Locality-Sensitive Orderings

Anil Maheshwari

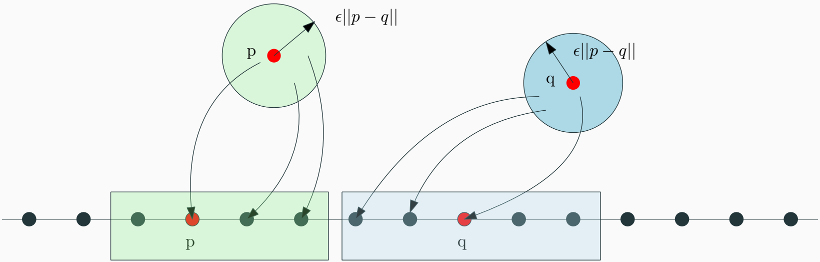
anil@scs.carleton.ca
School of Computer Science
Carleton University
Canada

Main Result

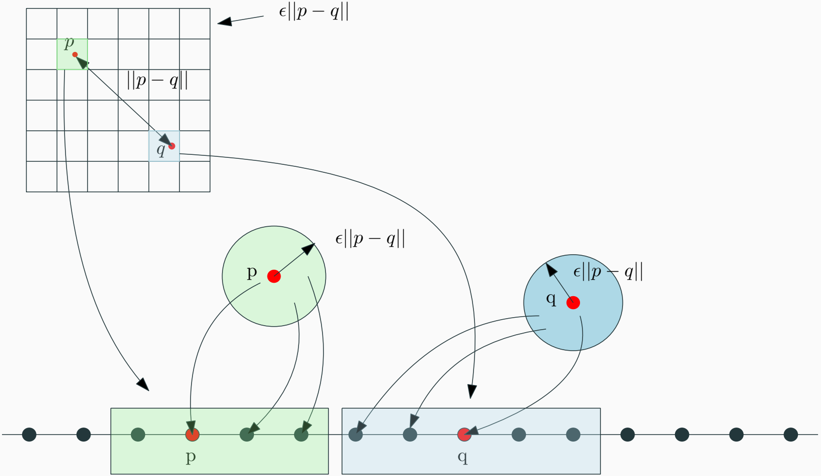
What we want to do?

Local Ordering Theorem (CHJ2020)

Consider a unit cube in d-dimensions. For any $\epsilon \in (0, \frac{1}{2}]$, there is a family of $O(\frac{1}{\epsilon^d}\log(\frac{1}{\epsilon}))$ orderings of $[0,1)^d$ such that for any pair of points $p, q \in [0,1)^d$, there is an ordering in the family where all the points between p and q are within a distance of at most $\epsilon||p-q||_2$ from p or q.



An Illustration



Properties of Orderings

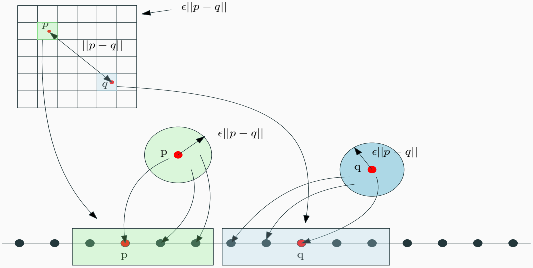
Let p and q be two points in a unit-cube.

Partition the unit-cube into sub-cubes of dimension $\epsilon||p-q||_2$.

Let C_p and C_q be two sub-cubes such that $p \in C_p$ and $q \in C_q$.

$\exists$ an ordering σ among $O(\frac{1}{\epsilon^d}\log(\frac{1}{\epsilon}))$ orderings such that

1. Points in C_p and C_q are mapped to two intervals on the real line by the ordering σ
2. These two intervals are adjacent.



Applications

An alternative to Quadtrees with applications in

- Dynamic approximate bichromatic closest pair
- dynamic spanners
- dynamic approximate Euclidean Minimum Spanning Trees
- Approximate Nearest Neighbors
- ...

Application: Closest Pair

Input: Set of n points in $\Re^d$
Output: Closest pair

Algorithm:

1. For every ordering, find the pair of consecutive points that have minimum distance.
2. Report the pair that has the least distance among all the orderings.

Time: $O(n \times \text{\# of orderings}) \approx O(n/\epsilon^d)$

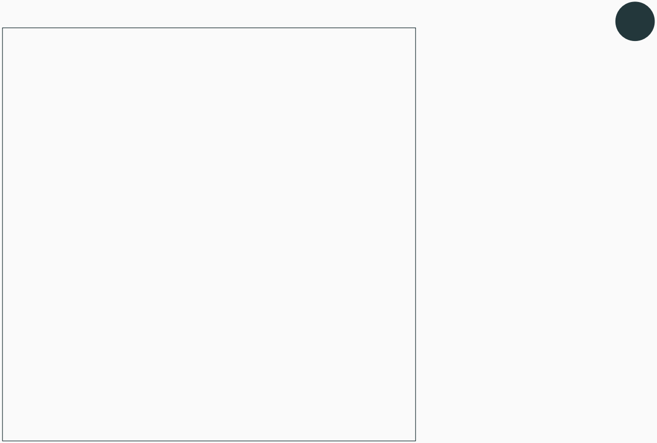
Dynamic: Insert/Delete points and maintain orderings (and hence the closest pair)

Old & New Concepts

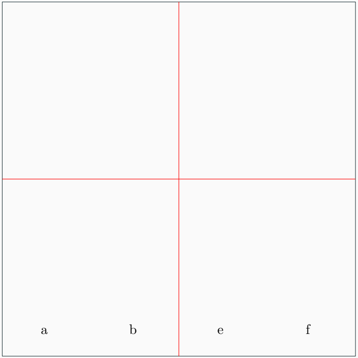
1. Quadtree.
2. Linear orderings of points in a Quadtree.
3. Shifted Quadtrees and ANN.
4. Quadtree as union of ϵ-Quadtrees.
5. (Wonderful) Walecki Construction from 19th Century.
6. Locality-Sensitive Orderings.
7. Applications in ANN, Bi-chromatic ANN, Spanners, ...

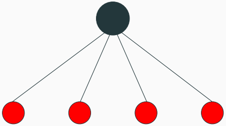
Quadtree

Quadtree of a point set

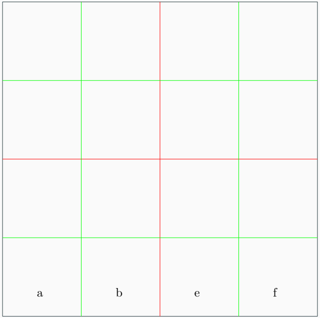


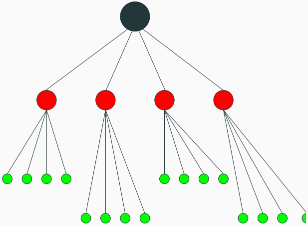
Quadtree of a point set



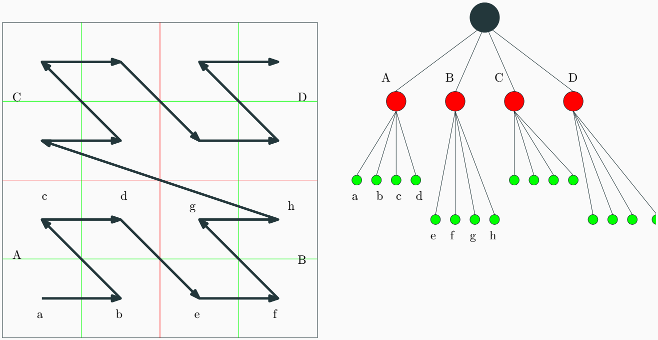


Quadtree of a point set



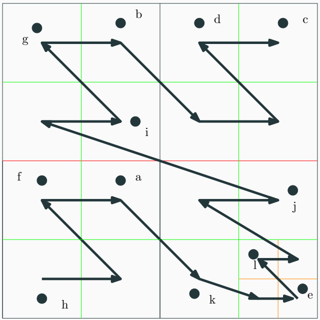


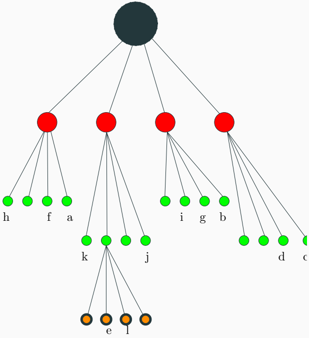
Quadtree of a point set



DFS traversal of Quadtree

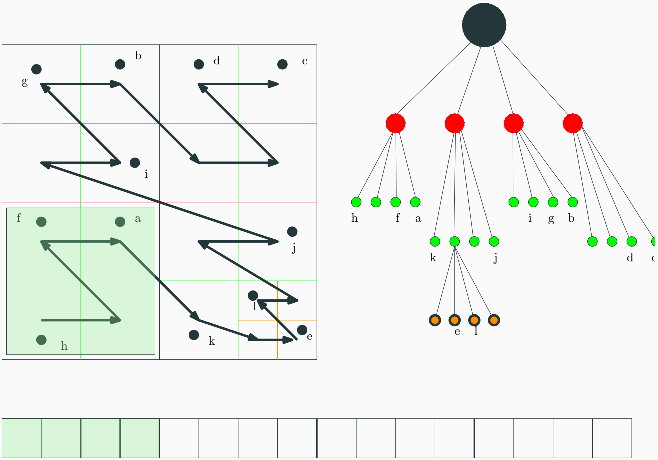
Obtain a linear order of points by performing the DFS traversal of the Quadtree.



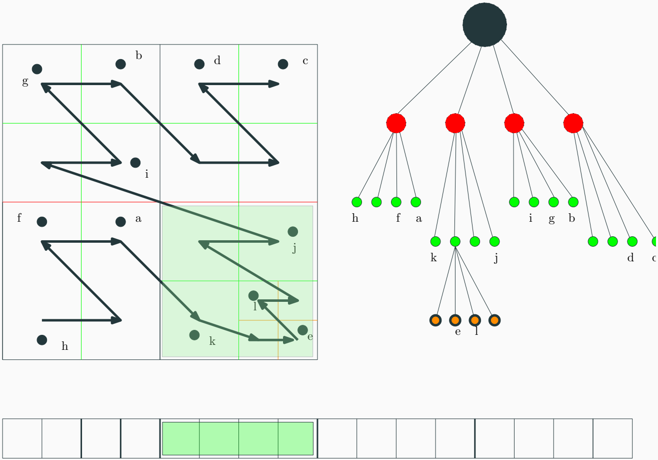


h	f	a	k	e	l	j	i	g	b	d	c

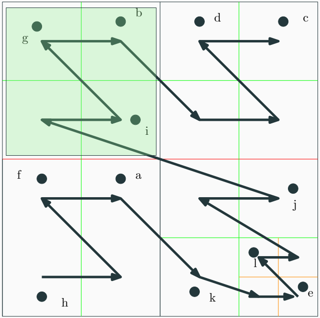
Quadtree Cells & DFS order

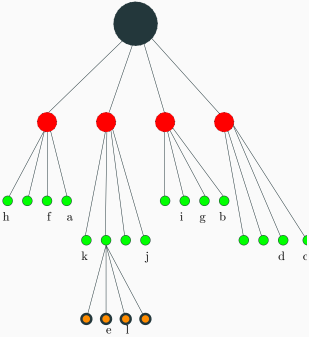


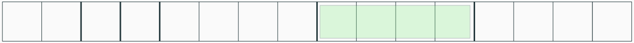
Quadtree Cells & DFS order



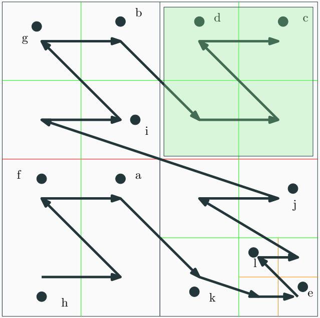
Quadtree Cells & DFS order

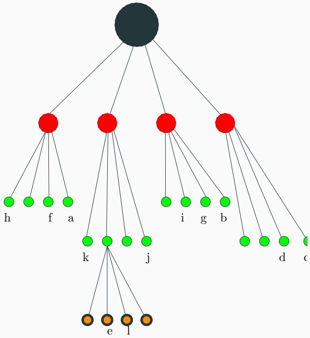




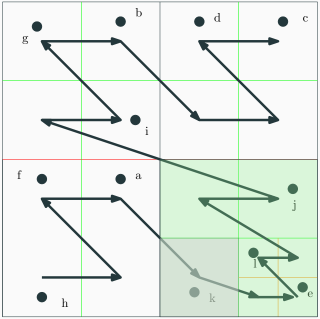


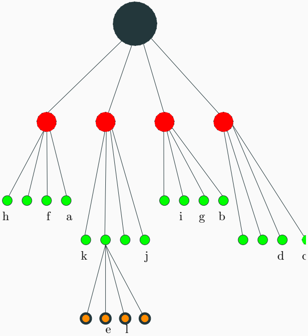
Quadtree Cells & DFS order

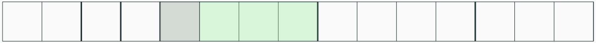




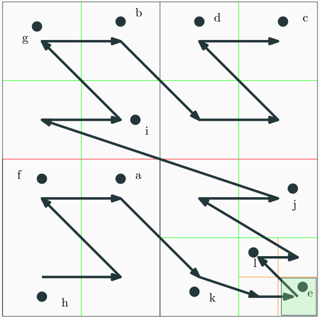
Quadtree Cells & DFS order

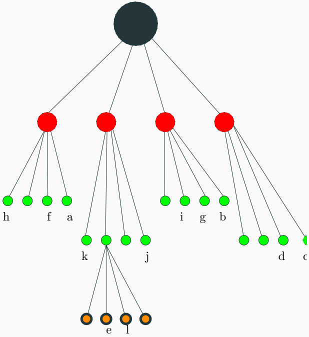






Quadtree Cells & DFS order

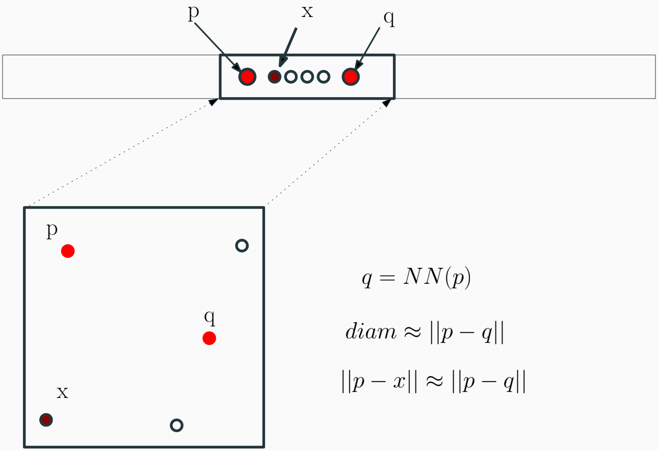




ANN

Approximate NN

Let q be nearest-neighbor of p. Assume that there is a cell containing p and q in Quadtree with diameter $\approx ||p-q||$.



$$q = NN(p)$$

$$diam \approx ||p-q||$$

$$||p-x|| \approx ||p-q||$$

Assumption?

How to ensure that the following is true?

There is a cell containing p and q in the Quadtree with diameter $\approx ||p - q||$

Quadtrees of Shifted Point Sets

Assume all points in $P \subset [0,1)^d$.
Construct $D = 2\lceil \frac{d}{2} \rceil + 1$ copies of P.

Shifted Point Sets

For $i = 0, \ldots, D$, define shifted point sets

$P_i = \{p_j + (\frac{i}{D+1}, \frac{i}{D+1}, \ldots, \frac{i}{D+1}) | \forall p_j \in P\}$

Let Quadtrees of $P_0, P_1, \ldots, P_D$ be $T_0, T_1, \ldots, T_D$.

Chan (DCG98)

For any pair of points $p, q \in P$, there exists a Quadtree $T \in \{T_0, T_1, \ldots, T_D\}$ such that the cell containing p, q in T has diameter $c||p - q||$ (for some constant $c \geq 1$).

Dynamic ANN

Chan's ANN Algorithm:

1. Construct linear (dfs) order for each of the Quadtrees $T_0, T_1, \ldots, T_D$.
2. For each point p, find its neighbor in each of the linear orders that minimizes the distance.
3. Let q be the neighbor of p with the minimum distance.
4. Report q as the ANN of p.

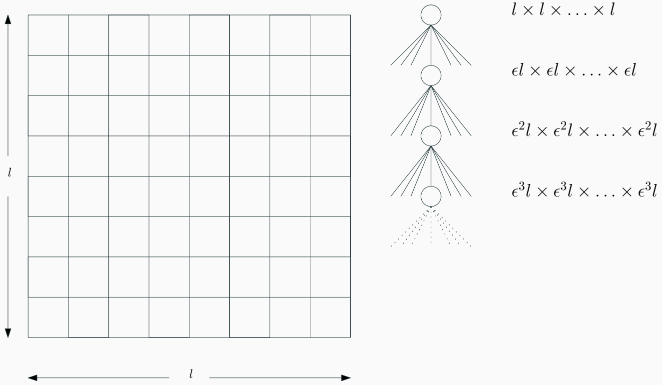
Chan (1998, 2006)

For fixed dimension d, in $O(n \log n)$ preprocessing time and $O(n)$ space, we can find a c-approximate nearest neighbor of any point in P in $O(\log n)$ time ($c = f(d)$).

ϵ-Quadtree

ϵ-Quadtree

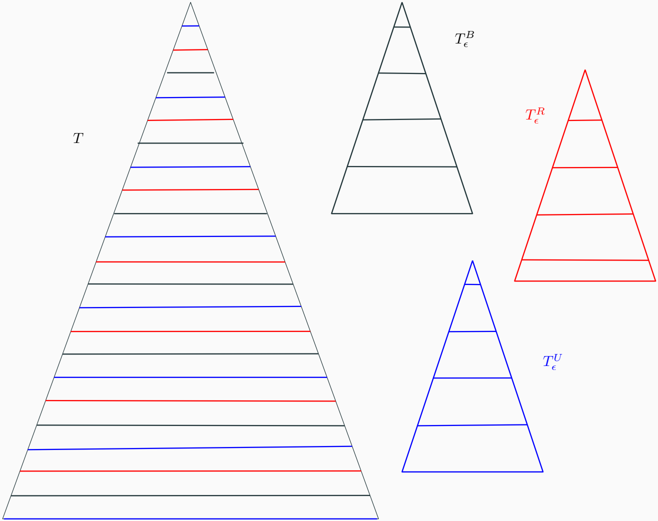
For a constant $\epsilon > 0$, recursively partition a cube $[0,1)^d$ evenly into $\frac{1}{\epsilon^d}$ sub-cubes ($\epsilon = 1/2 \implies$ Standard Quadtree).



Quadtree as union of ϵ-Quadtrees

Partitioning a Quadtree T into $\log \frac{1}{\epsilon}$ ϵ-Quadtrees

Let $\epsilon = 2^{-3}$. $T = T_\epsilon^B \cup T_\epsilon^R \cup T_\epsilon^U$.



Walecki Theorem

Walecki's Result

Permuting cells of a node of an ϵ-Quadtree

Let $\epsilon = 2^{-2}$. Any two cells are neighbors in at least one of the 8 permutations.

A	B	C	D
E	F	G	H
I	J	K	L
M	N	O	P

ABPCODNEMFLGKHJI

BCADPEOFNGMHLIKJ

CDBEAFPGOHNIMJLK

DECFBGAHPIOJNKML

EFDGCHBIAJPKOLNM

FGEHDICJBKALPMON

GHFIEJDKCLBMANPO

HIGJFKELDMCNBOAP

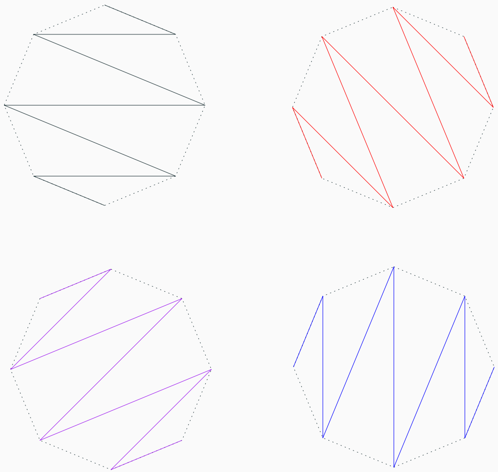
(Wonderful) Walecki Result

Walecki Theorem

For n elements $\{0, 1, 2, \ldots, n-1\}$, there is a set of $\lceil \frac{n}{2} \rceil$ permutations of the elements, such that, for all $i, j \in \{1, 2, \ldots, n-1\}$, there is a permutation in which i and j are adjacent.

Proof (by Figure)

Partition K_8 in 4 Hamiltonian Paths



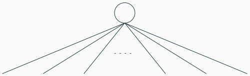
Linear orders of points of $P \subset [0,1)^d$

DFS Traversal of an ϵ-Quadtree T_ϵ

1. $\#$children of any node of $T_\epsilon = O(1/\epsilon^d)$.
2. Construct $O(1/\epsilon^d)$ permutations of cells using Walecki's construction.
3. Generate $O(1/\epsilon^d)$ linear orders of points in P by performing DFS traversal of T_ϵ with respect to each permutation.

Structure of Cells

A	B	C	D
E	F	G	H
I	J	K	L
M	N	O	P



A	B	P	C	O	D	N	E	M	F	L	G	K	H	J	I

What have we learnt so far?

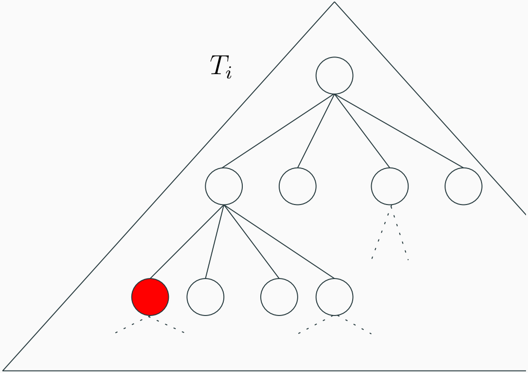
1. Point set $P \subset [0,1)^d$.
2. Shifted points sets $P_0, P_1, \dots, P_D$ and their Quadtrees $T_0, T_1, \dots, T_D$.
3. Each Quadtree T_i partitioned into $\log \frac{1}{\epsilon}$ ϵ-Quadtrees.
4. Permutations of cells of a node in an ϵ-Quadtree (Walecki's result).
5. Linear orders of points in P from DFS (for each permutation) of ϵ-Quadtrees.
6. Total $\#$Linear Orders $= O(D \times \log \frac{1}{\epsilon} \times \frac{1}{\epsilon^d}) = O(\frac{1}{\epsilon^d} \log \frac{1}{\epsilon})$.
7. These linear orders satisfy the “locality” condition.

Local-Sensitivity Theorem

Locality Property

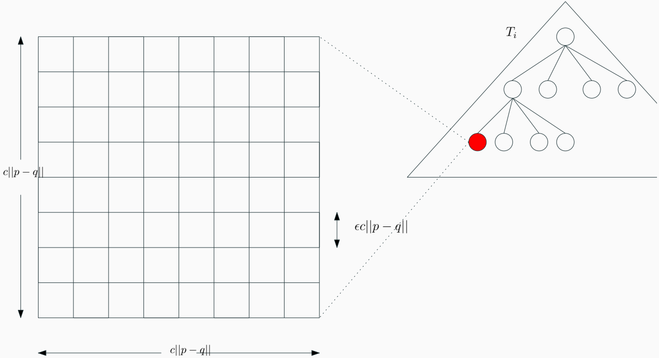
Locality-Sensitive Orderings

Let the Quadtree $T_i \in \{T_0, T_1, \ldots, T_D\}$ has a cell containing p and q with diameter $\approx ||p - q||$.



Locality-Sensitive Orderings

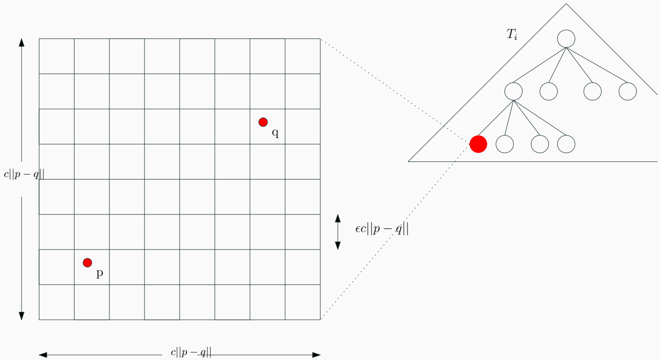
Let the Quadtree $T_i \in \{T_0, T_1, \ldots, T_D\}$ has a cell containing p and q with diameter $\approx ||p - q||$.



Locality Property

Locality-Sensitive Orderings

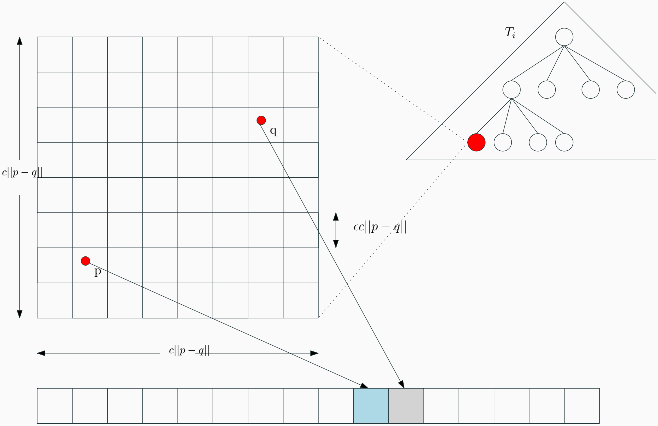
Let the Quadtree $T_i \in \{T_0, T_1, \ldots, T_D\}$ has a cell containing p and q with diameter $\approx ||p - q||$.



Locality Property

Locality-Sensitive Orderings

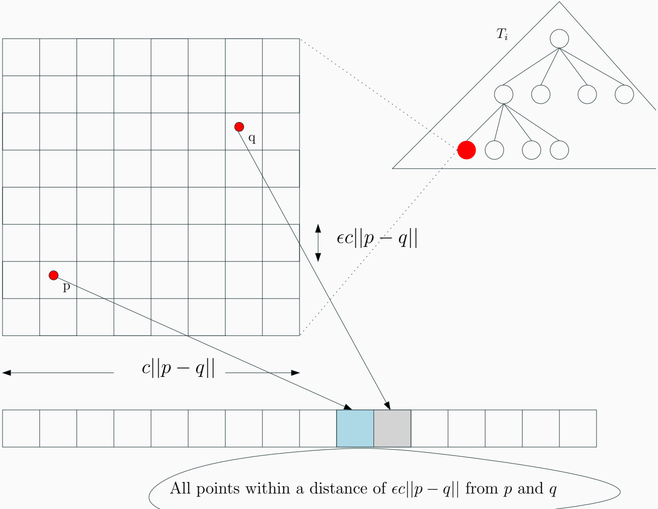
Let the Quadtree $T_i \in \{T_0, T_1, \ldots, T_D\}$ has a cell containing p and q with diameter $\approx ||p - q||$.



Locality Property

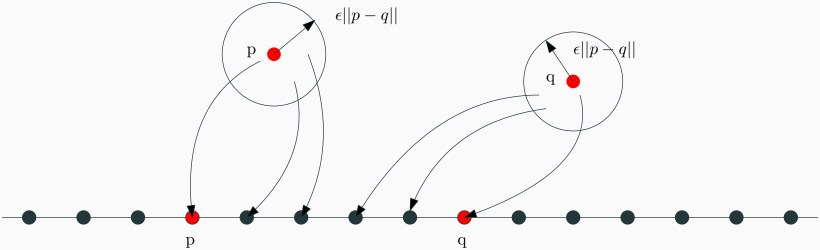
Locality-Sensitive Orderings

Let the Quadtree $T_i \in \{T_0, T_1, \ldots, T_D\}$ has a cell containing p and q with diameter $\approx ||p - q||$.



(CHJ 2019)

Consider a unit cube $[0,1)^d$. For $\epsilon > 0$, there is a family of $O(\frac{1}{\epsilon^d}\log\frac{1}{\epsilon})$ orderings of $[0,1)^d$ such that for any $p, q \in [0,1)^d$, there is an ordering in the family where all the points between p and q are within a distance of at most $\epsilon||p-q||_2$ from p or q.



Applications

Application Domain

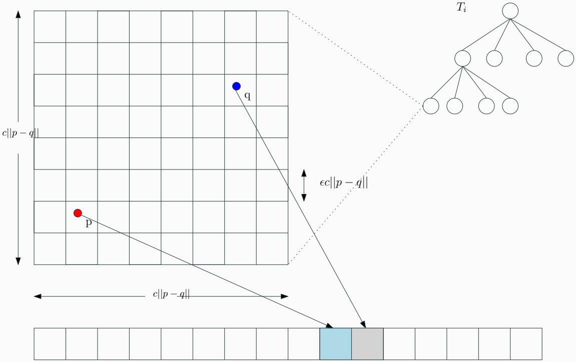
Consider problems where you may need to consider pairwise distances between points such as

1. Closest Pair
2. Nearest Neighbour of each point
3. MST
4. Sparse Spanners
5. & Updates
6. ...

Key Idea: Computation on sparse graph formed by joining adjacent points in linear orders rather than the complete graph

Approximate Bichromatic NN

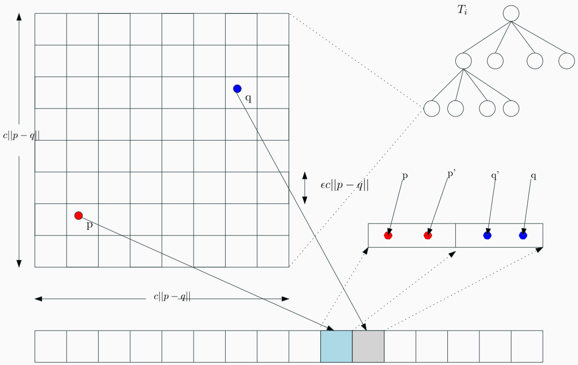
Let p and q constitute a red-blue Nearest Neighbor of the point set.



Bichromatic NN

Approximate Bichromatic NN

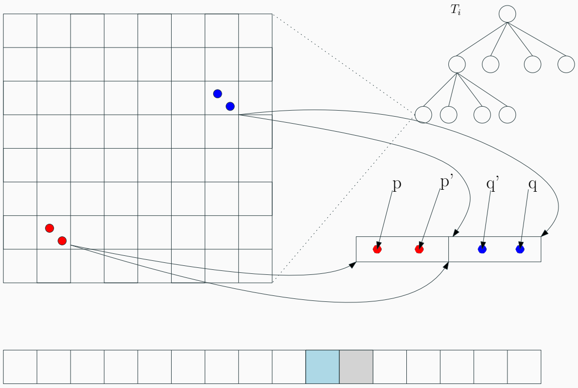
Let p and q constitute a red-blue Nearest Neighbor of the point set.



Bichromatic NN

Approximate Bichromatic NN

Let p and q constitute a red-blue Nearest Neighbor of the point set.



$$||p' - q'|| \leq ||p' - p|| + ||p - q|| + ||q - q'|| \leq (1 + 2\epsilon)||p - q||$$

Bichromatic ANN Algorithm

Input: Bichromatic point set $R \cup B \in [0,1)^d$.
Output: Bichromatic ANN pair (r, b), $r \in R, b \in B$.

For each of $D = O(d)$ quadtrees of shifted point sets &
For each of the $\log \frac{1}{\epsilon}$ ϵ-quadtrees

1. Construct $O(\frac{1}{\epsilon^d})$ Walecki's permutations.
2. For each permutation, perform DFS traversal of the ϵ-quadtrees, resulting in a linear order of points in P.
3. Among all pairs of consecutive red-blue points in all the linear orders, find the pair (r, b) that minimizes $||r - b||$.
4. Report (r, b) as Bichromatic ANN.

Bichromatic ANN

Bichromatic ANN Theorem (CHJ19)

Let R and B be two sets of points in $[0,1)^d$ and let $\epsilon \in (0,1)$ be a parameter. Then one can maintain a $(1+\epsilon)$-approximation to the bichromatic closest pair in $R \times B$ under updates (i.e., insertions and deletions) in $O(\log n \log^2 \frac{1}{\epsilon}/\epsilon^d)$ time per operation, where n is the total number of points in the two sets. The data structure uses $O(n \log \frac{1}{\epsilon}/\epsilon^d)$ space, and at all times maintains a pair of points $r \in R$, $b \in B$, such that $||r - b|| \leq (1+\epsilon)d(R,B)$, where $d(R,B) = \min_{r \in R, b \in B} ||r - b||$.

Conclusions

- Variants of linear orders are used to construct dynamic structures for ANN, Geometric Spanners, Approximate EMST, etc.
- Find more applications where this framework can be applied.

References

References

Timothy M. Chan, Sariel Har-Peled, Mitchell Jones: On Locality-Sensitive Orderings and Their Applications. SIAM Journal of Computing 49(3): 583-600, 2020.