An improved lower bound
for one-dimensional online unit clustering

Jun Kawahara\textsuperscript{1} and Koji M. Kobayashi\textsuperscript{2}
\textsuperscript{1}Graduate School of Information Science, Nara Institute of Science and Technology
\textsuperscript{2}National Institute of Informatics

Abstract

The online unit clustering problem was proposed by Chan and Zarrabi-Zadeh (WAOA2007 and Theory of Computing Systems 45(3), 2009), which is defined as follows: “Points” are given online in the $d$-dimensional Euclidean space one by one. An algorithm creates a “cluster,” which is a $d$-dimensional rectangle. The initial length of each edge of a cluster is 0. An algorithm can extend an edge until it reaches unit length independently of other dimensions. The task of an algorithm is to cover a new given point either by creating a new cluster and assigning it to the point, or by extending edges of an existing cluster created in past times. The goal is to minimize the total number of created clusters. Chan and Zarrabi-Zadeh proposed some method to obtain a competitive algorithm for the $d$-dimensional case using an algorithm for the one-dimensional case, and thus the one-dimensional case has been extensively studied including some variants of the unit clustering problem.

In this paper, we show a lower bound of $13/8 = 1.625$ on the competitive ratio of any deterministic online algorithm for the one-dimensional unit clustering, improving the previous lower bound $8/5 (= 1.6)$ presented by Epstein and van Stee (WAOA2007 and ACM Transactions on Algorithms 7(1), 2010). Note that Ehmsen and Larsen (SWAT2010 and Theoretical Computer Science, 500, 2013) showed the current best upper bound of $5/3$, and conjectured that the exact competitive ratio in the one-dimensional case may be $13/8$.

1 Introduction

Given some points, we consider the problem of partitioning points into some groups, namely clusters, which is generally called a clustering problem. The goal is to optimize an objective function according to applications. Clustering problems are applied for various purposes such as information retrieval, data mining and facility location. This problem is formulated as online problems, and some variants have been widely studied. The online unit covering problem, proposed by Charikar et al. \cite{Charikar}, is the most basic variant. In this problem, points are given in the $d$-dimensional Euclidean space one by one. An algorithm places balls with unit radius so that they cover points. The goal of the problem is to minimize the number of created balls.

Chan and Zarrabi-Zadeh have introduced the online unit clustering problem \cite{Chan}. In this problem, an input is a sequence of points. Points are given in the $d$-dimensional Euclidean space one by one and must be covered by rectangles, called clusters. Initially, the length of each edge of a cluster is 0, and each edge can be extended until unit length. Once a cluster is placed, it must not be shifted and removed. When a point is given outside existing clusters, the task of an algorithm is to determine whether creating a new cluster or extending some edges of an existing cluster to cover the point before the next one is given. The goal of the problem is to minimize the number of created clusters. The performance of an online algorithm is evaluated using competitive analysis.
For any algorithm $\text{ALG}$ and an input $\sigma$, let $C_{\text{ALG}}(\sigma)$ denote the number of clusters created by $\text{ALG}$. If, for any input $\sigma$, an online algorithm $\text{ON}$ creates clusters at most $c$ of the optimal offline algorithm, called $\text{OPT}$, for $\sigma$, then we say that $\text{ON}$ is $c$-competitive or that the competitive ratio of $\text{ON}$ is at most $c$.

Chan and Zarrabi-Zadeh [2] showed that a $2^{d-1}c$-competitive algorithm can be obtained by using a $c$-competitive algorithm for the one-dimensional case (i.e., the case of $d = 1$), which is the current best algorithm for the $d$-dimensional case when $d \geq 2$. This implies that by improving an algorithm for the one-dimensional case, we can also obtain a better algorithm for the $d$-dimensional case. Thus, the one-dimensional case has been extensively studied including some variants of the unit clustering problem. (The past results are summarized in Table 1.)

### Previous Results and Our Result.

For the one-dimensional case, the best known deterministic upper bound on the competitive ratio is $5/3 \approx 1.667$, presented by Ehmsen and Larsen [5], and the best known lower bound on the competitive ratio of any deterministic online algorithm is $8/5 = 1.6$, shown by Epstein and van Stee [7].

In this paper, for the one-dimensional deterministic case, we improve the lower bound from $8/5$ to $13/8 = 1.625$. (Ehmsen and Larsen [5] conjectured that the exact competitive ratio may be $13/8$.)

### Related Results.

For the two-dimensional case, Epstein and van Stee [7] gave a deterministic lower bound of $2$ and a randomized lower bound of $11/6 \approx 1.833$. Ehmsen and Larsen [5] showed a lower bound of $13/6 \approx 2.166$ for any deterministic algorithm. Epstein et al. [6] and Csirik et al. [4] studied some variants of the one-dimensional online unit clustering.

## 2 Improved Lower Bound

**Theorem 2.1** The competitive ratio of any deterministic online algorithm is at least $13/8$.

**Proof.** Let $\text{ON}$ be any deterministic online algorithm and we give $\text{ON}$ an instance $\sigma$ in Table 2. In the table, column “Point” presents the position of each given new point. “Cluster” shows the name of $\text{ON}$’s cluster assigned to the new point. If there exist other clusters which can be assigned at that time, column “Note” explains how $\text{ON}$ behaves and how much the costs of $\text{ON}$ and $\text{OPT}$ are if the other clusters are used. In addition, “Note” tells that $\text{ON}$ can assign an existing cluster $K$ or $L$ to the last given point and how much the costs are charged at each case. For example, when $\text{ON}$ is given the fourth point at position 6, $\text{ON}$ can choose either to create a new cluster $F$ or to assign the existing cluster $E$. If $\text{ON}$ creates $F$, then the next point arrives at position 2 and

| Years   | Deterministic | Randomized |
|---------|---------------|------------|
| Upper Bound |               |            |
| 2007 [2] | 2             | 15/8 (= 1.875) |
| 2007 [9] | 11/6 (≈ 1.834) |            |
| 2008 [7] | 7/4 (= 1.75)  | 7/4 (= 1.75) |
| 2010 [5] | 5/3 (≈ 1.667) |            |
| Lower Bound |               |            |
| 2015 [this paper] | 13/8 (= 1.625) |           |
| 2008 [7] | 8/5 (= 1.6)   | 3/2 (= 1.5)  |
| 2007 [2] | 3/2 (= 1.5)   | 4/3 (≈ 1.333) |

[1] [8].
the input goes on. Otherwise, three points are given one by one and the cost ratio is 5/3, which is larger than 13/8.

Therefore, we can show that $C_{ON}(\sigma)/C_{OPT}(\sigma) \geq 13/8$.

| Point | Cluster | Note |
|-------|---------|------|
| 3     | D       |      |
| 4     | D       | If $ON$ creates a new cluster, then we have $C_{ON}(\sigma) = 2$ and $C_{OPT}(\sigma) = 1$. |
| 5     | E       |      |
| 6     | F       | If $ON$ assigns $E$ to the given point, then three points are given at 2.5, 4.5 and 6.5 respectively and $C_{ON}(\sigma) = 5$ and $C_{OPT}(\sigma) = 3$ hold. |
| 2     | B       |      |
| 1     | B       | If a new cluster is created, we obtain the cost ratio of 5/3 $\approx$ 1.66. |
| 0     | A       |      |
| 2.5   | C       |      |
| 7     | F       | If $ON$ creates a new one, then the cost ratio becomes 7/4 $= 1.75$. |
| 8     | G       |      |
| 8.5   | H       | If $G$ is assigned to the given point, then four points arrive at 4.5, 5.6, 7.4 and 9.5 respectively and we have the ratio of 10/6 $\approx$ 1.66. ((i) in Fig. 1) |
| 9     | H       | If $ON$ newly creates a cluster, we obtain the cost ratio of 9/5 $= 1.8$. |
| 10    | J       |      |
| 11    | J       | If a new cluster is created, then $C_{ON}(\sigma) = 10$ and $C_{OPT}(\sigma) = 6$ hold. |
| 9.5   | I       | If $ON$ assigns $H$ to the given point, then six points are given at 4.5, 5.6, 7.2, 8.3, 9.7 and 11.5 respectively, and we have the cost ratio of $13/8 = 1.625$. ((ii) in Fig. 1) |
| 12    | K       |      |
| 13    | K       | Two points arrive at 11.5 and 14 respectively and the cost ratio becomes 13/8 $= 1.625$. ((iii) in Fig. 1) |
| L     | Two points are given at 4.5 and 5.6 respectively and the ratio of 13/8 $= 1.625$ is obtained. ((iv) in Fig. 1) |

3 Discussion

Developing an online algorithm $ON$ whose competitive ratio is at most $c$ in the one-dimensional case, we have little other choice than showing that for each $x$ clusters $ON$ creates, $OPT$ necessarily does at least $x/c$ ($= y$) clusters. This also appears in the analyses of the 7/4 and 5/3-competitive algorithms in [7] and [5], respectively. Ehmsen and Larsen [5] stated that if $x \leq 20$ and the optimal competitive ratio for the one-dimensional case is neither 8/5 nor 5/3, then the possible competitive ratio is only 13/8 or 18/11. They conjectured that these values may be optimal. (Note that pairs of integers $(x, y)$ such that $8/5 < x/y < 5/3$ holds are (13, 8), (18, 11), (21, 13), (29, 18), (34, 21),...
in ascending order of $x$.) Generally speaking, we need to carry out a more exhaustive case-analysis as $x$ grows or $c$ decreases. Thus, to obtain $ON$’s competitive ratio of $c = \frac{13}{8}$, i.e., $x = 13$, it seems that more complicated analysis and $ON$’s behaviors are required than $c = \frac{5}{3}$. Specifically, given our lower bound instance, $ON$ must behave in the same way described in Table 2. (Of course $ON$ has to be able to deal with the other inputs as well.)

References

[1] A. Borodin, and R. El-Yaniv, “Online computation and competitive analysis,” Cambridge University Press, 1998.

[2] T.M. Chan, and H. Zarrabi-Zadeh, “A randomized algorithm for online unit clustering,” Theory of Computing Systems, Vol. 45, No. 3, pp. 486–496, 2009.

[3] M. Charikar, C. Chekuri, T. Feder, and R. Motwani, “Incremental clustering and dynamic information retrieval,” SIAM Journal on Computing, Vol. 33, No. 6, pp. 1417–1440, 2004.

[4] J. Csanádi, L. Epstein, C. Imreh, and A. Levin, “Online clustering with variable sized clusters,” Algorithmica, Vol. 65, No. 2, pp. 251–274, 2013.

[5] M. Ehmsen, and K. Larsen, “Better bounds on online unit clustering,” Theoretical Computer Science, Vol. 500, pp. 1–24, 2013.

[6] L. Epstein, A. Levin, and R. van Stee, “Online unit clustering: Variations on a theme,” Theoretical Computer Science, Vol. 407, No. 1-3, pp. 85–96, 2008.
[7] L. Epstein, and R. van Stee, “On the online unit clustering problem,” *ACM Transactions on Algorithms*, Vol. 7, No. 1, 2010.

[8] D. Sleator, and R. Tarjan, “Amortized efficiency of list update and paging rules,” *Communications of the ACM*, Vol. 28, No. 2, pp. 202–208, 1985.

[9] H. Zarrabi-Zadeh, and T.M. Chan, “An improved algorithm for online unit clustering,” *Algorithmica*, Vol. 54, No. 4, pp. 490–500, 2009.