Theoretical Computer Science 609 (2016) 561–568

Contents lists available at ScienceDirect

Theoretical Computer Science

www.elsevier.com/locate/tcs

Shortest color-spanning intervals

Minghui Jiang *, Haitao Wang [1]

Department of Computer Science, Utah State University, Logan, UT 84322, USA

ARTICLE INFO

Article history:
Received 18 August 2014
Received in revised form 13 January 2015
Accepted 17 January 2015
Available online 22 January 2015

Keywords:
Color-spanning objects
Computational geometry
Exact algorithms
Parameterized complexity

ABSTRACT

Given a set of n points on a line, where each point has one of k colors, and given an integer $s_i \geq 1$ for each color i, $1 \leq i \leq k$, the problem Shortest Color-Spanning t Intervals (SCSI-t) aims at finding t intervals to cover at least s_i points of each color i, such that the maximum length of the intervals is minimized. Chen and Misiolek introduced the problem SCSI-1, and presented an algorithm running in $O(n)$ time if the input points are sorted. Khanteimouri et al. gave an $O(n^2 \log n)$ time algorithm for the special case of SCSI-2 with $s_i = 1$ for all colors i. In this paper, we present an improved algorithm with running time of $O(n^2)$ for SCSI-2 with arbitrary $s_i \geq 1$. We also obtain some interesting results for the general problem SCSI-t. From the negative direction, we show that approximating SCSI-t within any ratio is NP-hard when t is part of the input, is W[2]-hard when t is the parameter, and is W[1]-hard with both t and k as parameters. Moreover, the NP-hardness and the W[2]-hardness with parameter t hold even if $s_i = 1$ for all i. From the positive direction, we show that SCSI-t with $s_i = 1$ for all i is fixed-parameter tractable with k as the parameter, and admits an exact algorithm running in $O(2^k n \cdot \max\{k, \log n\})$ time.

1. Introduction

Given a set of n points on a line, where each point has one of k colors, and given an integer $s_i \geq 1$ for each color i, $1 \leq i \leq k$, the problem Shortest Color-Spanning t Intervals (SCSI-t) aims at finding t intervals to cover at least s_i points of each color i, such that the maximum length of the intervals is minimized.

Chen and Misiolek [3] introduced the problem SCSI-1, and presented an algorithm running in $O(n)$ time if the input points are sorted. Khanteimouri et al. [13] gave an $O(n^2 \log n)$ time algorithm for the special case of SCSI-2 with $s_i = 1$ for all colors i. Our first result in this paper is an improved algorithm for SCSI-2 with arbitrary $s_i \geq 1$:

Theorem 1. *SCSI-2 admits an exact algorithm running in $O(n^2)$ time.*

The problems SCSI-1 and SCSI-2 naturally generalize to SCSI-t for $t \geq 1$. Our next theorem shows that SCSI-t is intractable in a very strong sense:

Theorem 2. *Approximating SCSI-t within any ratio is NP-hard when t is part of the input, is W[2]-hard when t is the parameter, and is W[1]-hard with both t and k as parameters. Moreover, the NP-hardness and the W[2]-hardness with parameter t hold even if $s_i = 1$ for all i.*

* Corresponding author.
E-mail addresses: mjiang@cc.usu.edu (M. Jiang), haitao.wang@usu.edu (H. Wang).

[1] Supported in part by NSF under Grant CCF-1317143.

http://dx.doi.org/10.1016/j.tcs.2015.01.039
0304-3975/

Optimization problems that are hard to approximate within any ratio are no longer a novelty. A recent example is the exemplar distance problem in comparative genomics; see [11] and the references therein. The study of intractability combining both parameterized complexity and approximation hardness is not new either; see e.g. [15]. But to our best knowledge, SCSI-t is the first natural problem that is known to be intractable in the special way that obtaining any approximation is W[2]-hard.

In contrast to the very negative result in Theorem 2, our following theorem shows that the special case of SCSI-t with $s_i = 1$ for all i is fixed-parameter tractable when the parameter is the number k of colors:

Theorem 3. *The special case of SCSI-t with $s_i = 1$ for all i admits an exact algorithm running in $O(2^k n \cdot \max\{k, \log n\})$ time.*

In particular, we can solve SCSI-t with $s_i = 1$ for all i in $O(n \log n)$ time if k is a constant, and in polynomial time if $k = O(\log n)$. Thus the problem SCSI-t may still be manageable in practice.

1.1. Related work

Instead of finding t intervals to cover at least $s_i \geq 1$ points of each color i as in SCSI-t, another generalization of the problem SCSI-1 aims at finding one geometric object to cover at least $s_i \geq 1$ points of each color i in the plane rather than on a line. This planar problem is typically studied with $s_i = 1$ for all colors i. Abellanas et al. [1] proposed an $O(n(n-k)\log^2 k)$ time algorithm for computing the smallest (by perimeter or area) axis-parallel rectangle that contains at least one point of each color. Das et al. [6] gave an improved algorithm with $O(n(n-k)\log k)$ time for this problem, and moreover gave an $O(n^3 \log k)$ time algorithm for computing the smallest color-spanning rectangle of arbitrary orientation. Algorithms for computing the smallest color-spanning strips were also given in [1,6]. Recently, Khanteimouri et al. [14] gave an $O(n \log^2 n)$ time algorithm for computing the smallest color-spanning axis-parallel square, and Barba et al. [2] considered the related problem of computing a region (e.g., rectangle, square, or disk) that contains *exactly* s_i points of each color i.

Given a set of colored points, a *color-spanning set* is a subset of the input points including at least one point of each color. The various color-spanning problems for colored points with $s_i = 1$ for all colors i can be viewed as finding a color-spanning set such that certain geometric property of the set is optimized. In this framework, Fleischer and Xu [9,10] gave polynomial time algorithms for finding a minimum-diameter color-spanning set under the L_1 or L_∞ metric, and proved that the problem is NP-hard for all L_p with $1 < p < \infty$. Ju et al. [12] gave an efficient algorithm for computing a color-spanning set with the maximum diameter, and proved that several other problems are NP-hard, e.g., finding the color-spanning set with the largest closest-pair distance. Fan et al. [7] studied the problem of finding a color-spanning set with the minimum connection radius in the corresponding disk intersection graph.

2. An $O(n^2)$-time exact algorithm for SCSI-2

In this section we prove Theorem 1. We present an $O(n^2)$ time algorithm for solving the problem SCSI-2, which improves the $O(n^2 \log n)$ time algorithm in [13].

Let $P = \{p_1, p_2, \ldots, p_n\}$ be a set of n points given on a line L, say, the x-axis, sorted from left to right. Each point p_i has one of k colors. A line segment on L is also called an *interval* of L. We say an interval of L *covers* a point if the point is on the interval. The problem SCSI-2 is to find two intervals on L to cover at least s_i points of each color i with $1 \leq i \leq k$ such that the maximum length of the intervals is minimized. In the following, we assume that for any i, the number of points of color i in P is at least s_i, since otherwise there would be no solution for the problem.

If two intervals of L together cover at least s_i points of each color i in P, then we say the two intervals form a *feasible solution* for SCSI-2. For any interval I, let $d(I)$ denote the length of I. An interval I_1 is said to be *longer* than another interval I_2 if and only if $d(I_1) \geq d(I_2)$. We first prove the following lemma:

Lemma 1. *There must exist an optimal solution for the problem SCSI-2 that consists of two intervals such that the longer interval has both left and right endpoints in P.*

Proof. Consider any optimal solution for SCSI-2 that consists of two intervals I_1 and I_2. If both the left and right endpoints of both I_1 and I_2 are in P, then we are done with the proof. Otherwise, without loss of generality, assume the left endpoint of I_1 is not at any point of P. Then, we can shrink I_1 by moving its left endpoint rightwards for an infinitesimal distance such that the new interval I'_1 covers the same subset of points of P as I_1 does (e.g., see Fig. 1). Clearly, I'_1 and I_2 together still form a feasible solution.

If some endpoints of I'_1 and I_2 are not in P, then we use the same technique as above to shrink them. Eventually, we can obtain two intervals I''_1 and I''_2 whose endpoints are all in P and they form a feasible solution. Since $d(I''_1) \leq d(I_1)$, $d(I''_2) \leq d(I_2)$, and I_1 and I_2 form an optimal solution, the two new intervals I''_1 and I''_2 must also form an optimal solution. The lemma thus follows. □