Unit Covering in Color-Spanning Set Model

Ehsan Emamjomeh-Zadeh[1], Mohammad Ghodsi[2], Hamid Homapour[2], and Masoud Seddighin[2]

Department of Computer Science, University of Southern California
emamjome@usc.edu
Department of Computer Engineering, Sharif University of Technology
ghodsi@sharif.edu
homapour@ce.sharif.edu
mseddighin@ce.sharif.edu

Abstract. In this paper, we consider two new variants of the unit covering problem in color-spanning set model: Given a set of n points in d-dimensional plane colored with m colors, the *MinCSBC* problem is to select m points of different colors minimizing the minimum number of unit balls needed to cover them. Similarly, the *MaxCSBC* problem is to choose one point of each color to maximize the minimum number of needed unit balls. We show that MinCSBC is NP-hard and hard to approximate within any constant factor even in one dimension. For $d = 1$, however, we propose an $\ln(m)$-approximation algorithm and present a constant-factor approximation algorithm for fixed f, where f is the maximum frequency of the colors. For the MaxCSBC problem, we first prove its NP-hardness. Then we present an approximation algorithm with a factor of 1/2 in one-dimensional case.

Keywords: Unit Covering, Color-Spanning Set, Computational Geometry, Approximation Algorithm

1 Introduction

Given a set of n points, the **unit covering (UC)** problem is to cover them with minimum number of unit balls. This problem is NP-hard in Euclidean plane [3], while for constant-dimensional cases, it admits polynomial-time approximation schemes (PTAS) [5]. The UC problem has been studied extensively due to wide applications in many fields such as data management in terrains and wireless networks [2, 1, 4, 10].

Recently, many researchers address geometric problems in the situation where the input data is imprecise [8]. One common approach for modeling imprecise points is to use a set of finite points for possible locations that a single imprecise point may appear. In computational geometry, this problem is named *color-spanning set model*. In this model, we are given n points colored with m colors. Points with the same color refer to possible locations of an imprecise point. Imprecise inputs lead to imprecision of output. One of the widely studied problems in this model is to compute bounds on output [8].

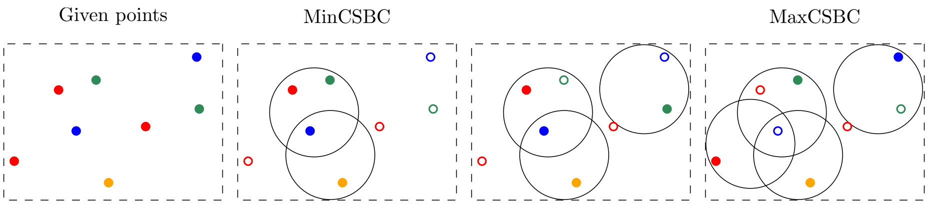

Fig. 1. Three different color selections of given points and their corresponding unit covering (i.e., to cover them using minimum number of unit balls).

In this paper, we discuss the unit covering problem in color-spanning set model. This model can be applied to the case when for each term, at least one of its alternatives should be covered. As an example, consider n different networks and suppose that we want to connect these networks to the Internet with minimum number of access points. Each access point can cover nodes in the certain distance, and a network is connected to the Internet if and only if at least one of its nodes is close enough to an access point, i.e., it is "covered" by the ball corresponding to the access point.

Given a set $P = \{p_1, p_2, ..., p_n\}$ of n points in d dimensions colored with $m \leq n$ colors in $C = \{c_1, c_2, ..., c_m\}$, a **color selection** of P is a subset of m points, one from each color.

We define the following two problems:

Problem 1 (MinCSBC). Find a color selection S of P that minimizes the number of balls in unit covering of S.

Problem 2 (MaxCSBC). Find a color selection S of P that maximizes the number of balls in unit covering of S.

In Figure 1, three different color selections for a set of points and their corresponding unit covering depicted.

2 Preliminaries and Notation

Suppose that P is a set of points given as the input for either MinCSBC or MaxCSBC and $C = \{c_1, \ldots, c_m\}$ is the set of colors of elements in P. For each $c_i \in C$, define the *frequency* of c_i as the number of points colored with c_i. We refer to the maximum frequency as f_P (and omit the subscript P when it is clear from the context), so that no more than f_P points are of the same color. Since there exists only one color selection for $f = 1$, we assume that $f \geq 2$.

Except explicitly specified, we restrict our discussion to one-dimensional case. In this case, a unit ball turns into a unit interval.

Given a color selection S of P, let $U(S)$ denote the set of intervals in the unit covering of S. Recall that unit covering uses the minimum number of intervals to cover the points. We have the following simple observation:

$$G = \{x_1, x_2, x_3, x_4\} \quad F = \{S_1, S_2, S_3, S_4\}$$
$$S_1 = \{x_1, x_3\} \qquad S_2 = \{x_2, x_3\}$$
$$S_3 = \{x_2, x_3, x_4\} \quad S_4 = \{x_1, x_4\}$$

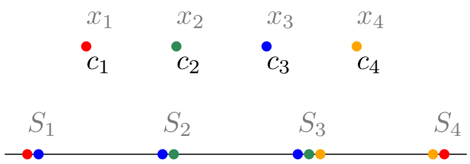

Fig. 2. An instance of MinCSBC.

Observation 1 *There is an optimal covering $U(S)$ such that the left endpoint of each interval corresponds to a point in S and all intervals in $U(S)$ are disjoint.*

Proof. Let $U(S) = \{I_1, I_2, ..., I_k\}$ be the set of intervals sorted by their left endpoints. Starting from I_1, for each interval I_i, shift I_i until its left endpoint lies on the first point that is not covered by intervals $I_1, ..., I_{i-1}$. Clearly, the resulting set of shifted intervals satisfies the required property. □

In the rest of the paper, $U(S)$ refers to an optimal covering with the property in Observation 1. We define OPT_{min} and OPT_{max} as the color selection regarding MinCSBC and MaxCSBC, respectively (for explicitly mentioned P or whenever it is clear from the context).

3 MinCSBC

3.1 Hardness of MinCSBC

Theorem 1. *MinCSBC is NP-hard.*

Proof. We show that the problem is NP-hard even in one dimension using a reduction from the Set Cover. Consider an instance of Set Cover with ground set $G = \{x_1, x_2, ..., x_m\}$, covering family $F = \{S_1, S_2, ..., S_k\}$ and $OPT_{sc} \subseteq F$ as the optimal cover. For each $x_j \in G$, consider color c_j in MinCSBC instance, and for each subset S_i, specify a unit segment on x-axis $Cell_i$ in a way that the distance between the endpoints of different segments is more than 1. Next, for each element $x_j \in S_i$, put a point with color c_j in $Cell_i$ as illustrated in Figure 2.

Suppose that P is the set of created points in the MinCSBC instance. Since the distance between each two cells is more than 1, each interval in $U(OPT_{min})$ covers points in only one cell. Moreover, if two intervals intersect the same cell, it is possible to replace them with one interval which includes the whole cell contradicting the minimality of OPT_{min}. We return the sets whose corresponding cells in MinCSBC instance intersect the intervals in $U(OPT_{min})$. Let R denote the set of returned subsets. Since at least one point of each color is covered by intervals in $U(OPT_{min})$, R is a feasible solution to the Set Cover instance. Consequently, we have $|U(OPT_{min})| \geq |OPT_{sc}|$.

On the other hand, consider the cells corresponding to subsets in OPT_{sc} and find a color selection using points in these cells. Since OPT_{sc} covers all elements in G, such a color selection exists. Obviously, this color selection can be covered by $|OPT_{sc}|$ unit intervals, so $|U(OPT_{min})| \leq |OPT_{sc}|$.

As a consequence, $|U(OPT_{min})| = |OPT_{sc}|$ which, keeping in mind NP-hardness of the Set Cover problem, implies MinCSBC to be NP-hard as well.

□

Note that the Set Cover problem is NP-hard even when the frequency of each $x_j \in G$ is at most 2, i.e., x_j appears in at most two subsets in F. Therefore, using the same reduction for this restricted version of the Set Cover problem, we can claim that one-dimensional MinCSBC is NP-hard even when $f = 2$.

Furthermore, it can be concluded from the above reduction that any constant-factor approximation algorithm for MinCSBC yields an approximation for the Set Cover problem with the same factor. Taking into account that there is no approximation algorithm with a constant factor for the Set Cover problem unless P=NP, we obtain the following corollary.

Corollary 1. *MinCSBC admits no polynomial-time approximation algorithm with a constant factor unless $P = NP$.*

3.2 Approximation Algorithms for MinCSBC

Theorem 2. *There is an $\ln(m)$-approximation algorithm for MinCSBC in one dimension.*

Proof. Let $I = \{I_1, I_2, ..., I_n\}$ be the set of intervals, where I_i is the unit interval whose left endpoint lies on point p_i. By Observation 1, for any color selection, there exists an optimal covering using intervals in I. Therefore, the MinCSBC problem is basically to find $\mathcal{I} \subseteq I$ of the minimum size such that $\mathcal{I}$ covers at least one point of each color, and then choose a color selection from the covered points.

In order to represent this problem with the Set Cover problem:

- let G be the set of all colors;
- for each $I_i \in I$, define a subset of G containing the colors covered by I_i.

There is a well-known greedy approximation algorithm for the Set Cover problem with factor $\ln(\Delta)$, where Δ is the maximum size of the subsets in the covering family [6]. Since in above reduction, the size of each covering subset is at most m, applying this $\ln(\Delta)$-approximation algorithm results in an $\ln(m)$-approximation algorithm for MinCSBC.

□

Theorem 3. *There is a $2f$-approximation algorithm for MinCSBC in one dimension.*

Proof. First, find a set I of unit intervals, with $|I| \leq n$, satisfying the following two conditions.

- All of the n points are covered.
- No two intervals in I intersect, i.e., all the intervals are disjoint.

Then, consider the following problem which is similar to MinCSBC but with an additional restriction.

The modified MinCSBC Problem: Find a subset of I with minimum size that covers at least one point of each color.

Lemma 1. $|S_I| \leq 2|U(OPT_{min})|$*, where* S_I *is the optimal solution to the modified MinCSBC problem with respect to* I*.*

Proof. Since intervals lie on x-axis, for each $u \in U(OPT_{min})$, there exists $I' \subseteq I$ with $|I'| \leq 2$ such that I' covers all points covered by u. Consequently, by replacing each interval in $U(OPT_{min})$ with at most two intervals in I, one can obtain a family $S' \subseteq I$ of intervals covering all points that are covered by $U(OPT_{min})$. Since $|S_I| \leq |S'|$ and $|S'| \leq 2|U(OPT_{min})|$, we obtain that $|S_I| \leq 2|U(OPT_{min})|$. □

As a consequence of Lemma 1, any f-approximation algorithm for modified MinCSBC yields an approximation algorithm for MinCSBC with factor $2f$.

Note that, an instance of the modified MinCSBC problem can be considered as an instance of the Set Cover problem. So, the f-approximation algorithm for Set Cover leads to an $2f$-approximation algorithm for MinCSBC.

□

4 MaxCSBC

4.1 Hardness of MaxCSBC

Theorem 4. *MaxCSBC is NP-hard.*

Proof. We show that MaxCSBC is NP-hard even in one dimension for $f > 2$ by reduction from a restricted version of normal CNF 3SAT in which each variable occurs at most twice in positive form and once in negative form. This problem which we name *3-Occurrence SAT* is known to be NP-complete [1].

[1] It is worth mentioning that in 3-Occurrence SAT problem, if each clause has to have exactly 3 distinct variables, the formula is always satisfiable and thus, the problem is not hard anymore. However, we allow clauses to have less than 3 variables, see [9] for more details.

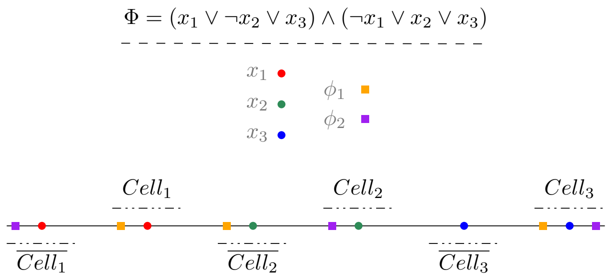

Fig. 3. Reduction to an instance of MaxCSBC.

Given an instance of 3-Occurrence SAT, for each variable x_i, specify two disjoint segments $Cell_i$ and $\overline{Cell_i}$, each of length 3. These two segments correspond to x_i and $\overline{x_i}$, respectively. The cells have to be placed in a way that no two cells intersect. Next, assign distinct colors to each variable and each clause. Denote the color assigned to variable x_i by c_i and the color which is corresponding to clause ϕ_j by c'_j.

For each variable x_i, place two points colored with c_i at the middle of $Cell_i$ and $\overline{Cell_i}$. Since these two points are the only points which are colored with c_i, any color selection must include at least one of them. Selecting the middle point of a cell is interpreted as setting the corresponding literal to 0. In other words, if the middle point of $Cell_i$ is selected, then $x_i = 0$, while selecting the middle point of $\overline{Cell_i}$ means that $\overline{x_i} = 0$ or, equivalently, $x_i = 1$.

Next, for each clause ϕ_j, place three points colored with c'_j in the cells corresponding to its literals at the distance of $\frac{3}{4}$ from the middle. Note that at most two clause-points are placed in the same cell (by the definition of 3-Occurrence SAT). If two points corresponding to different clauses are placed in the same cell, they have to be placed at the different sides of the middle point. See the example depicted in Figure 3.

Lemma 2. *The instance of 3-Occurrence SAT is satisfiable if and only if there exists a color selection for the corresponding MaxCSBC instance in which the distance between any pair of points is greater than* 1*.*

Proof. Suppose that a color selection S exists with the distance between each two points in S more than 1. For each color c_i, if the middle point of $Cell_i$ is in S, set $x_i = 0$, otherwise (i.e., if the middle point of $\overline{Cell_i}$ is in S) set $\overline{x_i} = 0$. Note that for each clause ϕ_j, there is one point p_j of color c'_j in S. Since the distance between p_j and the middle point of the cell that p_j lies in, is less than 1, this middle point cannot be in S, and so there exists a literal in ϕ_j whose value is 1.

On the other hand, we prove that any satisfying assignment for the 3-Occurrence SAT instance can result in a color selection in which the distance between any pair of points is greater than 1. For each variable x_i, the middle point of either $Cell_i$ or $\overline{Cell_i}$ should be chosen in order to have a color selection. If $x_i = 0$, select the middle point of $Cell_i$. Otherwise, select the middle point of

$\overline{Cell_i}$. Since for each clause ϕ_j, there is at least one literal in ϕ_j satisfying it, a cell containing a point with color c'_j exists whose middle point is not selected, so it is possible to select a point of color c'_j.

□

Observation 2 *In an instance of MaxCSBC,* $|U(OPT_{max})| = m$ *if and only if there exists a color selection in which the distance between any pair of points is greater than* 1*.*

Taking into account Observation 2, we can claim that the instance of 3-Occurrence SAT is satisfiable if and only if $|U(OPT_{max})| = m$. Notice that $|U(OPT_{max})|$ is never strictly larger than m.

□

4.2 Approximation Algorithm for MaxCSBC

Now, we present an $O(n \log n)$-time approximation algorithm with factor $\frac{1}{2}$ for MaxCSBC in one dimension.

Algorithm 1 Approximation Algorithm for MaxCSBC

Input: A set P of n points colored with m colors
Output: A color selection of P

$\mathcal{M} = \emptyset,\ T = \emptyset,\ T' = \emptyset$
while $|\mathcal{M}| < n$ **do**
  p = the leftmost point in $P \setminus \mathcal{M}$
  $\mathcal{M} = \mathcal{M} \cup \{p\}$
  $T = T \cup \{p\}$
  for each point $q \in P \setminus \mathcal{M}$ with the same color as p **do**
    $\mathcal{M} = \mathcal{M} \cup \{q\}$
  end for
  for each point $q \in P \setminus \mathcal{M}$ where $dist(p, q) \leq 1$ **do**
    $\mathcal{M} = \mathcal{M} \cup \{q\}$
  end for
end while
for each color c with no candidate in T **do**
  insert an arbitrary point of color c in T'
end for
return $T \cup T'$

Theorem 5. *Algorithm 1 is a* $\frac{1}{2}$*-approximation algorithm.*

Proof. Clearly $T \cup T'$ needs at least $|T|$ unit intervals to be covered since the distance between any two points in T is greater than 1. By Observation 1, all the intervals in $U(OPT_{max})$ are disjoint and the left endpoint of any interval in $U(OPT_{max})$ is one of the input points. Let $\mathcal{T}$ be a set of these points. We claim that $|T| \geq \frac{|\mathcal{T}|}{2}$.

To this end, we show that by adding p to T, at most two points of $\mathcal{T}$ that have been left *unmarked*[2] yet, can be inserted to $\mathcal{M}$. Note that when we add point p to T,

– only one of the points in $\mathcal{T}$ can be of the same color with p because all points in $\mathcal{T}$ have different colors;
– there is at most one unmarked point in $\mathcal{T}$ within the distance at most 1 to p as any two points in $\mathcal{T}$ are within the distance greater than 1. Recall that p is the left-most unmarked point, so all points at the left-hand side of p have been already marked.

Therefore, by adding p to T, at most two unmarked points of $\mathcal{T}$ might be inserted to $\mathcal{M}$. At the end of the algorithm, all points in $\mathcal{T}$ are marked, so $|T| \geq \frac{|\mathcal{T}|}{2}$. Thus the output of Algorithm 1 is within a factor $\frac{1}{2}$ of the optimal solution.

□

5 Conclusion

In this paper, we investigated on the problem of unit covering in the color-spanning set model.

For MinCSBC, we showed the NP-hardness and also hardness of approximation within any constant factor. In addition, we presented an $\ln(m)$-approximation algorithm for this problem and also an approximation algorithm for one-dimensional case with factor $2f$. While one-dimensional MinCSBC is NP-hard even when $f = 2$, the latter algorithm results in a constant-factor approximation algorithm for fixed f.

For MaxCSBC, we proved the NP-hardness and proposed an approximation algorithm with constant factor 2 when $d = 1$.

Here are some open questions.

1. Is there any algorithm with approximation factor better than $2f$ for MinCSBC? For special case when $f = 2$, the proposed algorithm leads to a 4-approximation algorithm. In this case ($f = 2$), a reduction from the Vertex Cover problem shows that assuming the Unique Game Conjecture, the problem does not admit any approximation algorithm with a factor better than 2 [7]. There is still a gap between these lower bounds and our factor, however.
2. Is there any approximation algorithm for MinCSBC and MaxCSBC in higher dimensions?
3. Having considered our reduction from 3-Occurrence SAT, we showed that MaxCSBC is NP-hard for $f > 2$ even in one-dimensional case, but the complexity of the problem for $f = 2$ is still unknown.

[2] A point is *marked* if it is in set $\mathcal{M}$ and *unmarked* otherwise.

References

1. Francisco Claude, Gautam K Das, Reza Dorrigiv, Stephane Durocher, Robert Fraser, Alejandro López-Ortiz, Bradford G Nickerson, and Alejandro Salinger. An improved line-separable algorithm for discrete unit disk cover. *Discrete Mathematics, Algorithms and Applications*, 2(01):77–87, 2010.
2. Gautam K Das, Robert Fraser, Alejandro Lòpez-Ortiz, and Bradford G Nickerson. On the discrete unit disk cover problem. In *WALCOM: Algorithms and Computation*, pages 146–157. Springer, 2011.
3. Robert J Fowler, Michael S Paterson, and Steven L Tanimoto. Optimal packing and covering in the plane are np-complete. *Information processing letters*, 12(3):133–137, 1981.
4. Stefan Funke, Alex Kesselman, Fabian Kuhn, Zvi Lotker, and Michael Segal. Improved approximation algorithms for connected sensor cover. *Wireless networks*, 13(2):153–164, 2007.
5. Dorit S Hochbaum and Wolfgang Maass. Approximation schemes for covering and packing problems in image processing and vlsi. *Journal of the ACM (JACM)*, 32(1):130–136, 1985.
6. David S Johnson. Approximation algorithms for combinatorial problems. In *Proceedings of the fifth annual ACM symposium on Theory of computing*, pages 38–49. ACM, 1973.
7. Subhash Khot and Oded Regev. Vertex cover might be hard to approximate to within 2- ε. *Journal of Computer and System Sciences*, 74(3):335–349, 2008.
8. Maarten Löffler. Data imprecision in computational geometry. 2009.
9. Craig A. Tovey. A simplified np-complete satisfiability problem. *Discrete Applied Mathematics*, 8(1):85 – 89, 1984.
10. Dejun Yang, Satyajayant Misra, Xi Fang, Guoliang Xue, and Junshan Zhang. Two-tiered constrained relay node placement in wireless sensor networks: efficient approximations. In *Sensor Mesh and Ad Hoc Communications and Networks (SECON), 2010 7th Annual IEEE Communications Society Conference on*, pages 1–9. IEEE, 2010.