Computing Minimum Diameter Color-Spanning Sets

Rudolf Fleischer* and Xiaoming Xu**

School of Computer Science and IIPL, Fudan University, Shanghai, China
{rudolf,xuxiaoming}@fudan.edu.cn

Abstract. We study the minimum diameter color-spanning set problem which has recently drawn some attention in the database community. We show that the problem can be solved in polynomial time for L_1 and L_∞ metrics, while it is NP-hard for all other L_p metrics even in two dimensions. However, we can efficiently compute a constant factor approximation.

1 Introduction

Assume we have a set of resources of one of several different types, or colors. We want to solve a task that requires simultaneous use of one resource of each color. There is a communication delay between any pair of resources. How should we allocate the resources so as to minimize the maximum delay between any two of our selected resources? We call this the *minimum diameter color-spanning set problem (*`MDCS`*)*. It arises in large computer networks with different types of servers (think of a large company trying to pool resources to solve a certain computational task). It also arises in spatial databases, where it has recently been studied by Zhang et al. [7]. For example, we may search for a holiday location that features skiing, sailing, golfing, and shopping, all within short distance of each other and of our hotel.

Modeling the Problem. We model `MDCS` as follows. We are given a set S of n points in d-dimensional space $\mathbb{R}^d$. We measure distances in the L_p metric, for some $1 \leq p \leq \infty$. Each point is colored in one of k colors, where $k \geq 1$. S may be a multiset, which means we can have scenarios where a point is colored simultaneously with several colors. We call a subset of k points of distinct colors a *rainbow set*. `MDCS` is the problem of finding a rainbow set of smallest diameter. If we want to emphasize the dimension and metric, we write `MDCS`(d, p). We denote the smallest diameter of a rainbow set of S in L_p metric by $r_p(S)$. See Fig. 1 for an example.

* This work was supported by a grant from the NSFC (no. 60973026), the Shanghai Leading Academic Discipline Project (no. B114), and the Shanghai Committee of Science and Technology (nos. 08DZ2271800 and 09DZ2272800).

** The order of authors follows the international standard of alphabetic order of the family name; first-author order would be Xu and Fleischer.

D.T. Lee, D.Z. Chen, and S. Ying (Eds.): FAW 2010, LNCS 6213, pp. 285–292, 2010.

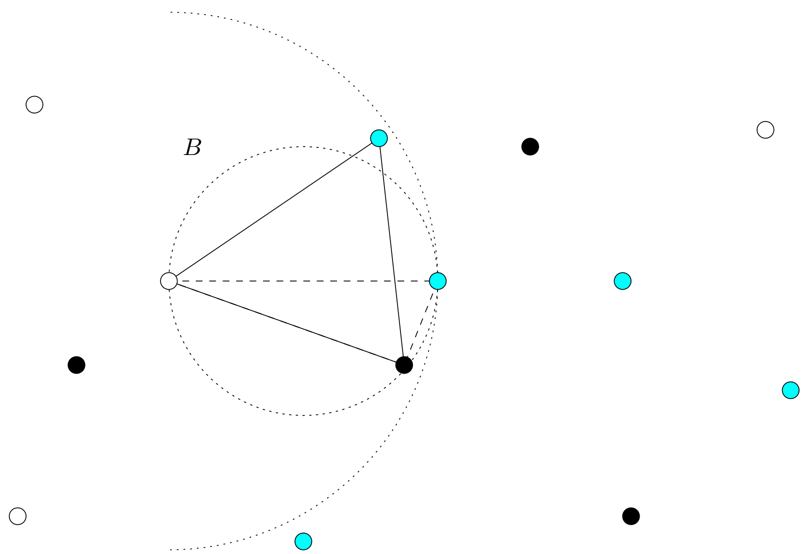

Fig. 1. An instance of $\texttt{MDCS}(2, 2)$ with 3 colors. The smallest color-spanning disc B, enclosing the dashed triangle, does not minimize the diameter of a rainbow set, but it is a good approximation (Lemma 4). The solid triangle spans the minimum-diameter rainbow set.

Previous Work. Zhang et al. proposed an $O(n^k)$-time algorithm for $\texttt{MDCS}(2, 2)$ based on a brute-force enumeration of all possible rainbow sets. Their algorithm was implemented in a geographical tagging system named MarcoPolo by Chen et al. [3]. They also list several other applications of this problem.

Our Results. It is straightforward to solve the problem for $d = 1$ in $O(n \log n)$ time. On the other hand, we show that $\texttt{MDCS}(2, 2)$ is NP-hard for $d \geq 2$ and $1 < p < \infty$. We do not know whether the problem is $W[1]$-hard or APX-hard for these metrics, but there are efficient approximation algorithms. For example, we can approximate $r_2(S)$ by a factor of 1.154 in time $O(n \log n)$ in two dimensions. For $p \in \{1, \infty\}$, we can solve $\texttt{MDCS}(2, p)$ optimally in time $O(n \log n)$, and $\texttt{MDCS}(3, p)$ in time $O(k^{1+\epsilon} n^2)$, for any $\epsilon > 0$. In $\mathbb{R}^d$, the running time is $O(n^{d+2})$.

Related Work. To the best of our knowledge, this problem has not been studied from a theoretical point of view. Several other color-spanning set problems have been studied by Abellanas et al. They called rainbow sets *color-spanning sets*. They gave efficient algorithms in two dimensions for the smallest color-spanning disc problem [1] (finding a smallest disc containing at least one point of each color) and the smallest color-spanning rectangle problem [2] (finding a smallest rectangle containing at least one point of each color). We denote the diameter of a smallest color-spanning disc (or ball in higher dimensions) of a k-colored set S by $b_p(S)$.

In two dimensions, the smallest color-spanning disc for any L_p metric can be computed in time $O(kn \log n)$ [1]. This algorithm uses a farthest color Voronoi diagram which can be computed using an algorithm for the computation of the upper envelope of Voronoi surfaces in one dimension higher [5, Theorem 8]. In three dimensions, the running time becomes $O(k^{1+\epsilon}n^2)$, for any $\epsilon > 1$ [5, Theorem 19]. To generalize these algorithms to higher dimensions would require to bound the complexity of the envelopes of higher-dimensional Voronoi surfaces, which required considerable effort even in two and three dimensions. However, we can always find the smallest enclosing ball in time $O(n^{d+2})$ by brute-force enumeration of all balls with some subset of $d+1$ points on their boundary.

Structure of the Paper. In Section 2, we state a few facts about smallest enclosing balls in higher dimensions. In Section 3, we present efficient algorithms for $\texttt{MDCS}(d, 1)$ and $\texttt{MDCS}(d, \infty)$ and approximation algorithms for the other L_p metrics. In Section 4, we show that $\texttt{MDCS}(2, p)$ is NP-complete for $1 < p < \infty$. We conclude the paper in Section 5.

2 Preliminaries

We first state a few facts about smallest enclosing balls of point sets in higher dimensions. For a point set S, we denote its diameter in L_p metric by $diam_p(S)$ and its smallest enclosing ball by $B_p(S)$.

As we will see later, the worst case ratio of the diameter of the smallest enclosing ball of a point set and the diameter of the point set determines the approximation ratio of our algorithms. We denote this ratio by α_p^d in d-dimensional space with L_p metric. The following proposition is illustrated in Fig. 2.

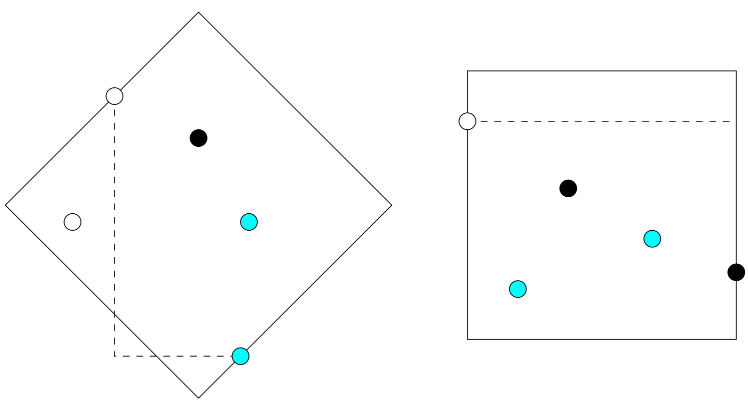

Fig. 2. The smallest spanning-color discs in L_1 metric (left) and L_∞ metric (right) for three-colored point sets. In both cases, the length of the dashed line (the distance of the two points on the boundary of the square) is equal to the diameter of the disc.

Proposition 1. *Let S be a point set in $\mathbb{R}^d$. For $p \in \{1, \infty\}$, $B_p(S)$ is a hypercube, and $diam_p(S) = diam_p(B_p(S))$, i.e., $\alpha_1^d = \alpha_\infty^d = 1$.* □

Proposition 2. *For $1 < p < \infty$, any two points on the boundary of a ball B in L_p metric have distance at most $diam_p(B)$, with equality if and only if the two points are antipodal.* □

Corollary 3. *Let S be a k-colored point set in $\mathbb{R}^d$. For any p, $r_p(S) \le b_p(S)$.* □

3 Approximating MDCS

Let S be a k-colored set of points in $\mathbb{R}^d$ with L_p metric. In this section we will show that the minimum diameter rainbow set can be approximated by computing the minimum color-spanning ball. Since we can compute $b_p(S)$ efficiently, this gives us polynomial-time approximation algorithms for $\texttt{MDCS}(d, p)$.

Lemma 4. *In any dimension d with any L_p metric, the smallest color-spanning ball is an α_p^d-approximation for the minimum diameter rainbow set.*

Proof. Let R be a minimum diameter rainbow set of S. Let B be a smallest enclosing ball of R. Then, by Cor. 3,

$$b_p(S) \ge r_p(S) = diam_p(R) \ge \frac{diam(B)}{\alpha_p^d} \ge \frac{b_p(S)}{\alpha_p^d} .$$ □

We can therefore now focus on determining better bounds for α_p^d. We have already seen in Prop. 1 that $\alpha_1^d = \alpha_\infty^d = 1$. We can therefore solve MDCS optimally by computing a smallest color-spanning ball.

Theorem 5. *We can solve $\texttt{MDCS}(d, 1)$ and $\texttt{MDCS}(d, \infty)$ by computing a smallest color-spanning ball of S.* □

For other L_p metrics, the minimum-color spanning ball is at least a 2-approximation for the minimum diameter rainbow set. It does not necessarily minimize the diameter, as can be seen in Fig. 1.

Theorem 6. *For any d and p, $\alpha_p^d \le 2$.*

Proof. Let B be a smallest enclosing ball of a point set R. Then there must be two points in R with distance at least $\frac{1}{2} diam_p(B)$. Otherwise, R would be completely contained in a half-ball of B, but then B would not be the smallest enclosing ball of R. □

For L_2 metric, we have a stronger bound.

Theorem 7. $\alpha_2^d \le \sqrt{\frac{2d}{d+1}}$. □

Proof. In L_2 metric, the smallest enclosing ball B of a regular simplex in $\mathbb{R}^d$ with unit edges is a *universal cover* [6], i.e., it contains *any* point set of diameter 1. B has diameter $\sqrt{\frac{2d}{d+1}}$. □

Corollary 8. *In two dimensions with L_2 metric, we can find a $\frac{2}{\sqrt{3}} \approx 1.154$-approximation to $diam_p(S)$ in time $O(kn \log n)$. In three dimensions, we can find a $\sqrt{\frac{3}{2}} \approx 1.225$-approximation to $diam_p(S)$ in time $O(k^{1+\epsilon} n^2)$, for any $\epsilon > 1$. In higher dimensions, the approximation factor is never more than $\sqrt{2} \approx 1.414$.* □

In two dimensions, we can improve the bound for α_p^2.

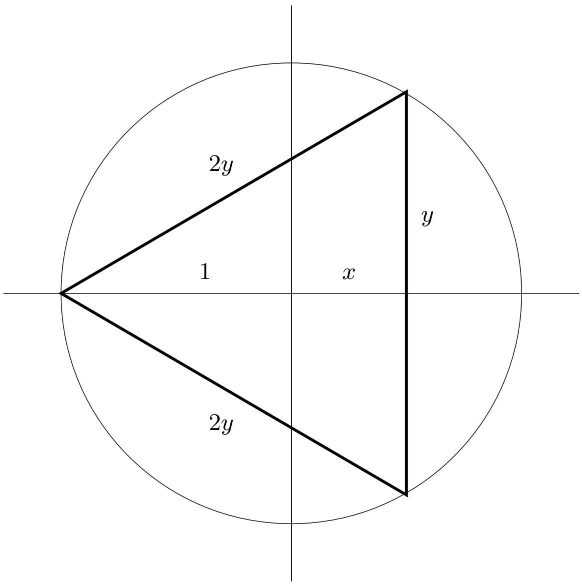

Fig. 3. An equilateral triangle with side length $2y$ in L_p metric in a ball of diameter 2

Theorem 9. *For $2 \leq p \leq \infty$, $\alpha_p^2 < \frac{4}{3}$.*

Proof. By Prop. 1, $\alpha_\infty^d = 1$, so we can assume that $2 \leq p < \infty$. It is straightforward to adapt the proof in [6] to L_p metrics to show that the smallest enclosing ball B of a regular simplex in $\mathbb{R}^d$ with unit edges is a universal cover. But we do not have a closed formula for the diameter of that ball. Fig. 3 shows a ball of diameter 2 in $\mathbb{R}^2$ with an inscribed equilateral triangle of side length $2y$ (the figure assumes L_2 metric, but the figure would be simular for any L_p metric, $2 \leq p < \infty$). We have two constraints for x and y:

$$x^p + y^p = 1 \tag{1}$$

$$\text{and} \quad (1+x)^p + y^p = (2y)^p \,. \tag{2}$$

Substituting $y^p = 1 - x^p$ in Eq. (2) yields

$$(1+x)^p + (2^p - 1)x^p = 2^p - 1 \,.$$

Note that the left-hand side of this equation is monotone increasing in x. If $p \geq 2$, then it is not larger than the right-hand-side if we set $x = \frac{1}{2}$. Thus, $x \geq \frac{1}{2}$. Then, Eq. 2 implies

$$y^p = \frac{(1+x)^p}{2^p - 1} > \left(\frac{3}{4}\right)^p \,.$$

Thus, $\alpha_p^2 = \frac{1}{y} < \frac{4}{3}$. □

4 Hardness of MDCS

We do not know whether the approximation algorithms in the previous section are optimal (we suspect they are not) or whether there exists a PTAS, but we cannot hope to solve the problem exactly in polynomial time because we will now show that MDCS is NP-hard.

We first give the *decision version* of MDCS: Given an instance of MDCS and a positive number d, decide whether there exists a rainbow set of diameter at most d. Clearly, this problem is in NP (we can compare the square of all pairwise point distances to d^2, avoiding costly square root calculations). Hardness of the decision problem implies hardness of the optimization problem.

Theorem 10. *The decision version of* MDCS *is NP-hard for L_p metric, for $1 < p < \infty$, in two or higher dimensions.*

Proof. We prove the hardness of MDCS by reduction from $3SAT$. We give the proof for L_2 metric in two dimensions and then show how to extend it to other L_p metrics. This implies hardness for higher dimensions.

We first sketch the proof under the assumption that we can easily compute coordinates of points on a circle. We will then show how to approximate these coordinates with low precision rational numbers.

Let F be a Boolean formula in conjunctive normal form with n variables $x_1, \ldots, x_n$ in m clauses $c_1, \ldots, c_m$ of size at most three. To construct an instance I of MDCS, we draw a circle C with diameter 1. For each variable x_i, we define two antipodal slots s_i and $\bar{s}_i$ on C, corresponding to the positive literal x_i and the negative literal $\bar{x}_i$, respectively. These slots should be pairwise distinct, but otherwise they can be placed arbitrarily; for example, we could create $2n$ equally-spaced slots on C.

For each clause c_j we create a new color col_j. If x_i appears in c_j, we place a point of color col_j at slot s_i Similarly, if $\bar{x}_i$ appears in c_j, we place a point of color col_j at slot $\bar{s}_j$. Note that several points can coincide if a literal appears in several clauses. Finally, we set $d = 1 - \epsilon$ for some ϵ defined below. See Fig. 4 for an example of the construction.

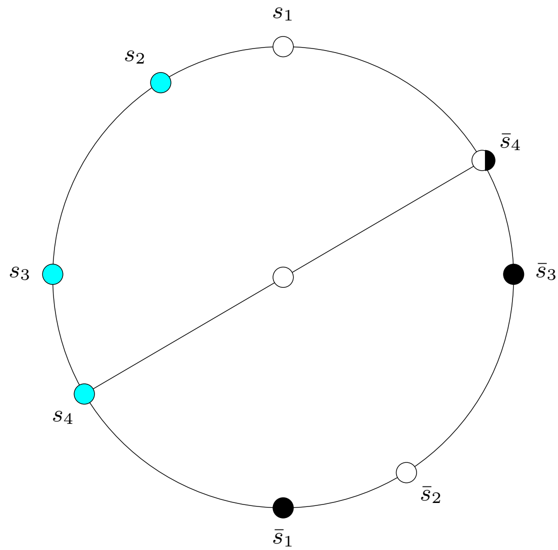

Fig. 4. An example for the construction in the NP-hardness reduction. Let $F = (x_1 \wedge \bar{x}_2 \wedge \bar{x}_4) \vee (x_2 \wedge x_3 \wedge x_4) \vee (\bar{x}_1 \wedge \bar{x}_3 \wedge \bar{x}_4)$. The literals of the first clause are colored white, the second clause colored cyan, and the third clause colored black. Note that a white and black point share slot $\bar{s}_4$.

If I has a solution, then there is a rainbow set R of m points with diameter at most $d = 1 - \epsilon$. In particular, R cannot contain both s_i and $\bar{s}_i$, for any i. Therefore, R induces a truth assignment for the variables x_i. If $s_i \in R$, we set $x_i = 1$, otherwise we set $x_i = 0$. Since R contains one point of each color, each clause will contain at least one true literal, i.e., F will be satisfied.

If F admits a satisfying assignment, then each clause c_j contains at least one true literal x_{i_j} (or $\bar{x}_{i_j}$). We add the corresponding point of color col_j at slot s_{i_j} (or $\bar{s}_{i_j}$) to the set R. Then, R is a rainbow set of diameter at most d. Thus, I has a solution.

We now discuss the problem of approximating the slot coordinates with low precision rationals. For example, if the slots are evenly spaced around the circle, we can set ϵ to be any value strictly smaller than $\frac{\pi}{n}$. If we chose $\epsilon < \frac{\pi}{2n}$, we can approximate the slot positions by choosing an arbitrary point inside a ball centered at the slot with diameter at most $\frac{\epsilon}{2}$.

For other L_p metrics, $1 < p < \infty$, observe that any two points on the boundary of a unit disc will have distance at most 1, with equality if and only if the two points are antipodal. Therefore, the proof above also works for arbitrary L_p metrics, except if $p = 1$ and $p = \infty$, since for these metrics the discs are actually squares and any two points on opposite sides of the square have distance equal to the diameter of the disc. $\square$

5 Conclusions

We have shown that MDCS can easily be solved for L_1 and L_∞ metrics, while it is NP-hard for other L_p metrics. Unfortunately, the approximation ratio of the smallest color-spanning disc deteriorates if the dimension increases. It would therefore be interesting to find better approximation algorithms for MDCS, in particular in higher dimensions. It may also be worthwhile to study FPT algorithms for MDCS, for example with parameter k, the number of colors. The problem is clearly polynomial-time for two colors, and our NP-hardness reduction required a large number of colors.

Acknowledgments

We thank the reviewers for their helpful suggestions how to improve the presentation of the paper.

References

1. Abellanas, M., Hurtado, F., Icking, C., Klein, R., Langetepe, E., Ma, L., Palop, B., Sacristan, V.: The farthest color Voronoi diagram and related problems. In: Proceedings of the 17th European Workshop on Computational Geometry (EWCG 2001), pp. 113–116 (2001)
2. Abellanas, M., Hurtado, F., Icking, C., Klein, R., Langetepe, E., Ma, L., Palop, B., Sacristan, V.: Smallest color-spanning objects. In: Meyer auf der Heide, F. (ed.) ESA 2001. LNCS, vol. 2161, pp. 278–289. Springer, Heidelberg (2001)
3. Chen, Y., Chen, S., Gu, Y., Hui, M., Li, F., Liu, C., Liu, L., Ooi, B.C., Yang, X., Zhang, D., Zhou, Y.: MarcoPolo: A community system for sharing and integrating travel information on maps. In: Proceedings of the 12th International Conference on Extending Database Technology (EDBT 2009), pp. 1148–1151 (2009)
4. El-Gebeily, M.A., Fiagbedzi, Y.A.: On certain properties of the regular n-simplex. International Journal of Mathematical Education in Science and Technology 35(4), 617–629 (2004)
5. Huttenlocher, D.P., Kedem, K., Sharir, M.: The upper envelope of voronoi surfaces and its applications. Discrete & Computational Geometry, 267–291 (1993)
6. Jung, H.: Über die kleinste Kugel, die eine räumliche Figur einschließt. Journal für die reine und angewandte Mathematik 123, 232–257 (1901)
7. Zhang, D., Chee, Y.M., Mondal, A., Tung, A.K.H., Kitsuregawa, M.: Keyword search in spatial databases: Towards searching by document. In: Proceedings of the 25th IEEE International Conference on Data Engineering (ICDE 2009), pp. 688–699 (2009)