CPS296.2 Geometric Optimization **March 22, 2007**

Lecture 19: Well Separated Pair Decomposition

Lecturer: Pankaj K. Agarwal *Scribe: Sharath Kumar*

19.1 Well Separated Point Sets

Let P be a set of n points in $\mathbb{R}^d$. For any parameter $s \geq 1$, $A, B \subseteq P$ are said to be s-separated or *well separated* if A and B can be enclosed by congruent d-spheres S_A and S_B of radius r such that the minimum distance between S_A and S_B is at least sr.

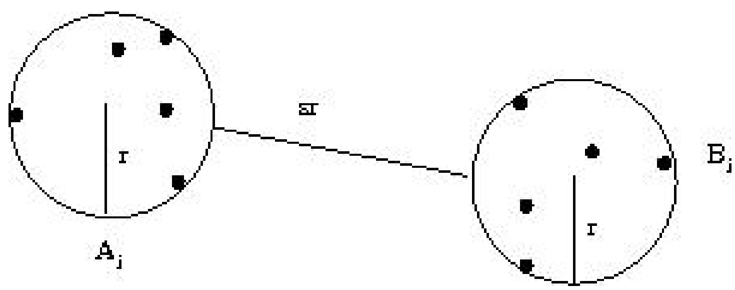

Figure 19.1: Well Separated Point Sets

There is a stronger notion of well-separated point sets. For any parameter $s \geq 1$, $A, B \subseteq P$ are said to be s-separated if the bounding boxes of A and B can be enclosed by congruent d-spheres S_A and S_B of radius r such that the minimum distance between S_A and S_B is at least sr.

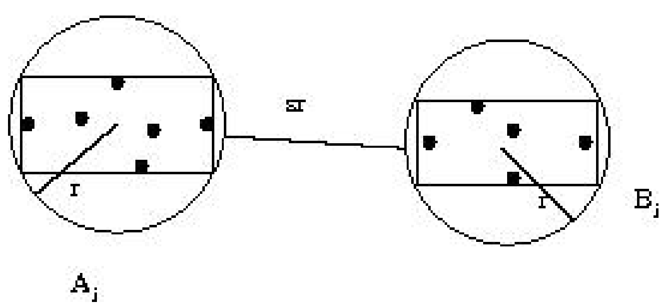

Figure 19.2: Well Separated Point Sets: A stronger notion

Let A and B be s-separated point sets and let $p, p' \in A$ and $q, q' \in B$. Since both p and p' are enclosed by a ball of radius r the maximum distance between p and p' can be at most $2r$

$$||p - p'|| \leq 2r \tag{19.1}$$

Similarly,

$$||q - q'|| \leq 2r \tag{19.2}$$

Now, since A and B are well separated, the distance from any point in A to any point in B is at least sr. Hence,

$$||p - q|| \geq sr \tag{19.3}$$

Now, from triangular inequality

$$||p' - q'|| \leq ||p' - p|| + ||p - q|| + ||q - q'||$$

From (19.1), (19.2) and (19.3)

$$||p' - q'|| \leq (1 + \frac{4}{s})\,||p - q|| \tag{19.4}$$

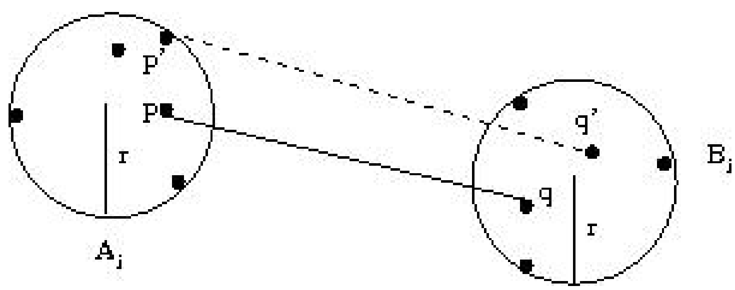

Figure 19.3: Well Separated Point sets approximate pairwise distances

19.2 Well Separated Pair Decomposition

A Well Separated Pair Decomposition of P of size m is a set of pairs $f = \{\{A_1, B_1\}, \{A_2, B_2\} \ldots, \{A_m, B_m\}\}$ such that

1. A_i and B_i are well-separated for every i
2. $\forall p, q \in P, p \neq q, \exists i$ such that
   $p \in A_i, q \in B_i$ or $p \in B_i, q \in A_i$.

Theorem 1 *[1] For $P \subseteq \mathbb{R}^d$ of n points and $s \geq 1$, a well separated pair decomposition of P of size $O(ns^d)$ can be found in time $O(n \log n + ns^d)$.*

We prove the theorem in section 19.4.

19.3 Applications of WSPD

Spanner Consider the problem of constructing sparse geoemtric graph of n points so that the shortest path distance between any two points in the graph is $(1+\epsilon)$ approximation of the Euclidean distance. Such a graph is called a $(1+\epsilon)$-spanner.

Theorem 2 *Given an n-point set $P \subset \mathbb{R}^d$, and a parameter $1 \geq \epsilon > 0$, one can compute a $(1+\epsilon)$-spanner of P with $O(n/\epsilon^d)$ edges, in time $O(n \log n + n/\epsilon^d)$.*

Proof: Consider a WSPD $F = \{\{A_1, B_1\}, \ldots \{A_m, B_m\}\}$ Arbitrarily pick $a_i \in A_i$ and $b_i \in B_i$ as representative elements of A_i and B_i for every i. Now consider the graph $G(P,E) = (P, \{(a_i, b_i) | (A_i, B_i) \in F\})$

Lemma 1 *For any $x, y \in P$, there is a path in between x and y in $G(P,E)$ of length $d_G(x,y) \leq (1+\epsilon)\,||x-y||$*

Proof: The proof is by induction on the pairs ordering according to their length.

Basis Let (r,s) be the closest pair of points. Then $(\{r\}, \{s\}) \in F$. Hence, $G(P,E)$ has an edge (r,s). This implies that $d_G(r,s) = ||r-s||$

Induction Hypothesis Fix a pair (p,q). For every pair (a,b) with $||a-b|| < ||p-q||$, let $d_G(a,b) \leq (1+\epsilon)\,||a-b||$.

Induction Step We need to show that $d_G(p,q) \leq (1+\epsilon)\,||p-q||$. There is an $(A_i, B_i) \in F$ such that $p \in A_i$ and $q \in B_i$. Thus,

$$d_G(p,q) \leq d_G(p,p') + ||p'-q'|| + d_G(q,q')$$

where p' and q' are representatives of A_i and B_i in $G(P,E)$.

From (19.4),

$$||a_i - b_i|| \leq (1 + \frac{4}{s})||p-q||$$

By induction hypothesis, p, p' and q, q' are connected in $G(P,E)$ by shorter edges, and $d_G(p, a_i) \leq (1+\epsilon)||p-p'||$, and $d_G(q, b_i) \leq (1+\epsilon)||q-q'||$. Thus,

$$d_G(p,q) \leq (1+\epsilon)||p-p'|| + (1+\frac{4}{s})||p-q|| + (1+\epsilon)||q-q'||$$

Note, however, that $||p-p'||, ||q-q'|| \leq (2/s)||p-q||$ by the WSPD properties. Hence, we can rewrite

$$d_G(p,q) \leq \frac{8}{s}||p-q|| + (1+\frac{4}{s})||p-q||$$

If we choose s sufficiently large, we get

$$d_G(p,q) \leq (1+\epsilon)||p-q||$$

■

Lemma 1 implies that $G(P, E)$ is a $(1+\epsilon)$-spanner. Since, s is $O(1/\epsilon)$, the size of the WSPD (and hence the number of edges in $G(P, E)$) is $O(n/\epsilon^d)$. Also, this WSPD (and hence $G(P, E)$) can be constructed in time $O(n \log n + n/\epsilon^d)$. ■

Minimum Spanning Tree Given a n-point set P, we can construct the Euclidean Minimum Spanning Tree of P using Well Separated Pair decomposition. Consider a WSPD $F = \{\{A_1, B_1\}, \ldots \{A_m, B_m\}\}$ of P. Pick $a_i \in A_i$ and $b_i \in B_i$ such that the pair (a_i, b_i) has the smallest distance among all possible pairs. Let a_i and b_i be the representative elements of A_i and B_i for every i. Now consider the graph $G(P, E) = (P, \{(a_i, b_i) | (A_i, B_i) \in F\})$. The graph $G(P, E)$ is of size $O(n)$ and has euclidean minimum spanning tree as its subgraph. We can obtain the EMST by running Kruskal's algorithm on $G(P, E)$. This is the fastest known algorithm for constructing the euclidean minimum spanning tree.

N-body simulation For many problems of physics, given a set P of n points it is required to compute a function

$$\forall i, \sum_{j=1, j \neq i}^{N} K(x_i, x_j)$$

where $K(x_i, x_j) \propto \frac{w_{x_i} w_{x_j}}{||x_i - x_j||^2}$. Computing the exact value requires time $O(n^2)$ since we need to compute n^2 pairwise distances. However, we can compute the approximate value of the function by approximating the pairwise distances. Consider a WSPD $F = \{\{A_1, B_1\}, \ldots \{A_m, B_m\}\}$ of P. Arbitrarily pick $a_i \in A_i$ and $b_i \in B_i$ to be the representative elements of A_i and B_i every i. Now consider the graph $G(P, E) = (P, \{(a_i, b_i) | (A_i, B_i) \in F\})$. For any pair of points (p, q), there exists an $(A_i, B_i) \in F$ such that $p \in A_i$ and $q \in B_i$. From (19.4), it follows that the weight of the edge (a_i, b_i) in $G(P, E)$ approximates the length of (p, q). Thus all the $O(n^2)$ pairs of points can be approximated by the $O(n)$ pairs corresponding to the well separated pair decomposition. This results in sub-quadratic time solutions [1] to the problems mentioned above.

19.4 Construction of WSPD

Now, we give an algorithm to construct a well separated pair decomposition of a given point set P in $\mathbb{R}^2$. Let us assume P to be uniformly distributed. Later, we show how to modify the algorithm to remove this assumption.

The algorithm essentially constructs a heirarchical partitioning of P. A simple way to do this is to construct a quadtree on the point set. Every node u in the quadtree is associated with a cell $\square_u$. Also, each node in the quadtree has four children corresponding to the four equal quadrants. We assume a binary tree representation of the quadtree where in, for any node u, the four equal quadrants of $\square_u$ are generated by first splitting $\square_u$ vertically into two equal parts and then splitting these two parts horizontally.

For any given internal node u of the quadtree,

$P_u = P \cap \square_u$,

$l_u =$ length of $\square_u$,

$R(u)$ and $L(u)$ are the right and left child of u.

Algorithm 1 $\mathbf{WSP}(u, v)$

if $\square_u, \square_v$ are well-separated **then**
  return (P_u, P_v)
else
  if $\square_u > \square_v$ **then**
    return $\mathbf{WSP}(L(u), v) \cup \mathbf{WSP}(R(u), v)$
  else
    return $\mathbf{WSP}(u, L(v)) \cup \mathbf{WSP}(u, R(v))$
  end if
end if

To generate the well separated pair decomposition of a point set, we first construct a quadtree Q on the point set. For every internal node u in Q, we call the function $\mathbf{WSP}(L(u), R(u))$. The union of all the results F would be a well separated pair decomposition.

Algorithm 2 $\mathbf{WSPD}(P)$

Construct Quadtree Q on P
for each internal node $u \in Q$ **do**
  $F = F \cup \mathbf{WSP}(R(u), L(u))$
end for
return F

Lemma 2 *F is a well separated pair decomposition of point set P.*

Proof: For any pair of points $u, v \in P$, consider the Least Common Ancestor s of u and v in the quadtree Q. Without loss of generality, let $u \in P_{L(s)}$ and $v \in P_{R(s)}$. $\mathbf{WSP}(L(s), R(s))$, invoked by Algorithm 2 would return a well separated pair $(A_i, B_i) \in F$ such that $p \in A_i$ and $q \in B_i$ or $p \in B_i$ and $q \in A_i$. Also, any pair $(A_i, B_i) \in F$ is a well-separated point set. Hence F is a well-separated pair decomposition. ■

Lemma 3 *$|F|$ is $O(n)$.*

Proof: It suffices to prove that for any internal node u, there can be at most $O(1)$ nodes v such that (u, v) would be returned as a well separated pair. $\mathbf{WSP}(u, v)$ can be invoked for one of the following reasons:

1. For some internal node t, both u and v are the right and left children. Since, $\mathbf{WSP}(L(t), R(t))$ is called for every internal node, $\mathbf{WSP}(u, v)$ would be invoked. There can be only one such v.
2. $\mathbf{WSP}(u, v)$ was invoked because we split the parent of u or the parent of v. Without loss of generality, let us assume we split the parent v' of v. Thus, $l_{v'} > l_u$ and the distance between $\square_u$ and $\square_v$ is at most $sl_{v'}$. In other words, $\square_v$ lies inside a square sq with the same center of $\square_u$ and width $(2s+3)l_{v'}$. But sq can have at most $(2s+3)^2$ disjoint squares of length $l_{v'}$. Thus the number of candidates for v' and hence v is is $O(1)$.

■

Lemma 4 *Algorithm 2 constructs a WSPD in $O(n \log n)$ time.*

Proof: It takes $O(n \log n)$ time to construct a quadtree on any given point set. For any internal node u, the function call $\mathbf{WSP}(L(u), R(u))$ takes $O(|P_u|)$ time. Now consider Q_d: the set of all internal nodes of depth d. The total time taken by the function $\mathbf{WSP}$ over all nodes in Q_d is $O\left(\sum_{v \in Q_d} |P_v|\right) = O(n)$. Given uniform distribution, the maximum depth of the quadtree is $O(\log n)$. Hence the total running time of the algorithm is $O(n \log n)$.

■

The uniform distribution assumption on the point set can be removed by storing the points in a fair-split tree instead of a quadtree. We describe the construction of a fair-split tree. The details of how to construct a well-separated pair decomposition from a fair-split tree can be read from [1]A fair-split tree on a point set $P \subset \mathbb{R}^d$ is constructed as follows.

$R(P)$: minimum axis-parallel rectangle that contains P,

$l_{max}(P)$: length of the longest edge of $R(P)$,

$l_{min}(P)$: length of the shortest edge of $R(P)$.

We construct a binary tree over the point set by recursively splitting the longest edge of the bounding rectangle of the current point set. Namely, unlike quadtree we pick in each node the dimension to use to split by and furthermore, we shrink the cell of the node of a point-set to its minimum bounding rectangle.

In fact, it is not necessary to split in the middle of the cell. Instead we can call a split *fair if it splits in the range* $1/3 \leq x \leq 2/3$.

Lemma 5 *[1] A fair split tree has linear size and depth in the worst-case. It can be computed in $O(n \log n)$ time.*

References

[1] *P. B. Callahan, S. R. Kosaraju. A decomposition of multidimensional Point sets with applications to k-nearest-neighbors and n-body potential fields.* Journal of ACM, 42, 1 1995, 67–90. 19-2, 19-4, 19-6