DCG, Wintersemester 2024/25	Lecture 18 (5 pages)
Well-Separated Pair Decomposition	
Anne Driemel and Herman Haverkort	updated: December 16, 2024

Consider a set of n points in $\mathbb{R}^d$ and let s be a real value. We are interested in representing all possible $\binom{n}{2}$ distances of this point set up to approximation factor s. We will see that for any constant s, we can do so using at most $O(n)$ distances. In addition, we will see that we can also represent a lot of structural information regarding these distances using only linear space.

1 Definition

Definition 18.1 (Well-separated pair)**.** *Let $s > 1$ be a real value and let A and B be two finite sets of points in $\mathbb{R}^d$. We say A and B are s-well-separated, if the distance between A and B is at least $(s/2)$ times the maximum diameter of A and B, that is:*

$$\min_{a\in A, b\in B} ||a-b|| \geq \frac{s}{2} \max\left(\max_{a,a'\in A} ||a-a'||, \max_{b,b'\in B} ||b-b'||\right).$$

For any $a \in A$ and $b \in B$, we say $\{A, B\}$ covers $\{a, b\}$.

Lemma 18.2. *If A and B are s-well-separated, then for any $a, a' \in A$ and $b, b' \in B$:*
(i) $\max(||a-a'||, \|b-b'\|) \leq (2/s)||a-b||$ *(distances within a set are small) and*
(ii) $||a'-b'|| \leq (1+4/s)||a-b||$ *(all distances between the sets are roughly the same).*

Proof. The first property (i) follows directly from the definition. For the second property (ii), we apply the triangle inequality as follows:

$$\begin{aligned} ||a'-b'|| &\leq ||a'-b|| + ||b-b'|| \\ &\leq ||a'-a|| + ||a-b|| + ||b-b'|| \\ &\leq (2/s)||a-b|| + ||a-b|| + (2/s)||a-b|| \end{aligned}$$

The last step follows (i). □

Definition 18.3 (WSPD)**.** *A well-separated pair decomposition (WSPD) for a set of points P with separation ratio s is a set $\{\{A_1, B_1\}, \{A_2, B_2\}, ..., \{A_m, B_m\}\}$ of s-well-separated pairs of subsets of P, such that for any two distinct points $p, q \in P$, there is at least one pair $\{A_i, B_i\}$ that covers $\{p, q\}$, that is, $p \in A_i$ and $q \in B_i$, or $p \in B_i$ and $q \in A_i$. The number of pairs m is called the size of the WSPD.*

2 Algorithm

Given a compressed quadtree on a set of points P, we can obtain a WSPD for P as follows. We say two nodes u and v are well-separated if $\mathrm{Cube}(u)$ and $\mathrm{Cube}(v)$ are well-separated. On each pair of siblings u and v, we run the following algorithm to generate a (non-exhaustive) set of well-separated pairs between nodes in the subtrees rooted at u and v, respectively: if $\mathrm{Cube}(u)$ and $\mathrm{Cube}(v)$ are well-separated, we report the pair $\{u, v\}$; otherwise, assume, without loss of generality, that the diameter $\mathrm{diam}(\mathrm{Cube}(u))$ of $\mathrm{Cube}(u)$ is at most the diameter $\mathrm{diam}(\mathrm{Cube}(v))$ of $\mathrm{Cube}(v)$, and we generate well-separated pairs between u and each child of v recursively. Finally, we interpret each pair $\{u, v\}$ of nodes reported as the pair $\{\mathrm{Cube}(u) \cap P, \mathrm{Cube}(v) \cap P\}$.

The following is the same algorithm in pseudocode.

Procedure AddWellSeparatedPairs(compressed quadtree nodes u and v, $s \in \mathbb{R}$)

```
1 if Cube(u) and Cube(v) are s-well-separated then
2     output the unordered pair of nodes {u, v};
3 else
4     if diam(Cube(u)) > diam(Cube(v)) then
5         swap u and v
6     let v_1, ..., v_k be the children of v;
7     for i ← 1 to k do
8         AddWellSeparatedPairs(u, v_i, s)
```

Procedure WellSeparatedPairDecomposition(compressed quadtree root u, $s \in \mathbb{R}$)

```
1 if u is a leaf, then return;
2 let u_1, ..., u_k be the children of u;
3 for i ← 1 to k do
4     for j ← i + 1 to k do
5         AddWellSeparatedPairs(u_i, u_j, s)
6 for i ← 1 to k do
7     WellSeparatedPairDecomposition (u_i, s)
8 return;
```

3 Analysis

Theorem 18.4. *The algorithm computes a well-separated pair decomposition of P with separation ratio s.*

Proof. Any pair output by the algorithm is an s-well-separated pair of the bounding boxes of the points stored in two different subtrees of the quadtree. We claim that any pair of distinct p, q in P is covered by some pair in the output.

Fist, we claim the following:

> a call to `AddWellSeparatedPairs` on two disjoint subtrees rooted at u and v results in a set of pairs D' output by the algorithm, such that for any two distinct points $p \in \text{Cube}(u) \cap P$ and $q \in \text{Cube}(v) \cap P$, there is a well-separated pair in D' that covers (p, q).

We prove this claim by induction on the recursive calls of the procedure, where the induction is on the height of a call in the tree of recursive calls. The base case of the induction is a call with height zero, that is, when no further recursive calls are made because u and v satisfy the condition in line 1, so they form a well-separated pair. (Note hat this is true if both nodes are leaves of the compressed quadtree, but it may also happen earlier, before the recursion reaches the leaves of the compressed quadtree). Now we take as the induction hypothesis that our claim is true for calls with height at most h, and we consider a recursive call of height $h + 1$ on two nodes u and v that do not satisfy the condition in line 1. Then the algorithm recurses on all children of the larger node, while keeping the smaller node fixed. These recursive calls have height at most h. Let u be the smaller node and let v be the larger node. For $q \in \mathrm{Cube}(v) \cap P$, it holds that q is contained in one of the children of v, say v_i. By the induction hypothesis, the claim is true for the recursive call on the pair of nodes u and v_i. Therefore, $\{p, q\}$ is covered by the output of the algorithm. This proves the claim for the recursive call on u and v, and thus, by induction, for any call of any height on any two disjoint subtrees.

Secondly, we claim that a call to `WellSeparatedPairDecomposition` on the root of a quadtree u computes a WSPD of the set P stored in the quadtree. Again, we show this by induction. The base case is when u is a leaf node. In this case, the WSPD is empty and nothing is output and this is a correct result. Now, assume that u is not a leaf node. In this case, the algorithm calls `AddWellSeparatedPairs` on any combination of children of u, thereby covering all pairs of points p, q of P stored in two different subtrees. This follows by the above analysis of `AddWellSeparatedPairs`. In addition, the algorithm calls itself recursively on each of the children of u, thereby covering all pairs of points p, q of P stored in the same subtree. This follows by induction. Now, for any two distinct points stored in the quadtree rooted at u, it must be that they are either stored in two different subtrees, or in the same subtree rooted at a child of u. $\square$

Lemma 18.5. *For any two distinct points $p, q \in P$, the algorithm outputs exactly one s-well-separated pair that covers $\{p, q\}$. (Proof $\rightarrow$ Exercise)*

What is the size of the WSPD, that is, how many pairs of nodes are output by the algorithm? We first show a lemma that will allow us to apply a packing argument to bound the number of pairs.

Lemma 18.6. *Let $\{u, v\}$ be a pair that is output by the algorithm. One of the following holds*

(i) the pair $\{\mathrm{Cube}(\mathrm{parent}(u)), \mathrm{Cube}(v)\}$ is not well-separated and

$$\mathrm{diam}(\mathrm{Cube}(v)) \leq \mathrm{diam}(\mathrm{Cube}(\mathrm{parent}(u))) \leq \mathrm{diam}(\mathrm{Cube}(\mathrm{parent}(v))).$$

(ii) the pair $\{\mathrm{Cube}(\mathrm{parent}(v)), \mathrm{Cube}(u)\}$ is not well-separated and

$$\mathrm{diam}(\mathrm{Cube}(u)) \leq \mathrm{diam}(\mathrm{Cube}(\mathrm{parent}(v))) \leq \mathrm{diam}(\mathrm{Cube}(\mathrm{parent}(u))).$$

Proof. If u and v are siblings, that is, if $\mathrm{parent}(u) = \mathrm{parent}(v)$, then the claim is trivially true, since any node's bounding box diameter is smaller than its parent's.

Otherwise, the pair $\{u, v\}$ must have been created by calling `AddWellSeparatedPairs` on u and v while running `AddWellSeparatedPairs` either (a) on $\mathrm{parent}(u)$ and v, or (b) on u and $\mathrm{parent}(v)$.

Assume it is case (a), then we claim that (i) from the lemma statement holds. We show this as follows. In case (a) we have

$$\mathrm{diam}(\mathrm{Cube}(v)) \leq \mathrm{diam}(\mathrm{Cube}(\mathrm{parent}(u))),$$

since we always recurse on the larger node (see lines 4 and 5 of the algorithm).

Now trace the recursive calls back to the call that introduced v as an argument, that is, the first call of `AddWellSeparatedPairs` on v and an ancestor u' of $\mathrm{parent}(u)$. There are two cases:

(a.i) this call was made by `WellSeparatedPairDecomposition`, so u' and v are siblings, and thus,

$$\mathrm{diam}(\mathrm{Cube}(\mathrm{parent}(u))) \leq \mathrm{diam}(\mathrm{Cube}(u')) \leq \mathrm{diam}(\mathrm{Cube}(\mathrm{parent}(u'))) = \mathrm{diam}(\mathrm{Cube}(\mathrm{parent}(v))).$$

(a.ii) the call was made by `AddWellSeparatedPairs` with u' and parent(v) as arguments, in which case we also have

$$\mathrm{diam}(\mathrm{Cube}(\mathrm{parent}(u)) \leq \mathrm{diam}(\mathrm{Cube}(u')) \leq \mathrm{diam}(\mathrm{Cube}(\mathrm{parent}(v))).$$

Now, assume it is case (b), then we claim that (ii) of the lemma statement holds. The argument is symmetric to the above. □

Definition 18.7 (Ordered pairs)**.** *We define an ordering on each pair $\{x, y\}$ output by the algorithm as follows. If part (i) of Lemma 18.6 applies with $u = x$ and $v = y$, then we consider the pair to be ordered as (x, y), otherwise we consider the pair to be ordered as (y, x).*

Note that in the latter case, part (ii) of Lemma 18.6 applies with $u = x$ and $v = y$, or to put it differently: part (i) of Lemma 18.6 applies with $u = y$ and $v = x$.

Theorem 18.8. *Let P be a set of n points stored in a compressed quadtree. The algorithm outputs a WSPD of P of size in $O(s^d n)$. The separation ratio of the WSPD is s. The running time of the algorithm is bounded by $O(s^d n)$.*

Proof. By Theorem 18.4 the output of the algorithm is a WSPD of the points stored in the compressed quadtree. It remains to prove the bound on the size of the WSPD. To this end, we count all ordered pairs in the WSPD, by charging each pair to the first node in the ordered tuple. Let u be a fixed node of the tree. We claim that there are at most $O(s^d)$ ordered pairs that have u at the first position. By the definition of ordered pairs, Lemma 18.6(i) holds for the ordered pair (u, v). This implies that for $A = \mathrm{Cube}(\mathrm{parent}(u))$ and $B = \mathrm{Cube}(v)$, it holds that

$$\min_{a \in A, b \in B} \|a - b\| < (s/2) \max(\mathrm{diam}(A), \mathrm{diam}(B))$$

since A and B are not well-separated by Lemma 18.6. Also, by Lemma 18.6(i) we have that $\mathrm{diam}(A) \geq \mathrm{diam}(B)$ and therefore the above implies that

$$\min_{a \in A, b \in B} \|a - b\| < (s/2)\,\mathrm{diam}(A) \tag{1}$$

Let C_v be the largest canonical subcube of Cube(parent(v)) that contains Cube(v). By definition, C_v has diameter

$$\mathrm{diam}(C_v) = \mathrm{diam}(\mathrm{Cube}(\mathrm{parent}(v)))/2$$

By Lemma 18.6(i), it follows that

$$\mathrm{diam}(C_v) \geq \mathrm{diam}(\mathrm{Cube}(\mathrm{parent}(u)))/2 = \mathrm{diam}(A)/2,$$

Therefore, C_v covers a cube C'_v of diameter $\mathrm{diam}(A)/2$ that includes the point $b \in \mathrm{Cube}(v)$ that lies closest to A, and thus, by Equation 1, C'_v lies entirely within the set U of points that lie within distance $(s/2 + 1/2)\,\mathrm{diam}(A)$ from A, that is:

$$U = \left\{ b \in \mathbb{R}^d \mid \exists a \in A : \|a - b\| \leq (s/2 + 1/2)\,\mathrm{diam}(A) \right\}$$

Furthermore, the regions C_v associated with the pairs (u, v) for a fixed u must all be disjoint from each other (otherwise, there would a pair of points in P that is covered by more than one well-separated pair, contradicting Lemma 18.5). Now we can apply a packing argument: the number of regions C_v for different v that intersect U is at most the volume of U divided by the volume of a cube of diameter $\mathrm{diam}(A)/2$. Thus, their number is at most:

$$\frac{\mathrm{vol}(U)}{(\mathrm{diam}(A)/2\sqrt{d})^d} \leq \frac{((s+2)\,\mathrm{diam}(A))^d}{(\mathrm{diam}(A)/2\sqrt{d})^d} = \left((2s+4)\sqrt{d}\right)^d = O(s^d).$$

This proves that for any fixed node u, the number of ordered pairs that have u at the first position is bounded by $O(s^d)$. By the analysis in Lecture 17, a compressed quadtree of a set of n points has $O(n)$ nodes. Thus, with $O(s^d)$ ordered pairs per node, the total number of pairs output by the algorithm is in $O(s^d n)$.

Now, consider the running time of the algorithm. The recursive calls to `WellSeparatedPairDecomposition` and `AddWellSeparatedPairs` form a tree of outdegree at least 2 (each recursive call produces at least two other recursive calls). Each leaf of the tree outputs a pair of the WSPD. Therefore, the total number of calls to `AddWellSeparatedPairs` is less than twice the size of the WSPD, which is in $O(s^d n)$. Assuming that the compressed quadtree stores $\mathrm{Cube}(u)$ with every internal node u, it takes constant time (for constant d) to determine for any pair of bounding boxes, whether they are s-well-separated. Therefore, the total running time of `WellSeparatedPairDecomposition` is in $O(s^d n)$. □

References

- Giri Narasimhan and Michiel Smid: *Geometric spanner networks*, Cambridge University Press, 2007.
- Sariel Har-Peled: *Geometric Approximation Algorithms*, Mathematical Surveys and Monographs, Volume 173, American Mathematical Society, 2011.