Locality-Sensitive Orderings & Their Applications

Timothy Chan, Sariel Har-Peled, <u>Mitchell Jones</u> September 11, 2018

University of Illinois at Urbana-Champaign

- Computing orderings:
 - Travelling salesman problem (hard)
 - Degeneracy of a graph (easy)

- Computing orderings:
 - Travelling salesman problem (hard)
 - Degeneracy of a graph (easy)
- Orderings are 1D embeddings
- Embedding into simpler structures

- Computing orderings:
 - Travelling salesman problem (hard)
 - Degeneracy of a graph (easy)
- Orderings are 1D embeddings
- Embedding into simpler structures
- ▶ E.g. $O(\log n)$ -apx for k-median clustering: Metric space \rightarrow tree

- Computing orderings:
 - Travelling salesman problem (hard)
 - Degeneracy of a graph (easy)
- Orderings are 1D embeddings
- Embedding into simpler structures
- ▶ E.g. $O(\log n)$ -apx for k-median clustering: Metric space \rightarrow tree
- ► Tradeoff between soln. quality and simpler algorithms

- Computing orderings:
 - Travelling salesman problem (hard)
 - Degeneracy of a graph (easy)
- Orderings are 1D embeddings
- Embedding into simpler structures
- ▶ E.g. $O(\log n)$ -apx for k-median clustering: Metric space \rightarrow tree
- Tradeoff between soln. quality and simpler algorithms
- ► In this talk: Orderings of points with special properties

For $\varepsilon \in (0,1)$, there is a set Π of size $O((1/\varepsilon^d) \log(1/\varepsilon))$ such that $\forall \rho, q \in [0,1)^d$, there exists $\sigma \in \Pi$ with:

Points between p and q in σ are distance at most $\varepsilon ||p-q||$ from p or q.

For $\varepsilon \in (0,1)$, there is a set Π of size $O((1/\varepsilon^d) \log(1/\varepsilon))$ such that $\forall \rho, q \in [0,1)^d$, there exists $\sigma \in \Pi$ with:

Points between p and q in σ are distance at most $\varepsilon || p - q ||$ from p or q.

Some applications

• New: $(1 + \varepsilon)$ -apx bichromatic closest pair

For $\varepsilon \in (0,1)$, there is a set Π of size $O((1/\varepsilon^d) \log(1/\varepsilon))$ such that $\forall \rho, q \in [0,1)^d$, there exists $\sigma \in \Pi$ with:

Points between p and q in σ are distance at most $\varepsilon ||p-q||$ from p or q.

Some applications

- New: $(1 + \varepsilon)$ -apx bichromatic closest pair
- Simpler: dynamic $(1 + \varepsilon)$ -spanners

For $\varepsilon \in (0, 1)$, there is a set Π of size $O((1/\varepsilon^d) \log(1/\varepsilon))$ such that $\forall \rho, q \in [0, 1)^d$, there exists $\sigma \in \Pi$ with:

Points between p and q in σ are distance at most $\varepsilon ||p-q||$ from p or q.

Some applications

- New: $(1 + \varepsilon)$ -apx bichromatic closest pair
- Simpler: dynamic $(1 + \varepsilon)$ -spanners
- New: dynamic k-vertex-fault-tolerant $(1 + \varepsilon)$ -spanners

For $\varepsilon \in (0,1)$, there is a set Π of size $O((1/\varepsilon^d) \log(1/\varepsilon))$ such that $\forall \rho, q \in [0,1)^d$, there exists $\sigma \in \Pi$ with:

Points between p and q in σ are distance at most $\varepsilon ||p-q||$ from p or q.

Some applications

- New: $(1 + \varepsilon)$ -apx bichromatic closest pair
- Simpler: dynamic $(1 + \varepsilon)$ -spanners
- New: dynamic k-vertex-fault-tolerant $(1 + \varepsilon)$ -spanners

► ...

Warmup: Constant factor approximation for bichromatic closest pair

Bichromatic closest pair



Bichromatic closest pair



Problem (*c***-approximation)**

Maintain a pair
$$(r', b')$$
 s.t. $\|r' - b'\| \le c \cdot \min_{(r,b)} \|r - b\|$.

Preliminaries: Quadtrees

"Hierarchy of grids"



Preliminaries: 2-order

Mapping points into 1D





Preliminaries: Quadtrees and $\ensuremath{\mathbb{Z}}\xspace$ -order

• DFS of a quadtree produces a \mathcal{Z} -order





Preliminaries: Quadtrees and $\ensuremath{\mathbb{Z}}\xspace$ -order

- DFS of a quadtree produces a \mathcal{Z} -order
- Only need to specify an order on 4 cells (or 2^d for higher dimensions)





Preliminaries: Computing the \mathcal{Z} -order

- Let $p = (x, y) \in [2^w] \times [2^w]$
- $x = x_w x_{w-1} \dots x_1$
- $y = y_w y_{w-1} \dots y_1$



Preliminaries: Computing the \mathcal{Z} -order

- Let $p = (x, y) \in [2^w] \times [2^w]$
- $x = x_w x_{w-1} \dots x_1$
- $y = y_w y_{w-1} \dots y_1$
- shuffle(p) = $y_w x_w y_{w-1} x_{w-1} \dots y_1 x_1$
- ▶ Position of *p* in *Z*-order = shuffle(*p*)



Preliminaries: Computing the \mathcal{Z} -order

• Let $p = (x, y) \in [2^w] \times [2^w]$

- $x = x_w x_{w-1} \dots x_1$
- $y = y_w y_{w-1} \dots y_1$
- shuffle(p) = $y_w x_w y_{w-1} x_{w-1} \dots y_1 x_1$
- ▶ Position of p in Z-order = shuffle(p)



Lemma

shuffle(p) and shuffle(q) can be compared in O(1) and/exclusive-or operations.

Map the point set to 1D



- Map the point set to 1D
- Maintain sorted order in binary tree



- Map the point set to 1D
- Maintain sorted order in binary tree
- Maintain min-heap of consecutive red/blue pairs



- Map the point set to 1D
- Maintain sorted order in binary tree
- Maintain min-heap of consecutive red/blue pairs
- Updates change O(1) consecutive pairs



- Map the point set to 1D
- Maintain sorted order in binary tree
- Maintain min-heap of consecutive red/blue pairs
- Updates change O(1) consecutive pairs
 - \implies Update time $O_d(\log n)$



► Points nearby in ℝ^d ⇒ nearby in 2-order



- ► Points nearby in ℝ^d ⇒ nearby in 2-order
- Idea: Shift the point set



- ► Points nearby in ℝ^d ⇒ nearby in 2-order
- Idea: Shift the point set



- ► Points nearby in ℝ^d ⇒ nearby in 2-order
- Idea: Shift the point set



- ► Points nearby in ℝ^d ⇒ nearby in 2-order
- Idea: Shift the point set



Lemma [Chan '98]

For i = 0, ..., d, let $v_i = (i/(d+1), ..., i/(d+1))$.

Let $p, q \in [0, 1)^d$ and \mathfrak{T} be a quadtree over $[0, 2)^d$.

There exists $i \in \{0, \ldots, d\}$ and $\Box \in \mathfrak{T}$:

1.
$$p + v_i, q + v_i \in \Box$$

2. $(d+1)||p-q|| < sidelength(\Box) \le 2(d+1)||p-q||$.





 $\implies O_d(\log n)$ update time



 $\implies O_d(\log n)$ update time

• Claim: $O_d(1)$ approximation




 $\mathsf{sidelength}(\Box) \leq 2(d+1)\|r-b\|$







 $(1+\varepsilon)$ -approximate bichromatic closest pair

• Assume $\varepsilon = 2^{-E}$ for $E \in \mathbb{N}$

- Assume $\varepsilon = 2^{-E}$ for $E \in \mathbb{N}$
- ► Idea: Pack many "ε-quadtrees" into a regular quadtree

- Assume $\varepsilon = 2^{-E}$ for $E \in \mathbb{N}$
- ► Idea: Pack many "ε-quadtrees" into a regular quadtree
- ▶ ε -quadtrees have $1/\varepsilon^d$ children

Reducing the approximation factor

- Assume $\varepsilon = 2^{-E}$ for $E \in \mathbb{N}$
- ► Idea: Pack many "ε-quadtrees" into a regular quadtree
- ▶ ε -quadtrees have $1/\varepsilon^d$ children



Reducing the approximation factor

- Assume $\varepsilon = 2^{-E}$ for $E \in \mathbb{N}$
- ► Idea: Pack many "ε-quadtrees" into a regular quadtree
- ▶ ε -quadtrees have $1/\varepsilon^d$ children
- ► "Partitions" a regular quadtree into lg(1/ε) ε-quadtrees



Reducing the approximation factor

- Assume $\varepsilon = 2^{-E}$ for $E \in \mathbb{N}$
- ► Idea: Pack many "ε-quadtrees" into a regular quadtree
- ▶ ε -quadtrees have $1/\varepsilon^d$ children
- ► "Partitions" a regular quadtree into lg(1/ε) ε-quadtrees
- \blacktriangleright Call them $\mathbb{T}^1_\varepsilon,\ldots,\mathbb{T}^E_\varepsilon$



O(1) problems

Extend $\operatorname{\mathfrak{Z}-order}$ to $\operatorname{\varepsilon-quadtrees}$ by ordering $1/\operatorname{\varepsilon}^d$ child cells



What \mathcal{Z} -order should we pick?



O(1) problems (cont.)



 $\mathsf{sidelength}(\Box) \leq 2(d+1) \|p-q\|$

O(1) problems (cont.)



 $\mathsf{sidelength}(\Box) \leq 2(d+1) \|p-q\|$

O(1) problems (cont.)



Idea

Pick a set \mathfrak{O} of orderings of the $1/\varepsilon^d$ cells such that:

For any \Box_1, \Box_2 , there is an ordering $\sigma \in \mathfrak{O}$ with \Box_1 adjacent to \Box_2

Lemma [Alspach '08]

For *n* elements $\{0, \ldots, n-1\}$, there is a set \mathfrak{O} of $\lceil n/2 \rceil$ orderings of the elements, such that, for all $i, j \in \{0, \ldots, n-1\}$, there exist an ordering $\sigma \in \mathfrak{O}$ in which *i* and *j* are adjacent.



Corollary

There is a set $\mathfrak{O}(1/\varepsilon)$ of $O(1/\varepsilon^d)$ orderings, such that for any \Box_1, \Box_2 , there exists an order $\sigma \in \mathfrak{O}(1/\varepsilon)$ where \Box_1 and \Box_2 are adjacent in σ .

• d + 1 shifted point sets

- d + 1 shifted point sets
- ► $lg(1/\varepsilon) \varepsilon$ -quadtrees

- d + 1 shifted point sets
- $\lg(1/\varepsilon) \varepsilon$ -quadtrees
- ► $O(1/\varepsilon^d)$ orderings

- d + 1 shifted point sets
- $\lg(1/\varepsilon) \varepsilon$ -quadtrees
- $O(1/\varepsilon^d)$ orderings
 - $\implies {\it O}_dig((1/arepsilon^d)\log(1/arepsilon)ig)$ different orderings of P

- d + 1 shifted point sets
- $\lg(1/\varepsilon) \varepsilon$ -quadtrees
- $O(1/\varepsilon^d)$ orderings
 - $\implies {\it O}_dig((1/arepsilon^d)\log(1/arepsilon)ig)$ different orderings of P
- Let Π denote these set of orderings

Main Theorem

For $\varepsilon \in (0,1)$, there is a set Π of size $O((1/\varepsilon^d) \log(1/\varepsilon))$ such that $\forall p, q \in [0,1)^d$, there exists $\sigma \in \Pi$ with:

Points between p and q in σ are distance at most $\varepsilon ||p-q||$ from p or q.

Main Theorem

For $\varepsilon \in (0, 1)$, there is a set Π of size $O((1/\varepsilon^d) \log(1/\varepsilon))$ such that $\forall \rho, q \in [0, 1)^d$, there exists $\sigma \in \Pi$ with:

Points between p and q in σ are distance at most $\varepsilon || p - q ||$ from p or q.

Lemma

Let $p, q \in [0, 1)^d$ and $\sigma \in \Pi$. Can decide if $p \prec_{\sigma} q$ using $O_d(\log(1/\varepsilon))$ bitwise-logical operations.

 \blacktriangleright Maintain the 1D data structure for all orderings Π

- \blacktriangleright Maintain the 1D data structure for all orderings Π
- $\blacktriangleright \ |\Pi| = O((1/\varepsilon^d) \log(1/\varepsilon))$

- \blacktriangleright Maintain the 1D data structure for all orderings Π
- $|\Pi| = O((1/\varepsilon^d) \log(1/\varepsilon))$
- ► Update time: $O(|\Pi| \cdot \log(n) \cdot \log(1/\varepsilon)) = O_d((1/\varepsilon^d) \log(n) \log^2(1/\varepsilon))$

- \blacktriangleright Maintain the 1D data structure for all orderings Π
- $|\Pi| = O((1/\varepsilon^d) \log(1/\varepsilon))$
- ► Update time: $O(|\Pi| \cdot \log(n) \cdot \log(1/\varepsilon)) = O_d((1/\varepsilon^d) \log(n) \log^2(1/\varepsilon))$
- Space: $O(|\Pi| \cdot n) = O_d((n/\varepsilon^d) \log(1/\varepsilon))$

- Maintain the 1D data structure for all orderings Π
- $|\Pi| = O((1/\varepsilon^d) \log(1/\varepsilon))$
- ► Update time: $O(|\Pi| \cdot \log(n) \cdot \log(1/\varepsilon)) = O_d((1/\varepsilon^d) \log(n) \log^2(1/\varepsilon))$
- Space: $O(|\Pi| \cdot n) = O_d((n/\varepsilon^d) \log(1/\varepsilon))$
- ▶ Claim: Maintains r', b' with $||r' b'|| \le (1 + \varepsilon)||r b||$



 $\mathsf{sidelength}(\Box) \leq 2(d+1)\|r-b\|$



 $\mathsf{sidelength}(\Box) \leq 2(d+1)\|r-b\|$



 $\mathsf{sidelength}(\Box) \leq 2(d+1) \|r-b\|$




A simple data structure for dynamic $(1 + \varepsilon)$ -spanners

Definition

For a set *n* of *P* points in \mathbb{R}^d and $t \ge 1$, a *t*-spanner of *P* is a graph G = (P, E) such that for all $p, q \in P$,

$$\|p-q\| \leq \operatorname{dist}_G(p,q) \leq t\|p-q\|.$$

▶ For each $\sigma \in \Pi$, connect the *n* consecutive points with *n* − 1 edges

- ▶ For each $\sigma \in \Pi$, connect the *n* consecutive points with *n* − 1 edges
- $(n-1)|\Pi| = O_d((n/\varepsilon^d)\log(1/\varepsilon))$ edges

- ▶ For each $\sigma \in \Pi$, connect the *n* consecutive points with *n* − 1 edges
- $(n-1)|\Pi| = O_d((n/\varepsilon^d)\log(1/\varepsilon))$ edges
- Maximum degree $O_d((1/\varepsilon^d)\log(1/\varepsilon))$

- ▶ For each $\sigma \in \Pi$, connect the *n* consecutive points with *n* − 1 edges
- $(n-1)|\Pi| = O_d((n/\varepsilon^d)\log(1/\varepsilon))$ edges
- Maximum degree $O_d((1/\varepsilon^d) \log(1/\varepsilon))$
- Update time $O_d((1/\varepsilon^d)\log(n)\log^2(1/\varepsilon))$

- ▶ For each $\sigma \in \Pi$, connect the *n* consecutive points with *n* − 1 edges
- $(n-1)|\Pi| = O_d((n/\varepsilon^d)\log(1/\varepsilon))$ edges
- Maximum degree $O_d((1/\varepsilon^d)\log(1/\varepsilon))$
- Update time $O_d((1/\varepsilon^d)\log(n)\log^2(1/\varepsilon))$
- Claim: G is a $(1 + \varepsilon)$ -spanner

Prove by induction on length of pairs:

 $\mathsf{dist}_{G}(p,q) \leq (1+\varepsilon) \|p-q\|$



- Prove by induction on length of pairs: dist_G(p,q) ≤ (1 + ε) ||p − q||
- ► G is a (1 + c_dε)-spanner for const. c_d



- Prove by induction on length of pairs: dist_G(p,q) ≤ (1 + ε) ||p − q||
- ► G is a (1 + c_dε)-spanner for const. c_d
- ► Readjust ε by c_d

 $\begin{aligned} & \begin{array}{c} & & \\ &$

Static & dynamic vertex-fault-tolerant spanners

Definition For a set of *n* points *P* in \mathbb{R}^d and $t \ge 1$, a *k*-vertex-fault-tolerant *t*-spanner of *P* is a graph G = (P, E) such that

- 1. *G* is a *t*-spanner, and
- 2. For any $P' \subseteq P$, $|P'| \leq k$, $G \setminus P'$ is a *t*-spanner for $P \setminus P'$.

For each σ ∈ Π and each p ∈ P, connect p to it's k + 1 predecessors and successors

- For each σ ∈ Π and each p ∈ P, connect p to it's k + 1 predecessors and successors
- $O(kn|\Pi|) = O_d((kn/\varepsilon^d)\log(1/\varepsilon))$ edges

- For each σ ∈ Π and each p ∈ P, connect p to it's k + 1 predecessors and successors
- $O(kn|\Pi|) = O_d((kn/\varepsilon^d)\log(1/\varepsilon))$ edges
- Maximum degree $O_d((k/\varepsilon^d)\log(1/\varepsilon))$









Update time $O_d((\log n \log(1/\varepsilon) + k) \log(1/\varepsilon)/\varepsilon^d)$

• G is already a $(1 + \varepsilon)$ -spanner



- G is already a $(1 + \varepsilon)$ -spanner
- Consider $P' \subseteq P$, $|P'| \leq k$



- G is already a $(1 + \varepsilon)$ -spanner
- Consider $P' \subseteq P$, $|P'| \leq k$



- G is already a $(1 + \varepsilon)$ -spanner
- Consider $P' \subseteq P$, $|P'| \leq k$
- Let $\sigma \in \Pi$ with P' removed



- G is already a $(1 + \varepsilon)$ -spanner
- Consider $P' \subseteq P$, $|P'| \leq k$
- Let $\sigma \in \Pi$ with P' removed



- G is already a $(1 + \varepsilon)$ -spanner
- Consider $P' \subseteq P$, $|P'| \leq k$
- Let $\sigma \in \Pi$ with P' removed
- Consecutive points in P \ P' remain in G \ P' (by construction)



- G is already a $(1 + \varepsilon)$ -spanner
- Consider $P' \subseteq P$, $|P'| \leq k$
- Let $\sigma \in \Pi$ with P' removed
- Consecutive points in P \ P' remain in G \ P' (by construction)

 \implies $G \setminus P'$ is a $(1 + \varepsilon)$ -spanner for $P \setminus P'$



Conclusion

Main Theorem

For $\varepsilon \in (0,1)$, there is a set Π of size $O((1/\varepsilon^d) \log(1/\varepsilon))$ such that $\forall \rho, q \in [0,1)^d$, there exists $\sigma \in \Pi$ with:

Points between p and q in σ are distance at most $\varepsilon ||p-q||$ from p or q.

 Approximate bichromatic closest pair (improved update time to O(log n))

- Approximate bichromatic closest pair (improved update time to O(log n))
- 2. Dynamic spanners (simpler data structure)

- Approximate bichromatic closest pair (improved update time to O(log n))
- 2. Dynamic spanners (simpler data structure)
- 3. Static vertex-fault-tolerant spanners (simple data structure)

- Approximate bichromatic closest pair (improved update time to O(log n))
- 2. Dynamic spanners (simpler data structure)
- 3. Static vertex-fault-tolerant spanners (simple data structure)
- 4. Dynamic vertex-fault-tolerant spanners (previous work?)

- Approximate bichromatic closest pair (improved update time to O(log n))
- 2. Dynamic spanners (simpler data structure)
- 3. Static vertex-fault-tolerant spanners (simple data structure)
- 4. Dynamic vertex-fault-tolerant spanners (previous work?)

Other applications:

1. Approximate nearest neighbor (not new)

- Approximate bichromatic closest pair (improved update time to O(log n))
- 2. Dynamic spanners (simpler data structure)
- 3. Static vertex-fault-tolerant spanners (simple data structure)
- 4. Dynamic vertex-fault-tolerant spanners (previous work?)

Other applications:

- 1. Approximate nearest neighbor (not new)
- 2. Dynamic approximate MST (uses dynamic spanners)