NP-Hardness of the Problem of Optimal Box Positioning

Alexei V. Galatenko, Stepan A. Nersisyan * and Dmitriy N. Zhuk

Faculty of Mechanics and Mathematics, Lomonosov Moscow State University, Leninskie gory 1, 119991 Moscow, Russia
* Correspondence: s.a.nersisyan@gmail.com

Received: 12 June 2019; Accepted: 4 August 2019; Published: 6 August 2019

Abstract: We consider the problem of finding a position of a $d$-dimensional box with given edge lengths that maximizes the number of enclosed points of the given finite set $P \subset \mathbb{R}^d$, i.e., the problem of optimal box positioning. We prove that while this problem is polynomial for fixed values of $d$, it is NP-hard in the general case. The proof is based on a polynomial reduction technique applied to the considered problem and the 3-CNF satisfiability problem.

Keywords: optimal box positioning; NP-hardness; computational geometry

1. Introduction

We consider the problem of optimal box positioning, that is, finding a position of a $d$-dimensional box with given edge lengths that maximizes the number of enclosed points of a given $n$-element set $P \subset \mathbb{R}^d$. In this paper, we prove that this problem is NP-hard when integers $n, d$ are not fixed and treated as parameters of the problem.

The problem of optimal box positioning has wide applications in computational geometry, data mining, and pattern recognition (e.g., see [1–3]). In [4], the authors presented a clustering approach based on the greedy algorithm that finds an approximate solution of the optimal box positioning problem. The algorithm was inspired by the apparatus of maximum interval pattern concepts (see, e.g., [4,5]), a technique that allows one to select patterns from fuzzy contexts. This approach was successfully applied to the dataset of tactile images registered by the Medical Tactile Endosurgical Complex [6–8], which allows intraoperative tactile examination of tissues. Comparison of the proposed clustering approach with the conventional $k$-means clustering resulted in a statistically significant advantage of the proposed method over $k$-means in clustering quality. Note that the result proved in the present paper justifies developing algorithms to solve an approximate version of the optimal box positioning over the exact one.

The rest of the paper is organized as follows. In Section 2, we describe some known results. In Section 3, we introduce formal definitions and formulate the problem of optimal integer box positioning and the auxiliary problem of the existence of an integer $m$-box. In Section 4, we prove the NP-hardness of the problem of optimal box positioning. In Section 5, we summarize the results.

2. Previous Results

Despite the fact that, to the best of our knowledge, a proof of the NP-hardness of the optimal box positioning problem is not available so far, some results are known for related problems. For example, in [3], Eckstein et al. considered a generalization of the problem of optimal box positioning: given two finite sets $P_+, P_-$ in $\mathbb{R}^d$, find a box $B$ with arbitrary edge lengths such that
1. $B$ does not intersect with $P_-$, and
2. the cardinality of $B \cap P_+$ is maximal over all boxes that satisfy the first condition.

The authors proved the NP-hardness of the problem by applying a polynomial reduction of the classical NP-hard problem of finding a maximum independent set of vertices in a graph (e.g., see [9]) to the considered problem.

Barbay et al. considered a weighted generalization of the problem: given a finite set $P \subset \mathbb{R}^d$ and a function $w : P \to \mathbb{R} \cup \{-\infty\}$, find a box $B$ with arbitrary edge lengths that maximizes the sum $\sum_{p \in B \cap P} w(p)$ [10]. This problem is also NP-hard since it generalizes the previous problem in which points from $P_+$ have weight $+1$ and points from $P_-$ have weight $-\infty$.

De Figueiredo and da Fonseca considered the weighted problem of the optimal unit ball positioning with non-negative weight function [11]. They obtained a lower bound $\Omega(n^d)$ under an additional restriction: an algorithm decides which operation to apply to a point based only on the coordinates of the point, ignoring its weight, so it processes input points in an order that does not depend on a weight function. Under this restriction, an algorithm must calculate the weight for each ball that is optimal for some weight function. Note that this restriction is not met in the unweighted version of the problem since we use the fact that the weight of each point is equal to $+1$. Also note that a unit box is a unit ball in $\ell^\infty$ metric.

3. Formal Definitions

Definition 1. A $d$-dimensional box with edge lengths $\delta_1, \delta_2, \ldots, \delta_d$ is a Cartesian product of the intervals $[a_1, b_1] \times \cdots \times [a_d, b_d]$, where $b_i - a_i = \delta_i$ ($i \in \{1, \ldots, d\}$).

Furthermore, we consider only boxes with integer edge lengths and vertex coordinates, i.e., $\delta_i, a_i, b_i \in \mathbb{Z}$ ($i \in \{1, \ldots, d\}$). We call such boxes integer boxes.

Definition 2. The problem of optimal integer box positioning is defined as follows: find an integer box with given edge lengths that maximizes the number of enclosed points of a set $P = \{p_i\}_{i=1}^n \subset \mathbb{Z}^d$.

In Section 4, we obtain NP-hardness of the problem of optimal integer box positioning as a corollary of the theorem about NP-completeness of the problem of the existence of an integer $m$-box.

Definition 3. The problem of the existence of an integer $m$-box is a problem of the existence of an integer box with given edge lengths that contains at least $m$ points from a set $P = \{p_i\}_{i=1}^n \subset \mathbb{Z}^d$.

In general, case parameters of both problems are integers $n, d, \delta_1, \delta_2, \ldots, \delta_d$ and a set $P$. The number $m$ is considered as a function of $n, d$ or as a constant.

It is easy to see that both problems belong to the P complexity class if the parameter $d$ is fixed. Indeed, without loss of generality, we can consider only boxes for which each $a_i$ is equal to the $i$-th coordinate of some point $p_i$ from the set $P$. So to solve the problem, we can count the number of points in at most $n^d$ boxes. Since each count can be performed in $O(nd)$ operations, the total number of operations for solving the problem is $O(dn^{d+1})$, which is polynomial in $n$.

Definition 4. The 3-CNF satisfiability problem is the problem of the existence of an assignment $(s_1, \ldots, s_d) \in \{0, 1\}^d$ to the Boolean variables $x_1, \ldots, x_d$ which turns formula $\wedge_{i=1}^n l_{1i} \lor l_{2i} \lor l_{3i}$ in the conjunctive normal form to 1 (here, $l_{ij}$ denotes literals over variables from the set $\{x_1, \ldots, x_d\}$). For further details, see e.g., [9].

Without loss of generality, assume that variables of every conjunctive clause are distinct. Indeed, otherwise a clause is either identically equal to 1 (if it contains both a variable and its negation) or can be replaced with at most four clauses with the required property such that the conjunction of these clauses is identically equal to the initial clause.
Cook’s theorem [12] states that the 3-CNF satisfiability problem is NP-complete. This fact will give ground for our proof of NP-hardness of the problem of the existence of an integer m-box.

4. NP-Hardness of the Problem of Optimal Box Positioning

Theorem 1. The problem of the existence of an integer m-box belongs to the NP complexity class.

Proof. Suppose we have a certificate: a box \( B \) which encloses at least \( m \) points from the set \( P \). Then the certificate validation can be performed by counting cardinality of \( B \cap P \), which can be done by iterating over the set \( P \) and checking whether the current point lies in the box \( B \). Since \( P \) contains \( n \) elements and each check can be done with \( O(d) \) comparisons, counting cardinality of \( B \cap P \) will take \( O(dn) \) operations, which is polynomial in parameters \( n,d \). \( \blacksquare \)

Theorem 2. The problem of the existence of an integer m-box is NP-hard.

Proof. We will prove this theorem by employing a polynomial reduction of the 3-CNF satisfiability problem (which is NP-hard [12]) in the problem of the existence of an integer m-box. Consider an arbitrary formula \( F \) in conjunctive normal form with \( d \) variables \( x_1, \ldots, x_d \) and \( n \) disjunctive clauses \( D_1, \ldots, D_n \), each containing exactly 3 literals: \( F = \bigwedge_{i=1}^n D_i \), where \( D_i = l_{i,1} \lor l_{i,2} \lor l_{i,3} \); \( l_{i,j} \) denotes a literal over one of the variables \( x_1, \ldots, x_d \).

We construct the set \( P = \{ p_i \} \subset \mathbb{Z}^d \) by the following procedure. Consider the disjunctive clause \( D_i \) with variables \( x_{i,1}, x_{i,2}, x_{i,3} \) and the set of its satisfying assignments \( S_i = \{ S_{ij} \} \) over the variable set \( \{ x_{i,1}, x_{i,2}, x_{i,3} \} \). Since each disjunctive clause contains exactly 3 literals corresponding to distinct variables, it holds that \( |S_i| = 7 \). We map the pair \((D_i, S_{ij})\) to the point \( z_{ij} \in \mathbb{Z}^d \) with coordinates \((z^1, \ldots, z^d)\) by the following rule:

\[
\begin{align*}
  z^i &= \begin{cases} 
    0, & \text{if } x_i \in \{ x_{i,1}, x_{i,2}, x_{i,3} \} \text{ and the value of } x_i \text{ in } S_{ij} \text{ is 0;} \\
    1, & \text{if } x_i \notin \{ x_{i,1}, x_{i,2}, x_{i,3} \}; \\
    2, & \text{if } x_i \in \{ x_{i,1}, x_{i,2}, x_{i,3} \} \text{ and the value of } x_i \text{ in } S_{ij} \text{ is 1.}
  \end{cases}
\end{align*}
\]

We define the set \( P \) as an image of this map over all clauses \( D_1, \ldots, D_n \) and their sets of satisfying assignments \( S_1, \ldots, S_n \), so \( |P| \leq 7n \). For further convenience, we also introduce sets \( Z_i = \{ z_{ij} \}, i \in \{ 1, \ldots, n \} \), as subsets of \( P \) that consist of all points associated with \( D_i \).

To complete the proof of the theorem, we prove the following lemmas.

Lemma 1. In the above notation, for an arbitrary unit cube \( C \subset \mathbb{Z}^d \) and for all \( i \in \{ 1, \ldots, n \} \), the intersection \( C \cap Z_i \) contains zero points or one point.

Proof. Consider an arbitrary \( i \in \{ 1, \ldots, n \} \) and points \( Z_i = \{ z_{ij} \}_{j=1}^7 \) associated with \( D_i \). Since for any \( j,k \in \{ 1, \ldots, 7 \}, j \neq k \) satisfying assignments \( S_{ij} \) and \( S_{ik} \) are different, there exists \( l \in \{ 1, \ldots, d \} \) such that the values of variable \( x_l \in \{ x_{i,1}, x_{i,2}, x_{i,3} \} \) in \( S_{ij} \) and \( S_{ik} \) are opposite. Hence, the \( l \)th coordinates of \( z_{ij} \) and \( z_{ik} \) differ by 2 (one of these coordinates equals 0, and the other equals 2). Thus, points \( z_{ij} \) and \( z_{ik} \) cannot belong to the same unit cube. \( \blacksquare \)

Lemma 2. In the above notation, a formula \( F \) is satisfiable if and only if there exists a unit cube \( C \subset \mathbb{Z}^d \) such that \( |C \cap P| = n \).

Proof. Let us first prove that if \( F \) is satisfiable, then a box \( C \) with \( |C \cap P| = n \) exists. Let \( S = (s^1, \ldots, s^d) \) be a satisfying assignment for \( F \). We construct a subset \( P \subset P \) consisting of the points that correspond to the satisfying assignments \( S_{ij} \) matching the satisfying assignment \( S \). Since for each \( i \in \{ 1, \ldots, n \} \) there exists exactly one satisfying assignment \( S_{ij} \in S_i \) that matches \( S \), we have \( |P| = n \). Let \( z = (z^1, \ldots, z^d) \) be an arbitrary point in \( P \) and \( l \in \{ 1, \ldots, d \} \). If \( x_l \) is not met in the
respective clause, the value of $z^l$ will be equal to 1. Otherwise, the value of $z^l$ will be equal to $2 \cdot s^l$. This means that if $s^l = 0$, the value of $z^l$ will lie in the interval $[0, 1]$, and otherwise in the interval $[1, 2]$. Thus, the cube $C = [a_1, a_1 + 1] \times \ldots \times [a_d, a_d + 1]$, where

$$a_l = \begin{cases} 
0, & \text{if } s^l = 0, \\
1, & \text{otherwise,}
\end{cases}$$

covers the $n$-element set $\tilde{P} \subset P$. Note that $\tilde{P}$ contains exactly one point corresponding to each clause, so according to Lemma 1, the cube $C$ has no common points with $P \setminus \tilde{P}$. Thus $|C \cap P| = n$.

Now we prove that if a unit cube with $|C \cap P| = n$ exists, then $F$ is satisfiable. Let $C$ be the specified unit cube. By Lemma 1, we conclude that $C \cap P$ contains exactly one point corresponding to each clause. Since each edge length of $C$ is equal to 1 and the cube vertex coordinates are integers, the list of $l$-th coordinates of the points from $C \cap P$ (for fixed $l \in \{1, \ldots, d\}$) contains exactly one value from the set $\{0, 2\}$, and we denote this value by $2 \cdot s^l$. From the procedure of construction of the set $P$, we conclude that $S = (s^1, \ldots, s^d)$ is a satisfying assignment for $F$.

Lemmas 1 and 2 directly imply the following assertion.

**Lemma 3.** In the above notation, a formula $F$ is satisfiable if and only if there exists a unit $m$-cube for $m = n$ and the set $P$.

To complete the proof of Theorem 2, we consider the problem of the existence of an integer $m$-box (with $m$ equal to $n$) in $d$-dimensional space for a box with all edge lengths equal to 1 (i.e., for the unit cube) and the constructed set $P$. Lemma 3 states that $F$ is satisfiable if and only if there exists a unit cube that encloses $n$ points. This statement in combination with the fact that set $P$ can be constructed in time polynomial in $n, d$ completes the proof of the theorem.

Since the class of NP-complete problems is the intersection of the class NP and the class NP-hard, Theorems 1 and 2 immediately lead to the following theorem.

**Theorem 3.** The problem of the existence of an integer $m$-box is NP-complete.

Now we are ready to prove the main theorem.

**Theorem 4.** The problem of optimal integer box positioning is NP-hard.

**Proof.** This theorem is a trivial corollary of Theorem 3. Consider a set $P = \{p_i\}_{i=1}^n \subset \mathbb{Z}^d$. Then, finding the optimal position of an integer box $B$ with edge lengths $\delta_1, \delta_2, \ldots, \delta_d$ immediately leads to an answer to the problem of the existence of an integer $m$-box (by simply counting the number of points in the found box in $O(nd)$ operations and comparing it with $m$), which is proved to be NP-complete. Thus, we made a polynomial reduction of the problem of the existence of an integer $m$-box to the problem of optimal integer box positioning.

Note that the above proofs actually lead to stronger results, namely to NP-completeness of the problem of the existence of an integer unit $m$-cube and the NP-hardness of the problem of optimal integer unit cube positioning.

**Corollary 1.** The problem of optimal integer box positioning with a set of prohibited points $P_-$ (i.e., box should have an empty intersection with it) is NP-hard.

**Proof.** This statement immediately follows from the NP-hardness of the problem of optimal integer box positioning since it is a particular case of the considered problem with $P_- = \emptyset$. 


Corollary 2. The weighted problem of optimal integer box positioning with the range of the weight function in \( \mathbb{R} \cup \{-\infty\} \) is NP-hard.

Proof. This is also a corollary of the NP-hardness of the problem of optimal integer box positioning since we obtain an unweighted version of the problem by setting the weight function to +1 for all points. \( \square \)

5. Conclusions

The problem of optimal box positioning finds its applications in computer science, pattern recognition, and data analysis [1–4]. In this paper, we have proved that this problem is NP-hard.

On the one hand, this result means that algorithms based on optimal box positioning are in general inefficient for the analysis of high-dimensional data, thus it makes sense to develop algorithms that look for an approximately optimal box position. An example of such an algorithm used for data clustering can be found in [4].

On the other hand, NP-hardness does not necessarily imply average-case hardness. For example, the canonical NP-complete problem of CNF satisfiability (the one used in the proof of Cook’s theorem about the existence of NP-complete problems, [12]) can be solved using an algorithm with polynomial average time [13]. Thus, the problem of estimation of average complexity for finding an optimal box position remains an interesting open challenge.

Author Contributions: All authors contributed equally to the writing of this paper. All authors read and approved the final manuscript.

Funding: The research was supported by the Russian Science Foundation (project 16-11-00058 “The development of methods and algorithms for automated analysis of medical tactile information and classification of tactile images”).

Acknowledgments: The authors thank Vladimir V. Galatenko for valuable comments and discussions.

Conflicts of Interest: The authors declare no conflict of interest.

References


© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).