Approximate Nearest neighbor problem amidst box obstacles in $\mathbb{R}^d$

October 11, 2024

Input

- B: the set of n disjoint axis-aligned boxes in $\mathbb{R}^d$
- $S(=\{s_1, \dots, s_k\})$: the set of k sites in $\mathbb{R}^d$
- Query point $q \in \mathbb{R}^d$
- Approximation factor ε

Output

- Return one of s_i such that $d(q, s_i) \le (1+\varepsilon) \cdot d(q, s_{i^*})$ where $i^* = \arg\min_i d(q, s_i)$ and $d(x, y)$ is the Euclidean geodesic distance from x to y.

Sketch of the algorithm

1. For any $(d-2)$-face of the boxes, generate Steiner points in the following way.
   - Without loss of generality, suppose that a $(d-2)$-face is represented as $\Pi_{i=1}^{d-2}[a_{ki}, b_{ki}] \times a_{k(d-1)} \times a_{kd}$.
   - The interval is defined as $\varepsilon' = \frac{r_{\min}}{(2d+1)^{\frac{3}{2}}}\varepsilon$ where $r_{\min}$ is the distance of the closest pair on $B \cup S \cup \{q\}$.
   - Then, we place Steiner points on each point in $\Pi_{i=1}^{d-2}\{a_{ki} + m_{bi} + j\varepsilon' \mid j = 0, 1, ..., \lfloor \frac{b_{ki}-a_{ki}}{\varepsilon} \rfloor\} \times \{a_{k(d-1)}\} \times \{a_{kd}\}$, where the i-th *margin* m_{bi} is defined as $\{(b_{ki} - a_{ki}) - \lfloor \frac{b_{ki}-a_{ki}}{\varepsilon'} \rfloor \cdot \varepsilon'\}/2$.
2. Construct the visibility graph $G = (V_\varepsilon, E_\varepsilon)$ with respect to B where V_ε is the union of $S \cup \{q\}$ and the Steiner points generated in 1.
3. Find s_i that minimizes $d_G(q, s_i)$ by running Dijkstra algorithm on $G = (V_\varepsilon, E_\varepsilon)$.

Correctness of the algorithm

We show that $d_G(q, s_i) \leq (1+\varepsilon)d(q, s_i)$ for any $s_i \in S$. Let $\pi = p_1 \dots p_m (p_1 = q$ and $p_m = s_i)$ be the shortest Euclidean geodesic path from q to s_i.

1. For any $1 < l < m$, p_k is on a $d-2$-dimensional face of an obstacle. If p_l is on the interior of a $(d-1)$-dimensional face, $|p_{l-1}p_l| + |p_l p_{l+1}|$ can be reduced by moving p_l on the face.

2. For any $1 < l < m$, there exists a Steiner point p'_i contained in $Q_l := \Pi_{i=1}^{d}[p_{ki} - \frac{\varepsilon'}{2}, p_{li} + \frac{\varepsilon'}{2}]$ where p_{li} is the i-th coordinate of p_l.

3. For $1 \leq l < m$, the maximum distance from Q_l to Q_{l+1} is not more than $|p_{l-1}p_l| + \frac{\varepsilon}{2d+1} \cdot r_{\min}$.

4. Suppose that π hits the box b_1 and b_2 in B concecutively. Let $p_{l_1} \in \pi$ be the first point hitting b_1 and $p_{l_2} \in \pi$ be the first point hitting b_2. Then the number of segments of π between p_{l_1} and p_{l_2} is at most $2d+1$ since the number of $(d-1)$-dimensional faces of a box is $2d$ and each face contain at most one segment of π.

5. Therefore, $d_G(p_{l_1}, p_{l_2}) \leq \sum_{i=l_1}^{l_2-1} |p'_i p'_{i+1}| \leq d(p_{l_1}, p_{l_2}) + \varepsilon \cdot r_{\min} \leq (1+\varepsilon) \cdot d(p_{l_1}, p_{l_2})$.

Complexity of the algorithm

Let denote $C := (\frac{l_{\max}}{r_{\min}})^d$ where $l_{\max}$ is the longest side of the boxes in B. Then, the number of Steiner points is $O(\frac{d^{1.5d-1}}{\varepsilon^{d-2}}n)$ and the number of edges in the visibility graph is $O(\frac{d^{3d-2}}{\varepsilon^{2(d-2)}}n^2)$. So, we can construct the visibility graph in $O(\frac{d^{3d-2}}{\varepsilon^{2(d-2)}}n^3)$ time since we can determine that a segment intersects with B in $O(n)$ time and the Dijkstra from the query point spends $O(\frac{d^{3d-2}}{\varepsilon^{2(d-2)}}n^2 \log n)$ time.