

Approximation algorithms for free-label maximization

Mark de Berg*

Dirk H.P. Gerrits*

Abstract

Inspired by applications where moving objects have to be labeled, we consider the following (static) point labeling problem: given a set P of n points in the plane and labels that are unit squares, place a label with each point in P in such a way that the number of free labels (labels not intersecting any other label) is maximized. We develop efficient constant-factor approximation algorithms for this problem, as well as PTASs, for various label-placement models.

1 Introduction

Map labeling involves associating textual *labels* with certain *features* on a map such as cities (points), roads (polylines), and lakes (polygons). Manually performing this task is estimated to take cartographers 50% of the total time in creating a map [11]. It is therefore unsurprising that map labeling was listed as an important research area in the ACM Computational Geometry Impact Task Force report [4], and has generated a lot of algorithmic research, especially for point features. See for instance the on-line Map Labeling Bibliography [15], currently containing 371 references.

Label models. A good labeling for a point set has legible labels, and an unambiguous association between the labels and the points. The latter puts restrictions on the shape of labels and the way they can be placed in relation to points. Various such *label models* have been proposed, most often with labels assumed to be axis-aligned rectangles slightly larger than the text they contain.

In the *fixed-position models*, every point has a finite number of *label candidates* (often 4 or 8), each being a rectangle having the point on its boundary. In particular, in the 1-position (1P) model one designated corner of the label must coincide with the point, in the 2-position (2PH, 2PV) models there is a choice between two adjacent corners, and the 4-position (4P) model allows any corner of the label to coincide with the point (see the upper-left 2x2 block in Figure 1). The *slider models*, introduced by Van Kreveld et al. [14] generalize this. In the 1-slider (1SH, 1SV) models one side of the label is designated, but the label may contain the point anywhere on this side. In the 2-

slider (2SH, 2SV) models there is a choice between two opposite sides of the label, and in the 4-slider (4S) model the label can have the point anywhere on its boundary (see the fourth row and column in Figure 1). Erlebach et al. [5] introduced terminology analogous to the slider models for fixed-position models with a non-constant number of positions (1MH, 1MV, 2MH, 2MV, 4M; see the third row and column in Figure 1).

















$y \setminus x$	1	2	k	∞
1	 optimal	 1/4-approx.	 1/4-approx.	 1/4-approx.
2	 1/4-approx.	 1/16-approx.	 1/16-approx.	 1/12-approx.
k	 1/4-approx.	 1/16-approx.	 1/32-approx.	 1/32-approx.
∞	 1/4-approx.	 1/12-approx.	 1/32-approx.	 1/24-approx.

Figure 1: The fixed-position and slider models, and our approximation results for them for the free-label-maximization problem (assuming unit-square labels). The x -axis (y -axis) indicates the number of allowed horizontal (vertical) positions for a label.

Static labeling. Intersecting labels and small font sizes hinder legibility. The *size-maximization problem* asks to label all points with pairwise non-intersecting labels of maximum size. For a given placement of the labels it is a fairly simple geometric task to find the optimal scale factor, so the problem can be solved optimally for the 1P model. For two or more label candidates the problem is APX-hard, even for unit-square labels [6]. Constant-factor approximation algorithms exist for various label models [6, 9].

The more widely studied *number-maximization problem* asks to label a maximum-cardinality subset of the n points with pairwise non-intersecting labels of *given* dimensions. Even if all labels are unit squares, this problem is known to be strongly NP-hard for the 1P [7], 4P [6, 10], and 4S models [14]. A generalization of this problem concerns weighted points [12] and asks for a maximum-weight subset

*Dept. of Computer Science, TU Eindhoven, the Netherlands, mdeberg@win.tue.nl, dirk@dirkgerrits.com

of the points to be labeled so that, for example, a big city will more likely get a label than a small town. For unit-height rectangular labels this problem admits a polynomial-time approximation scheme (PTAS) for static points in all fixed-position and slider models, both in the weighted [5, 12] and the unweighted case [1, 14]. For arbitrary rectangles in the unweighted case an $O(1/\log \log n)$ -approximation algorithm is known for the fixed-position models [3], but the slider models, the weighted case, and the (non-)existence of a PTAS remain open problems.

Labeling of moving points. Mobile devices with interactive displays and GPS have vastly increased both the availability of motion data and our ability to view them. An important aspect of displaying such data is the association of textual labels with points of interest. Yet, despite the large body of work on labeling static points, virtually no results have been published on labeling moving points. Been et al. [2] studied the unweighted number-maximization problem for static points under continuous zooming in and out by the viewer, which can be seen as points moving on a very specific kind of trajectories. Rostamabadi and Ghodsi [13] studied how to quickly flip and scale the labels of static points to avoid one moving point.

There is not only a lack of results on labeling moving points, but in fact the size- and number-maximization problems are ill-suited to moving points. Continuously scaling labels under size maximization would be hard to read, and the (dis)appearance of a label under number maximization can be disturbing for the viewer. We instead propose the *free-label-maximization problem*, where the labels have given dimensions (as in the number-maximization problem), but all points need to be labeled (as in the size-maximization problem). Instead of disallowing intersections, we want to maximize the number of labels which are not intersected, and call such labels *free*. Ideally, an algorithm for free-label maximization on moving points would move the labels continuously in such a way that the number of free labels is close to the static optimum at all times.

Our results. As a first step towards this goal we have studied the free-label-maximization problem for static points. For unit-square labels we have developed a simple $O(n \log n)$ -time, $O(n)$ -space constant-factor approximation algorithm, as well as a PTAS. This makes free-label maximization easier than size maximization, as the latter is APX-hard even for unit-square labels. In contrast, techniques used for (approximate) number maximization for unit-square labels easily extend to unit-height labels of different widths, which seems not to be the case for free-label maximization. Thus the complexity of free-label maximization seems to fall in between that of the size- and number-maximization problems.

We present our constant-factor approximation algorithm in the next section, leaving our PTAS to the full version of this paper. The algorithm's approximation guarantees for the various label models are listed in Figure 1. We prove them for the 2PH, 4P, 1SH, 2SH, and 4S label models, the other models being analogous. We assume that no two points have the same x - or y -coordinate, and that labels are open sets (their boundaries may intersect). Neither assumption is essential, but they make our exposition simpler.

2 Constant-factor approximations for unit squares

Consider the algorithm GREEDYSWEEP, which works as follows. Going through the points from left to right, we label them one-by-one. We call a label candidate ℓ for a point being processed *freeable* if none of the previously placed labels intersect ℓ , and every point still to be labeled has at least one label candidate that does not intersect ℓ or any previously placed freeable label. We always choose a freeable label candidate if possible, and then also call the resulting label freeable. If a point has no freeable label candidate we pick a non-freeable label candidate that does not intersect any previously placed freeable label (which is always possible by the definition of freeable). In case of ties, we pick the label candidate farthest to the left. (Further ties between equally leftmost label candidates can be broken arbitrarily.)

Lemma 1 *For the free-label-maximization problem with unit-square labels, algorithm GREEDYSWEEP gives a $1/4$ -approximation for the 2PH and 1SH models and this ratio is tight.*

Proof. Let OPT be some optimal solution, and let ALG be the solution computed by GREEDYSWEEP. Now suppose a point p is labeled with a free label ℓ_p^{OPT} in OPT, but that the label candidate ℓ_p^{OPT} was not freeable when p was being processed by GREEDYSWEEP. Call a label candidate for a point *rightmost* if it is farthest to the right of all label candidates for that point, and define *leftmost* analogously. Since p and all points that already have a label lie to the left of every unprocessed point p' , their labels cannot intersect the rightmost label candidate for p' without intersecting all other label candidates for p' as well. Thus all unprocessed points can be labeled with their rightmost label candidate without intersecting ℓ_p^{OPT} . Hence, ℓ_p^{OPT} not being freeable must be caused by a label $\ell_{p'}^{\text{ALG}}$ (either freeable or not) that was placed earlier. We note that $\ell_{p'}^{\text{ALG}}$ cannot be leftmost. (If the leftmost label candidate for a point p' left of p intersects ℓ_p^{OPT} , then all other label candidates for p' do as well, contradicting that ℓ_p^{OPT} is free in OPT.) That $\ell_{p'}^{\text{ALG}}$ is not leftmost can mean two things. Either $\ell_{p'}^{\text{ALG}}$ is freeable, in which case we charge ℓ_p^{OPT} to $\ell_{p'}^{\text{ALG}}$, or

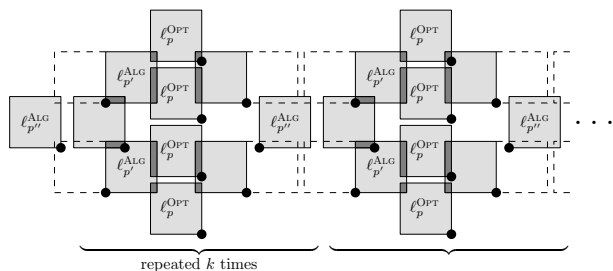


Figure 2: A labeling computed by GREEDYSWEEP for the 2PH model, where the $k + 1$ labels marked $\ell_{p''}^{\text{ALG}}$ are free. In the optimal solution the $4k$ labels marked ℓ_p^{OPT} are free. Thus the $1/4$ -approximation is tight for the 2PH model, and the same can be shown for the 1SH model by moving the points closer together horizontally.

making $\ell_{p'}^{\text{ALG}}$ leftmost will cause it to intersect some freeable label $\ell_{p''}^{\text{ALG}}$, in which case we charge ℓ_p^{OPT} to $\ell_{p''}^{\text{ALG}}$. With a careful case analysis one can argue that at most four free labels of OPT get charged to a single freeable label of ALG by the above scheme. Fig. 2 shows that resulting approximation ratio is tight.

We still need to consider the case where a point p has a free label ℓ_p^{OPT} in OPT that is also a freeable label candidate when p is being processed by GREEDYSWEEP. Then ℓ_p^{ALG} must be a freeable label, and we charge ℓ_p^{OPT} to ℓ_p^{ALG} . The label ℓ_p^{ALG} can at most be as far to the right as ℓ_p^{OPT} , otherwise GREEDYSWEEP would have picked ℓ_p^{OPT} over ℓ_p^{ALG} . One can argue that this implies ℓ_p^{ALG} will only be charged once. \square

Already for the 4P model, GREEDYSWEEP can be as bad as an $O(1/\sqrt{n})$ -approximation. We instead take the best solution over running GREEDYSWEEP several times with different sweep directions. For the 4P model we do one left-to-right sweep (as before) and one right-to-left sweep (preferring rightmost label candidates). For the 2SH model we do one top-to-bottom sweep (preferring topmost label candidates) and one bottom-to-top sweep (preferring bottommost label candidates). For the 4S model we sweep in all four of these directions. This yields the following:

Theorem 2 *There are $O(n \log n)$ -time and $O(n)$ -space algorithms for free-label maximization on n points with unit-square labels, having the following approximation ratios: $1/4$ (tight) for the 2PH and 1SH models, $1/12$ for the 2SH model, $1/16$ for the 4P model, and $1/24$ for the 4S model.*

Proof. We will prove the approximation ratio for the 4P model; the proofs for the 2SH and 4S models are similar, and the ratio for the 2PH and 1SH models was proved in Lemma 1. Let OPT be an optimal solution for the 4P model, and consider the solution ALG computed in the left-to-right sweep. We can assume that

at least half of the labels in OPT are placed in one of the two rightmost positions. (If not, at least half must be placed in one of the two leftmost positions and we can instead consider the right-to-left sweep in a completely symmetric way.) We will argue that the rightmost free labels in OPT can be charged to free labels of ALG so that no label receives more than eight charges, yielding the stated $1/16$ -approximation.

Suppose p is a point with a rightmost free label ℓ_p^{OPT} in OPT, but with a non-free label ℓ_p^{ALG} in ALG. At the time p was being processed, the label candidate ℓ_p^{OPT} must not have been freeable, either because some unprocessed point would inevitably get a label intersecting ℓ_p^{OPT} , or because some processed point already had a label intersecting ℓ_p^{OPT} . We consider these two cases separately.

(1) Suppose every label candidate of some unprocessed point p' intersects either ℓ_p^{OPT} or some previously placed freeable label. (This cannot occur in the 2PH and 1SH models.) Of the rightmost label candidates for p' one must be topmost, say $\ell_{p'}^{\wedge}$, and one must be bottommost, say $\ell_{p'}^{\vee}$. Since p and all points that already have a label lie to the left of p' , if ℓ_p^{OPT} or a freeable label intersects a rightmost label candidate for p' with the same y -coordinate but lying more to the left. So if all rightmost label candidates for p' are intersected by previously placed freeable labels, then all label candidates for p' are intersected by previously placed freeable labels, meaning that at least one of them was in fact not freeable. Thus ℓ_p^{OPT} must intersect some rightmost label candidate of p' . This implies that ℓ_p^{OPT} does not intersect the horizontal line through p' , for otherwise ℓ_p^{OPT} would contain p' . Thus ℓ_p^{OPT} intersects either $\ell_{p'}^{\wedge}$ or $\ell_{p'}^{\vee}$, but not both, so there must be a freeable label $\ell_{p''}^{\text{ALG}}$ in ALG which intersects $\ell_{p'}^{\vee}$ if ℓ_p^{OPT} intersects $\ell_{p'}^{\wedge}$, or vice versa. Charge ℓ_p^{OPT} to $\ell_{p''}^{\text{ALG}}$. One can argue that any freeable label can be charged at most twice this way (see Figure 3(b)).

(2) Suppose some already processed point p' has a label $\ell_{p'}^{\text{ALG}}$ (either freeable or not) that intersects ℓ_p^{OPT} . Because ℓ_p^{OPT} is rightmost, $\ell_{p'}^{\text{ALG}}$ cannot be leftmost. So either $\ell_{p'}^{\text{ALG}}$ is freeable, and we charge ℓ_p^{OPT} to $\ell_{p'}^{\text{ALG}}$, or making $\ell_{p'}^{\text{ALG}}$ leftmost will cause it to intersect some freeable label $\ell_{p''}^{\text{ALG}}$, and we charge ℓ_p^{OPT} to $\ell_{p''}^{\text{ALG}}$. One can argue that any freeable label can be charged at most six times this way for the 4P model (see Figure 3(a)).

Combining the charges of these two cases yields at most eight charges per free label for the 4P model, and we argued that at least one half the free labels in OPT could be charged, yielding the claimed $1/16$ -approximation. We have not yet charged free labels in OPT which label points that also have a free label in ALG. One can argue that charging such labels does

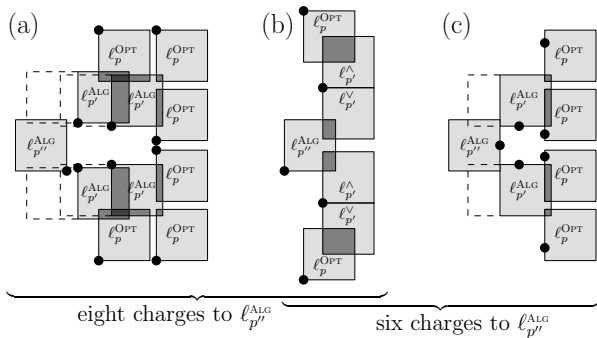


Figure 3: (b) If every ℓ_p^{OPT} charged to $\ell_{p''}^{\text{ALG}}$ is intersected by labels placed *later*, $\ell_{p''}^{\text{ALG}}$ is charged at most twice. If every ℓ_p^{OPT} charged to $\ell_{p''}^{\text{ALG}}$ is intersected by labels placed *earlier*, $\ell_{p''}^{\text{ALG}}$ is charged (a) at most six times for the 4P model, and (c) at most four times for the 2SH and 4S models.

not cost us extra charges, as one of the charges to $\ell_{p''}^{\text{ALG}}$ in Figure 3(b) must disappear if $\ell_{p''}^{\text{OPT}}$ is free.

The proofs for the 2SH and 4S models are similar, but each free label can get at most six charges (see Figure 3(b)–(c)). In the 2SH model every free label in OPT is either topmost or bottommost so that we can again charge at least half of them, but in the 4S model a label can also be leftmost or rightmost so that we can charge only one fourth.

With some clever use of standard data structures, similar to the $1/2$ -approximation algorithm for number maximization by Van Kreveld et al. [14], GREEDY-SWEEP can be implemented to run in $O(n \log n)$ time and $O(n)$ space. We omit the details. \square

3 Conclusion

We have presented a simple and efficient constant-factor approximation algorithm for a new variant of the labeling problem motivated by the wish to label moving points. Our algorithm works for the case where all labels are unit squares (or, equivalently, if all labels are rectangles of the same dimensions). For this case we also developed a PTAS using a variation on the “shifting technique” due to Hochbaum and Maass [8]. Details can be found in the full version of this paper. The cases of labels being unit-height rectangles or arbitrary rectangles are still open. For the number-maximization problem these cases allow, respectively, a PTAS and an $O(1/\log \log n)$ -approximation, and it would be interesting to see if these results can be matched. If not, the free-label-maximization problem is strictly harder than the number-maximization problem, while easier than the size-maximization problem. The weighted version of the free-label-maximization problem is another interesting direction for future research.

References

- [1] P. Agarwal, M. van Kreveld, and S. Suri. Label placement by maximum independent set in rectangles. *Comput. Geom. Theory Appl.*, 11:209–218, 1998.
- [2] K. Been, M. Nöllenburg, S.-H. Poon, and A. Wolff. Optimizing active ranges for consistent dynamic map labeling. *Comput. Geom. Theory Appl.*, 43(3):312–328, 2010.
- [3] P. Chalermsook and J. Chuzhoy. Maximum independent set of rectangles. In C. Mathieu, editor, *Proc. 20th ACM-SIAM Sympos. on Discrete Algorithms (SODA'09)*, pages 892–901, New York, 2009.
- [4] B. Chazelle and 36 co-authors. Application challenges to computational geometry: CG impact task force report. Technical Report TR-521-96, Princeton University, 1996.
- [5] T. Erlebach, T. Hagerup, K. Jansen, M. Minzlaff, and A. Wolff. Trimming of graphs, with an application to point labeling. In S. Albers and P. Weil, editors, *Proc. 25th Internat. Sympos. Theoretical Aspects Comput. Sci. (STACS'08)*, pages 265–276, Bordeaux, 2008.
- [6] M. Formann and F. Wagner. A packing problem with applications to lettering of maps. In *Proc. 7th Annu. ACM Sympos. Comput. Geom. (SoCG'91)*, pages 281–288, North Conway, 1991.
- [7] R. Fowler, M. Paterson, and S. Tanimoto. Optimal packing and covering in the plane are NP-complete. *Information Processing Letters*, 12:133–137, 1981.
- [8] D. Hochbaum and W. Maass. Approximation schemes for covering and packing problems in image processing and VLSI. *Journal of the ACM*, 32(1):130–136, 1985.
- [9] M. Jiang, S. Bereg, Z. Qin, and B. Zhu. New bounds on map labeling with circular labels. In *Proc. 15th Annu. Internat. Sympos. Algorithms Comput. (ISAAC'04)*, pages 606–617, 2004.
- [10] J. Marks and S. Shieber. The computational complexity of cartographic label placement. Technical Report TR-05-91, Harvard CS, 1991.
- [11] J. Morrison. Computer technology and cartographic change. In D. Taylor, editor, *The Computer in Contemporary Cartography*. Johns Hopkins University Press, 1980.
- [12] S.-H. Poon, C.-S. Shin, T. Strijk, T. Uno, and A. Wolff. Labeling points with weights. *Algorithmica*, 38(2):341–362, 2003.
- [13] F. Rostamabadi and M. Ghodsi. A fast algorithm for updating a labeling to avoid a moving point. In *Proc. 16th Canadian Conf. Comput. Geom. (CCCG'04)*, pages 204–208, 2004.
- [14] M. van Kreveld, T. Strijk, and A. Wolff. Point labeling with sliding labels. *Comput. Geom. Theory Appl.*, 13:21–47, 1999.
- [15] A. Wolff and T. Strijk. The Map Labeling Bibliography. <http://liinwww.ira.uka.de/bibliography/Theory/map.labeling.html>, 2009.