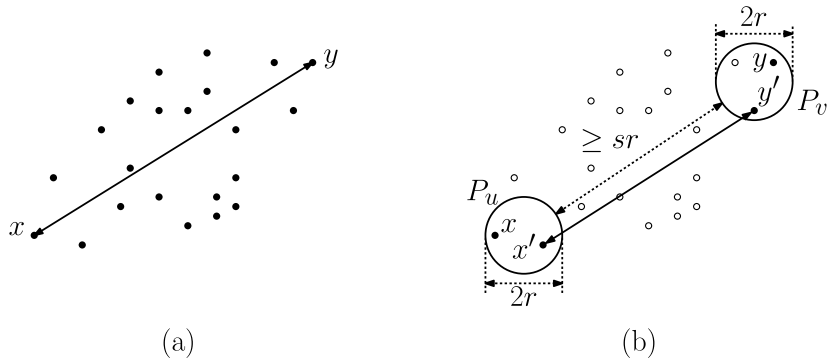

Fig. 3: Approximating the diameter.

idea to higher dimensions results in an $O(n^2)$ running time, which is no better than brute force search.

Using the WSPD construction, we can easily compute an ε-approximation to the diameter of a point set P in linear time. Given ε, we let $s = 4/\varepsilon$ and construct an s-WSPD. As mentioned above, each pair $(P(u), P(v))$ in our WSPD construction consists of the points descended from two nodes, u and v, in a compressed quadtree. Let $x' = \text{rep}(u)$ and $y' = \text{rep}(v)$ denote the representative points associated with u and v, respectively. For every well separated pair $\{P(u), P(v)\}$, we compute the distance $\|x' - y'\|$ between their representative, and output the pair achieving the largest such distance.

To prove correctness, let x and y be the points of P that realize the diameter. Let $\{P(u), P(v)\}$ be the well separated pair containing these points, and let x' and y' denote their respective representatives. By the WSPD Utility Lemma we have

$$\|x - y\| \;\leq\; \left(1 + \frac{4}{s}\right) \|x' - y'\| \;=\; (1 + \varepsilon)\|x' - y'\|.$$

Since $\{x, y\}$ is the diametrical pair, we have

$$\frac{\|x - y\|}{1 + \varepsilon} \;\leq\; \|x' - y'\| \;\leq\; \|x - y\|,$$

which implies that the output pair $\{x', y'\}$ is an ε-approximation to the diameter.[2] The running time is dominated by the time to construct the WSPD, which is $O(n \log n + s^d n) = O(n \log n + n/\varepsilon^d)$. If we treat ε as a constant, this is $O(n \log n)$.

Closest Pair (Exact!): The same sort of approach we used for the diameter (farthest pair) could be applied to produce an ε-approximation to the closest pair as well. For example, we could first compute the s-WSPD where $s \approx 1/\varepsilon$ and compute the distances between all representatives and take the pair with the smallest value. However, we can in fact do much better and obtain an exact solution. The reason is given by the following lemma.

[2] You might wonder, why we didn't select the representatives more carefully, so that x' and y' are the diametrical pair for the WSPD. Remember that there is only one representative for each node u, and there are many WSPs involving u. So, no one representative will work for all WSPs.

Lemma: Let P be any point set in $\mathbb{R}^d$ and let x and y be the closest pair of points in P. For any separation factor $s > 2$ an s-WSPD for P contains a pair $\{\{x\}, \{y\}\}$. (That is, x and y show up as singleton sets in some pair.)

Assuming this lemma for now, this yields the following simple solution. First, take any $s > 2$ (e.g., $s = 2.000001$) and compute an s-WSPD for P. Then for each WSP $\{u, v\}$, compute $\|\mathrm{rep}(u) - \mathrm{rep}(v)\|$ and return the minimum of all of these. By the above lemma, there must exist a pair where $P(u) = \{x\}$ and $P(v) = \{y\}$, and therefore, the associated representatives must be the closest pair x and y. Therefore, the algorithm will correctly find the closest pair (exactly!). Since s is a constant, the running time of the WSPD construction is $O(n \log n + s^d n) = O(n \log n)$. This is essentially optimal (in the algebraic decision tree model of computing).

To prove the above lemma, let u and v be the WSP that separated x and y. Such a pair must exist by definition of WSPDs. We assert that $P(u) = \{x\}$ and $P(v) = \{y\}$ (that is, u and v are leaves containing x and y, respectively.) Suppose to the contrary that either $P(u)$ or $P(v)$ contains another point. Without loss of generality, suppose that $P(u)$ has another point x'.

By the utility lemma and the fact that $s > 2$, we have

$$\|x - x'\| \;\leq\; \frac{2}{s} \cdot \|x - y\| \;<\; \|x - y\|,$$

but this contradicts the hypothesis that x and y are the closest pair in P. Therefore, the lemma holds.

Low-Stretch Spanners: In an earlier lecture we defined the notion of a spanner graph. Recall that any set P of n points in $\mathbb{R}^d$ defines a complete weighted graph, called the *Euclidean graph*, in which each point is a vertex, and every pair of vertices is connected by an edge whose weight is the Euclidean distance between these points. This graph is *dense*, meaning that it has $\Theta(n^2)$ edges. Intuitively, a spanner is a *sparse graph* (having only $O(n)$ edges) in which shortest paths are not significantly longer than the Euclidean distance between points. Such a graph is called a (Euclidean) *spanner*.

More formally, suppose that we are given a set P in $\mathbb{R}^d$ and a parameter $t \geq 1$, called the *stretch factor*. A *t-spanner* is any weighted graph G whose vertex set is P and, given any pair of points $x, y \in P$ we have

$$\|x - y\| \;\leq\; \delta_G(x, y) \;\leq\; t \cdot \|x - y\|,$$

where $\delta_G(x, y)$ denotes the length of the shortest path between x and y in G.

In an earlier lecture, we showed that the Delaunay triangulation of P is an $O(1)$-spanner. This was only really useful in the plane, since in dimension 3 and higher, the Delaunay triangulation can have a quadratic number of edges. Here we consider the question of how to produce a spanner in any space of constant dimension that achieves any desired stretch factor $t > 1$. There are many different ways of building spanners. Here we will discuss a straightforward method based on a WSPD of the point set.

WSPD-based Spanner Construction: Given the point set P and a (constant) stretch factor t, the idea is to build an s-WSPD for P, where s is an appropriately chosen separation factor (which will depend on t). We will then create one edge in the spanner from each well-separated pair.

Given t, we set $s = 4(t+1)/(t-1)$. (Later we will justify the mysterious choice.) For each well-separated pair $\{P(u), P(v)\}$ associated with the nodes u and v of the quadtree, let $p_u = \text{rep}(u)$ and let $p_v = \text{rep}(v)$. Add the undirected edge $\{p_u, p_v\}$ to our graph. Let G be the resulting undirected weighted graph (see Fig. 4). G will be the desired spanner. Clearly the number of edges of G is equal to the number of well-separated pairs, which is $O(s^d n) = O(n)$, and it can be built in the same $O(n \log n + s^d n) = O(n \log n)$ running time as the WSPD construction.

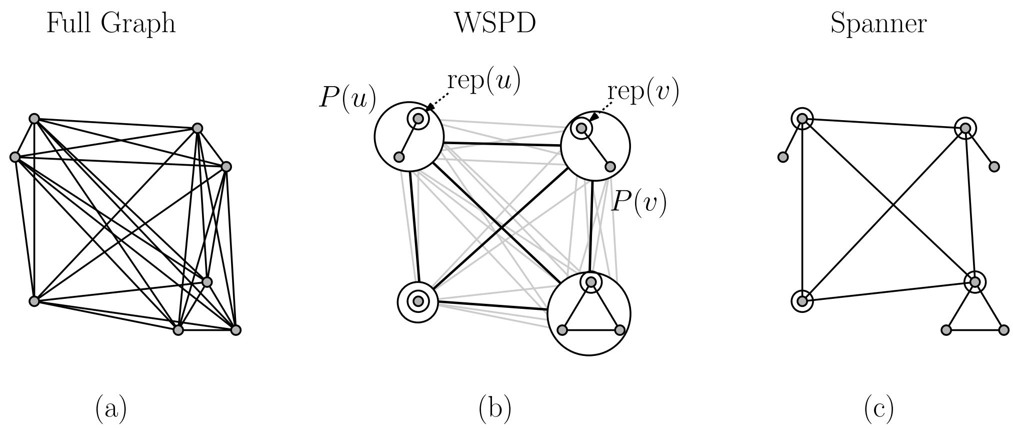

Fig. 4: WSPD-based spanner construction.

Correctness: To establish the correctness of our spanner construction algorithm, it suffices to show that for all pairs $x, y \in P$, we have

$$\|x - y\| \;\leq\; \delta_G(x, y) \;\leq\; t \cdot \|x - y\|.$$

Clearly, the first inequality holds trivially, because (by the triangle inequality) no path in any graph can be shorter than the distance between the two points. To prove the second inequality, we apply an induction based on the number of edges of the shortest path in the spanner.

For the basis case, observe that, if x and y are joined by an edge in G, then clearly $\delta_G(x, y) = \|x - y\| \leq t \cdot \|x - y\|$ for all $t \geq 1$.

If, on the other hand, there is no direct edge between x and y, we know that x and y must lie in some well-separated pair $\{P(u), P(v)\}$ defined by the pair of nodes $\{u, v\}$ in the quadtree. let $x' = \text{rep}(u)$ and $y' = \text{rep}(v)$ be the respective representative points. (It might be that $x' = x$ or $y' = y$, but not both.) Let us consider the length of the path from x to x' to y' to y. Since the edge $\{x', y'\}$ is in the graph, we have

$$\begin{aligned} \delta_G(x, y) &\leq \delta_G(x, x') + \delta_G(x', y') + \delta_G(y', y) \\ &\leq \delta_G(x, x') + \|x' - y'\| + \delta_G(y', y). \end{aligned}$$

(See Fig. 5.)

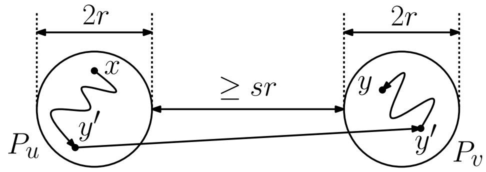

Fig. 5: Proof of the spanner bound.

The paths from x to x' and y' to y are subpaths of the full spanner path from x to y, and hence they use fewer edges. Thus, we may apply the induction hypothesis, which yields $\delta_G(x, x') \leq t\|x - x'\|$ and $\delta_G(y', y) \leq t\|y' - y\|$, yielding

$$\delta_G(x, y) \;\leq\; t\,(\|x - x'\| + \|y' - y\|) + \|x' - y'\|. \tag{1}$$

By the WSPD Utility Lemma (with $\{x, x'\}$ from one pair and $\{y, y'\}$ from the other) we have

$$\max(\|x - x'\|, \|y' - y\|) \;\leq\; \frac{2}{s} \cdot \|x - y\| \qquad \text{and} \qquad \|x' - y'\| \;\leq\; \left(1 + \frac{4}{s}\right) \|x - y\|.$$

Combining these observations with Eq. (1) we obtain

$$\delta_G(x, y) \;\leq\; t\left(2 \cdot \frac{2}{s} \cdot \|x - y\|\right) + \left(1 + \frac{4}{s}\right)\|x - y\| \;=\; \left(1 + \frac{4(t+1)}{s}\right)\|x - y\|.$$

To complete the proof, observe that it suffices to select s so that $1 + 4(t+1)/s \leq t$. Towards this end, let us set

$$s \;=\; 4\left(\frac{t+1}{t-1}\right).$$

This is well defined for any $t > 1$. By substituting in this value of s, we have

$$\delta_G(x, y) \;\leq\; \left(1 + \frac{4(t+1)}{4(t+1)/(t-1)}\right)\|x - y\| \;=\; (1 + (t-1))\|x - y\| \;=\; t \cdot \|x - y\|,$$

which completes the correctness proof.

Because we have one spanner edge for each well-separated pair, the number of edges in the spanner is $O(s^d n)$. Since spanners are most interesting for small stretch factors, let us assume that $t \leq 2$. If we express t as $t = 1 + \varepsilon$ for $\varepsilon \leq 1$, we see that the size of the spanner is

$$O(s^d n) \;=\; O\left(\left(4\frac{(1+\varepsilon)+1}{(1+\varepsilon)-1}\right)^d n\right) \;\leq\; O\left(\left(\frac{12}{\varepsilon}\right)^d n\right) \;=\; O\left(\frac{n}{\varepsilon^d}\right).$$

In conclusion, we have the following theorem:

Theorem: Given a point set P in $\mathbb{R}^d$ and $\varepsilon > 0$, a $(1+\varepsilon)$-spanner for P containing $O(n/\varepsilon^d)$ edges can be computed in time $O(n \log n + n/\varepsilon^d)$.

Approximating the Euclidean MST: The Euclidean Minimum Spanning Tree (EMST) of a point set P is the minimum spanning tree of the complete Euclidean graph on P. In an earlier lecture, we showed that the EMST is a subgraph of the Delaunay triangulation of P. This provided an $O(n \log n)$ time algorithm in the plane. Unfortunately, the generalization to higher dimensions was not interesting because the worst-case number of edges in the Delaunay triangulation is quadratic in dimensions 3 and higher.

We will now that for any constant approximation factor ε, it is possible to compute an ε-approximation to the minimum spanning tree in any constant dimension d. Given a graph G with v vertices and e edges, it is well known that the MST of G can be computed in time $O(e + v \log v)$. It follows that we can compute the EMST of a set of points in any dimension by first constructing the Euclidean graph and then computing its MST, which takes $O(n^2)$ time. To compute the approximation to the EMST, we first construct a $(1 + \varepsilon)$-spanner, call it G, and then compute and return the MST of G (see Fig. 6). This approach has an overall running time of $O(n \log n + s^d n)$.

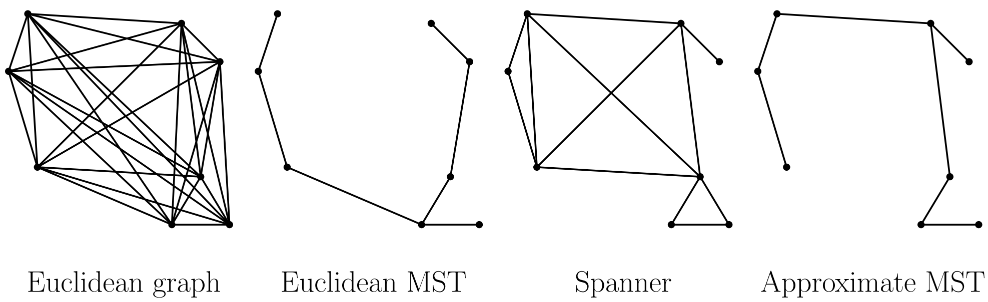

Fig. 6: Approximating the Euclidean MST.

Let EMST(P) denote P's EMST, and let emst(P) denote its total weight. Similarly, let MST(G) denote G's (exact) MST, and let mst(G) be its weight. We define a spanning subgraph of G, call it H, as follows. For each edge (x, y) in EMST(P), find the shortest path between x and y in G and add all these edges to H. Let $\delta_G(x, y)$ be the total weight of these edges, and let $w(H)$ denote the total weight of all of H's edges. By definition of the spanner, $\delta_G(x, y) \leq (1 + \varepsilon)\|x - y\|$. (Note that some of these paths may share common edges, and so the total weight of H may be smaller than the sum of the weights of these paths). Since the edges of the EMST span all the points of P, union of paths from G also spans these points. Since the MST is the spanning structure of minimum weight, we have $\mathrm{mst}(G) \leq w(H)$. Putting this all together, we obtain

$$\begin{aligned}
\mathrm{mst}(G) \;\leq\; w(H) \;\leq\; & \sum\nolimits_{(x,y)\in \mathrm{EMST}(P)} \delta_G(x, y) \\
=\; & (1 + \varepsilon) \sum\nolimits_{(x,y)\in \mathrm{EMST}(P)} \|x - y\| \\
=\; & (1 + \varepsilon) \cdot \mathrm{emst}(P),
\end{aligned}$$

as desired.