A New Approximation Algorithm for Labeling Weighted Points with Sliding Labels[*]

Thomas Erlebach[†] Torben Hagerup[‡] Klaus Jansen[§] Moritz Minzlaff[¶] Alexander Wolff[‖]

Abstract

This paper presents a polynomial-time approximation algorithm for labeling some of the points in a given set of weighted points with horizontally sliding labels of unit height and given lengths to maximize the weight of the labeled points. The approach is based on a discretization, and two results are established: In general, the algorithm has an approximation factor of $2/3 - \epsilon$, for arbitrary fixed $\epsilon > 0$. If the ratio of maximal to minimal label lengths is bounded by a constant, the approximation factor becomes $1 - \epsilon$.

1 Introduction

Map labeling is the problem of placing a set of labels next to a given set of points in the plane while meeting certain conditions. Most often, the label associated with a point is of a specified rectangular shape and must be placed in the plane without rotation so that its boundary touches the point. One distinguishes *fixed-position models* and *slider models*. In fixed-position models, the labeled point must belong to a predetermined finite set of points on the boundary of the label. Slider models allow the labeled point to touch the label anywhere along a certain segment of the label's boundary. Poon et al. [3] introduce a hierarchy of fixed-position and slider models.

In this paper we consider the slider model 1SH [3] that is defined as follows. Let P be a set of n points in the Euclidian plane $\mathbb{R}^2$. The x and y coordinates of a point p are denoted by p_x and p_y, respectively. We associate with each point $p \in P$ an axes-parallel open rectangular shape L_p of unit height and length $l_p \in \mathbb{R}_{>0}$, the *label* of p. Each point $p \in P$ has a *weight* $w_p \in \mathbb{R}_{>0}$. An instance of a 1SH-labeling problem is given by the triple $I = (P, l, w)$. A *(1SH-) labeling* of I is a family $\mathcal{L} = \{r_p\}_{p \in Q}$, indexed by the elements of $Q \subseteq P$, where $r_p \in [0, l_p]$ places L_p in the plane with its right edge at the x-coordinate $p_x + r_p$ and its lower edge at the y coordinate p_y; see Fig. 1. For any two points $p, q \in Q$, the values of r_p and r_q must be such that $L_p \cap L_q = \emptyset$. Points in Q are said to be *labeled*. We define $w_{\mathcal{L}} := \sum_{p \in Q} w_p$ to be the *weight* of the labeling $\mathcal{L} = \{r_p\}_{p \in Q}$. A labeling of I is *optimal* if no labeling of I has a larger weight. We denote the weight of an optimal labeling of I by $w_{\text{opt}}(I)$.

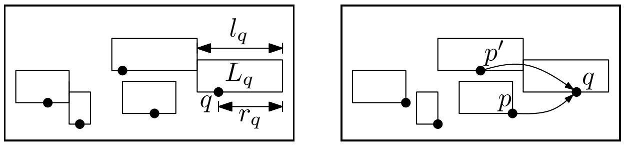

Figure 1: *Left*: A 1SH-labeling $\mathcal{L}$. *Right*: The corresponding normalization $\mathcal{L}^*$ and $G_{\mathcal{L}^*}$.

Poon et al. [3] show that finding an optimal 1SH-labeling is NP-hard even if all points lie on a horizontal line and the weight of each point equals the length of its label. For the one-dimensional case, they give a fully polynomial-time approximation scheme, which yields an $O(n^2/\epsilon)$-time $(1/2-\epsilon)$-approximation for the two-dimensional case. Poon et al. also give a polynomial-time approximation scheme (PTAS) for unit-square labels. They raise the question of whether a PTAS exists for rectangular labels of arbitrary length and unit height. This is known to be the case for fixed-position models [1] and for sliding labels with unit weight [4].

In this paper we make a step towards settling the question. We present a new approximation algorithm for 1SH-labeling. If the ratio of maximal to minimal label lengths is bounded by a constant, our algorithm is a PTAS. In the general case our algorithm has an approximation factor of $2/3 - \epsilon$, for arbitrary fixed $\epsilon > 0$. This is an improvement over the approximation factor of $1/2 - \epsilon$ of Poon et al. [3].

In Section 2 we discretize the problem. The idea is to compute, for a given problem instance $I = (P, l, w)$, "suitable" sets of discrete label positions for each point in P. "Suitable" means that the weight $w_{\text{opt}}(I_{\text{fix}})$ of an optimal labeling of the resulting instance I_{fix} of a fixed-position labeling problem must be close enough to $w_{\text{opt}}(I)$. This leads to a two-step approximation algorithm for the slider model 1SH:

[*]Work supported by grant WO 758/4-2 of the German Research Foundation (DFG).

[†]Department of Computer Science, University of Leicester, Leicester LE1 7RH, England. te17@mcs.le.ac.uk

[‡]Institut für Informatik, Universität Augsburg, D–86135 Augsburg, Germany. hagerup@informatik.uni-augsburg.de

[§]Institut für Informatik und Praktische Mathematik, Universität Kiel, D–24098 Kiel, Germany. kj@informatik.uni-kiel.de

[¶]moritz.minzlaff@stud.uni-karlsruhe.de

[‖]Fakultät für Informatik, Universität Karlsruhe, P.O. Box 6980, D-76128 Karlsruhe, Germany. WWW: i11www.ira.uka.de/people/awolff

First discretize the given problem instance. Then apply an algorithm, e.g., the one in [1], to the computed fixed-position problem instance. In Section 3 we consider the requirements for and the quality of the discretization.

2 Approximation by a Fixed-Position Model

By the *diameter* of a directed graph G we mean the maximum number of nodes on any path in G. As a first step towards discretization, we associate with each labeling $\mathcal{L}$ a directed graph $G_{\mathcal{L}}$. If, for some constant $t \in \mathbb{N}$, $G_{\mathcal{L}}$ has a subgraph of weight at least $(1 - 1/t)w(\mathcal{L})$ and of diameter bounded by a constant g, then we can compute an instance I of a fixed-position labeling problem such that an optimal labeling of I has weight at least $(1 - 1/t)w(\mathcal{L})$. Roughly speaking, in order to compute the instance I, we enumerate all positions that a label can have if it is part of a chain of at most g labels that succeed each other in the following sense.

Definition 1 *Let $\mathcal{L} = \{r_p\}_{p \in Q}$ be a labeling and let $p, q \in Q$. The label L_q* succeeds *L_p (with respect to $\mathcal{L}$), written $L_p \to L_q$, if (a) the left edge of L_q touches the right edge of L_p other than at a corner, (b) the vertical line supporting the left edge of L_q contains no point in Q, and (c) $r_q \neq 0$.*

The position of each label in a chain of labels that succeed each other depends on the position of the first label in the chain. The following definition discretizes the position of the first label and thus of all its successors.

Definition 2 *Let $\mathcal{L} = \{r_p\}_{p \in Q}$ be a labeling. For $q \in Q$, a label L_q is in* normal *position (with respect to $\mathcal{L}$) if (a) $r_q = 0$, (b) the vertical line supporting the left edge of L_q contains a point in Q, or (c) L_q succeeds L_p for some $p \in Q$. If all labels of $\mathcal{L}$ are in normal position, then $\mathcal{L}$ is* normal.

We can obtain a normal labeling from an arbitrary labeling by processing the labeled rectangles in the order from left to right and moving each rectangle left until it first reaches a normal position. We call this process *normalizing* the labeling. For an example see Fig. 1.

We could associate a directed graph $G = (Q, E)$ with a labeling $\{r_p\}_{p \in Q}$ by defining $(p, q) \in E$ if and only if $L_p \to L_q$. However, this is not entirely satisfactory: On the one hand, we want to bound the diameter of the graph. On the other hand, we want to normalize labelings. However, normalization may increase the diameter by too much. For this reason we also define an edge whenever there is the "possibility" that a label L_q succeeds a label L_p, namely if a normalization after the removal of some other labels can make this happen. This gives rise to the edge (p, q) in Fig. 1.

Definition 3 *Let $\mathcal{L} = \{r_p\}_{p \in Q}$ be a labeling. The* labeling graph *$G_{\mathcal{L}} = (Q, E)$ of $\mathcal{L}$ has the edge (p, q) if*

(E1) $p_x < q_x$,

(E2) $p_x + r_p > q_x - l_q$,

(E3) $(p_y, p_y + 1) \cap (q_y, q_y + 1) \neq \emptyset$, *and*

(E4) there is no $p' \in Q$ such that
$p_x + r_p \leq p'_x \leq q_x - l_q + r_q$.

For a label L_q to succeed a label L_p, the point p must lie to the left of q (E1). Properties (E2) and (E3) ensure that L_p and L_q overlap if L_q is shifted as far left as possible. Finally, (E4) implies that there is no point in Q between the right edge of L_p and the left edge of L_q. Note that $L_p \to L_q$ implies (E1)–(E4). We define the *weight* of a labeling graph $G_{\mathcal{L}}$ as the weight of the labeling $\mathcal{L}$. The following lemma lists properties of labeling graphs that we will need later.

Lemma 1 *Every labeling graph is a planar directed acyclic graph. If (p, q) is an edge of the graph, then this edge is the only path from p to q.*

Now we can state precisely the condition under which we can discretize a 1SH-labeling problem such that the optimal weight of the resulting discrete instance is close enough to that of the original instance.

Theorem 2 *Assume that for some instance $I = (P, l, w)$ of the 1SH-labeling problem, there are $g \in \mathbb{N}$ and $\alpha \in \mathbb{R}$ such that for every labeling $\mathcal{L}$ of I, there is a normal labeling $\mathcal{L}^*$ of I with $w_{\mathcal{L}^*} \geq \alpha w_{\mathcal{L}}$ for which the diameter of $G_{\mathcal{L}^*}$ is bounded by g. Then, in $O(n^{g+1})$ time, we can compute for each $p \in P$ a set $M(p) \subseteq [0, l_p]$ of cardinality $O(n^g)$ such that the instance I_{fix} of the fixed-position labeling problem defined by the sets $M(p)$ fulfills $w_{\text{opt}}(I_{\text{fix}}) \geq \alpha w_{\text{opt}}(I)$.*

Proof. For $\tau = 1, \ldots, g$, we compute for each $p \in P$ a set $M(p, \tau)$ that contains all possible values of r_p in a normal labeling whose longest chain of labels ending in L_p contains τ labels.

$$M(p, 1) = \big(\{q_x - (p_x - l_p) \mid q \in P\} \cup \{0\}\big) \cap [0, l_p]$$

for all $p \in P$, and for $\tau = 2, \ldots, g$ and for all $p \in P$,

$$M(p, \tau) = \{q_x + r - (p_x - l_p) \mid q \in P, r \in M(q, \tau - 1)\} \cap [0, l_p].$$

Finally, let $M(p) = \bigcup_{\tau=1}^{g} M(p, \tau)$. Clearly $|M(p, 1)| \leq n + 1$ and, by induction, $|M(p, \tau)| = O(n^\tau)$ for $\tau = 1, \ldots, g$. Thus $|M(p)| = O(n^g)$ for all $p \in P$, and $M(p)$ can be computed for all $p \in P$ in $O(n^{g+1})$ time.

If $\mathcal{L}$ is an optimal labeling of I, then, by assumption on g, there is a normal labeling $\mathcal{L}^* = \{r_p^*\}_{p\in Q}$ ($Q \subseteq P$) of I whose weight is at least $(1-1/t)w_{\text{opt}}(I)$ and whose labeling graph $G_{\mathcal{L}^*}$ has diameter at most g. By the construction above, $r_p^* \in M(p)$ for each $p \in Q$. Thus $\mathcal{L}^*$ is a labeling of the fixed-position problem instance I_{fix} defined by the sets $M(p)$, so $w_{\text{opt}}(I_{\text{fix}}) \geq (1-1/t)w_{\text{opt}}(I)$. □

3 Simplifiable Graphs

To satisfy the prerequisites of Theorem 2, we are interested in families of labeling graphs for which, for each constant $t \in \mathbb{N}$, there is a constant $g = g(t)$ such that every labeling graph G in the family is *trimmable* for g, i.e., contains a subgraph G' with weight $w' \geq (1-1/t)w$ and diameter at most $g(t)$. As a tool we use the stronger notion of simplifiable graphs, which we now define.

Definition 4 *Let $G = (V, E)$ be a directed graph. A* simplification *of G is a function $f : V \to \mathbb{Z}$ with the following properties:*

(S1) $\forall (v,w) \in E : \; 0 \leq f(w) - f(v) \leq 1.$

(S2) For all $v \in V$, there exists at most one $w \in V$ such that $(v,w) \in E$ and $f(v) = f(w)$.

The graph G is said to be simplifiable *by f and for each $i \in \mathbb{Z}$, the set $V_f(i) = \{v \in V \mid f(v) = i\}$ is called the i*th level of G with respect to f.

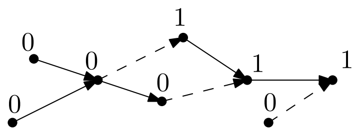

Figure 2: A graph with a simplification. Dashed arrows indicate a change of levels.

We write $V(i)$ instead of $V_f(i)$ whenever f is clear from the context. Given a graph $G = (V, E)$ and a subset V' of V, we denote by $G[V']$ the subgraph of G induced by V'. Now we exploit property (S1) to reduce the trimming of graphs to that of levels.

Lemma 3 *Let $G = (V, E)$ be a labeling graph with simplification f and let $t \in \mathbb{N}$. Assume that for each $i \in \mathbb{Z}$, there is a subset $V'(i)$ of $V(i)$ such that $G[V'(i)]$ has weight at least $(1-1/(2t))w(G[V(i)])$ and diameter at most $g'(t)$. Then there is a subgraph G' of G with weight at least $(1-t)w(G)$ and diameter at most $(2t-1)g'(t)$.*

Proof. Consider the subgraphs

$$G_\tau := G[V \setminus \bigcup_{i\in\mathbb{Z}} V(\tau + 2ti)], \; \tau = 0, \ldots, 2t-1,$$

in which we remove every $(2t)$th level of G. By the pigeon-hole principle, there is a $\tau_0 \in \{0, \ldots, 2t-1\}$ such that G_{τ_0} has weight at least $(1-1/2t)w(G)$. Due to property (S1), the nodes of any path in G_{τ_0} span at most $2t-1$ distinct levels of G. Hence, with $V' = \bigcup_{i\in\mathbb{Z}} V'(i)$, the graph

$$G' := G[V' \setminus \bigcup_{i\in\mathbb{Z}} V'(\tau_0 + 2ti)]$$

corresponding to G_{τ_0} has the required property. □

Property (S2) implies that each node has at most one successor within its own level. Thus a level $V(i)$ is an in-forest; see Fig. 2. For $r = 0, \ldots, 2t-1$, let $V'_r(i)$ be the subset of $V(i)$ obtained by removing all nodes whose depth in the forest modulo $2t$ is r. By the pigeon-hole principle, the set $V'(i)$ with maximum weight among the sets $V'_r(i)$ satisfies the requirements of Lemma 3 with $g'(t) = 2t-1$.

Theorem 4 *Every simplifiable labeling graph is trimmable for $g(t) = (2t-1)^2$.*

3.1 Outerplanar labeling graphs

Lemma 5 *Let F be the subgraph of a labeling graph spanned by the nodes on the boundary of one of its faces. For every edge (v, w) of F, there are simplifications f and f' of F with $f(w) = f(v)$ and $f'(w) = f'(v) + 1$.*

This lemma depends on the last property noted in Lemme 1. The proof is simple and can be found in [2].

Theorem 6 *Every outerplanar labeling graph is simplifiable.*

Proof. A simplification can be constructed essentially as follows. Begin by choosing an arbitrary inner face of the graph. By Lemma 5, a simplification of this face exists. Then extend the simplification face by face to a simplification of the entire graph. The clue is that by outerplanarity, there is always a face that shares either exactly one node or one edge with the current subgraph. Hence extending the simplification can be done with a repeated application of Lemma 5. Tree parts of the graph can also be handled easily. We refer to [2] for a detailed proof. □

We apply this result in the context of *stabbing lines*. A set $\mathcal{S}$ of stabbing lines with respect to a labeling $\{r_p\}_{p\in Q}$ is a set of horizontal lines with the following properties: Each line has distance greater than 1 from every other line, each line intersects at least one label, and each label is intersected by exactly one line. With each stabbing line l we may associate a "sub-labeling" consisting of all r_p such that L_p intersects l. For more on stabbing lines, see, e.g., [1, 2].

Proposition 7 *If $\mathcal{L}$ is a labeling with two stabbing lines, then the labeling graph $G_{\mathcal{L}}$ is outerplanar.*

Proof. In a natural planar embedding of $G_{\mathcal{L}}$, each node can be reached from above or from below by a vertical ray from infinity that does not cross any edges. Therefore all nodes of $G_{\mathcal{L}}$ lie on the boundary of the outer face. □

The following corollary is an immediate consequence of the results of this section, Theorem 2 and the two-step algorithm outlined in Section 1. Its proof again relies on the pigeon-hole principle: Consider removing all labels stabbed by every third line. An optimal fixed-position labeling for k stabbing lines can be computed in $A_k(\tilde{n}) = O(\tilde{n}^{2k-1})$ time, where $\tilde{n}$ is the size of the fixed-position instance [1].

Corollary 8 *There exists an approximation algorithm with factor $2/3(1-1/t)$ for the slider model that runs in $O(n^{g(t)+1}) + A_2(n^{g(t)+1}) = O(n^{3(g(t)+1)})$ time, where $g(t) = (2t-1)^2$.*

3.2 Bounded Ratio of Label Lengths

The labeling graph of Fig. 3 is not simplifiable: As the numbers next to the nodes indicate, on the path from a via b to c any simplification would need to "jump" at least two levels.

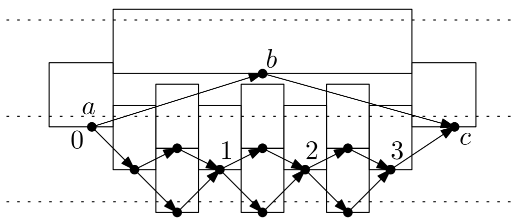

Figure 3: A labeling graph whose labeling needs three stabbing lines.

The example easily extends in such a way as to force a simplification to make arbitrarily large "jumps". However, the ratio of the maximal label length M to the minimal label length m would tend to infinity in the process. This gives rise to the following idea: let $\lambda \in \mathbb{N}$ and replace (S1) in the definition of a simplification by

(S1$_\lambda$) $\forall (v,w) \in E: \; 1 \le f(w) - f(v) \le \lambda$

We call a function f with properties (S1$_\lambda$) and (S2) a λ-simplification and the corresponding graph λ-simplifiable. Note that $f(w) \ge f(v) + 1$ for each edge (v,w) according to this definition. Therefore, each level $V(i)$ induces a subgraph of diameter 1 (consisting of isolated nodes). In the proof of Lemma 3 we can set $V'(i) = V(i)$ and have $g'(t) = 1$. If further not only one but λ consecutive levels out of t levels are removed from the graph to get G_τ, the pigeon-hole principle yields the following analog of Theorem 4.

Theorem 9 *If G is a λ-simplifiable labeling graph of a labeling $\mathcal{L}$ and t is an integer with $t > \lambda$, then G contains a subgraph with weight at least $(1-\lambda/t)w(\mathcal{L})$ and diameter at most $g(t) = t - \lambda$.*

Theorem 10 *Let I be a problem instance with $M/m \le \rho$ for some $\rho \in \mathbb{N}$. Then the labeling graph of every labeling of I is (2ρ)-simplifiable.*

Proof. Consider a labeling $\mathcal{L} = \{r_p\}_{p \in Q}$ of I. The idea is to divide the x-axis into intervals of length m and to assign a node p of $G_{\mathcal{L}}$ to level i if the x-coordinate of the left edge of L_p belongs to the interval $[im, (i+1)m)$. This is achieved by the following function $f : Q \to \mathbb{Z}$:

$$f(p) = \left\lfloor \frac{p_x - l_p + r_p}{m} \right\rfloor, \; p \in Q.$$

Let (p,q) be an edge of the labeling graph. By properties (E1), (E2) and (E3) and the fact that L_p and L_q do not overlap, the left edge of L_q is at least $l_p \ge m$ and at most $l_p + l_q \le 2M$ to the right of the left edge of L_p. Therefore $1 \le f(q) - f(p) \le 2\rho$, which shows f to be a (2ρ)-simplification of $G_{\mathcal{L}}$. □

Plugging Theorem 10 into Theorem 9 and using a PTAS for fixed-position models [1] yields a PTAS for instances with bounded ratios of label lengths.

Corollary 11 *For all instances with $M/m \le \rho$ for some $\rho \in \mathbb{N}$ and for all integers $t > 2\rho$ and $k \ge 1$, there exists a factor-$(1-2\rho/t)(1-1/(k+1))$ approximation algorithm that runs in $O(A_k(n^{g(t)+1}))$ time, where $g(t) = t - 2\rho$.*

Acknowledgment

We thank an anonymous referee, whose recommendations led to greater readability.

References

[1] P. K. Agarwal, M. van Kreveld, and S. Suri. Label placement by maximum independent set in rectangles. *Computational Geometry: Theory and Applications*, 11:209–218, 1998.

[2] M. Minzlaff. Beschriften gewichteter Punkte mit verschiebbaren Labeln. Student research report, Fakultät für Informatik, Universität Karlsruhe, March 2005. Available at `http://i11www.ira.uka.de/teaching/theses/files/studienarbeit-minzlaff-05.pdf`.

[3] S.-H. Poon, C.-S. Shin, T. Strijk, T. Uno, and A. Wolff. Labeling points with weights. *Algorithmica*, 38(2):341–362, 2003.

[4] M. van Kreveld, T. Strijk, and A. Wolff. Point labeling with sliding labels. *Computational Geometry: Theory and Applications*, 13:21–47, 1999.