Approximate Map Labeling is in $\Omega(n \log n)$

Frank Wagner*

B 93–18
December 1993

Abstract

Given n real numbers, the ϵ-CLOSENESS problem consists in deciding whether any two of them are within ϵ of each other, where ϵ is a fixed parameter of the problem. This is a basic problems with an $\Omega(n \log n)$ lower bound on the running time in the algebraic computation tree model.
In this paper we show that for a natural approximation version thereof the same lower bound holds. The main tool used is a lower bound theorem of Ben-Or. We introduce a new interpolation method relating the approximation version of the problem to two corresponding exact versions.
Using this result, we are able to prove the optimality of the running time of a cartographic algorithm, that determines an approximate solution to the so-called MAP LABELING problem. MAP LABELING was shown to be $\mathcal{NP}$-complete. The approximation algorithm discussed here is of provably optimal approximation quality. Our result is the first $n \log n$ lower bound for an *approximate* problem. The proof method is general enough to be potentially helpful for further results of this type.

*Institut für Informatik, Fachbereich Mathematik und Informatik, Freie Universität Berlin, Takustraße 9, 14195 Berlin-Dahlem, Germany, e-mail: wagner@math.fu-berlin.de

1 Introduction

Map labeling is one of the classical key problems that has to be solved in the process of map production. Mostly the map producer wants to include in a map not only the exact geographic positions of the features depicted, but also informations, typically in written form, explaining properties of these features. She has to arrange these informations on the map such that

— for every information it is intuitively clear, which feature is described;
— the information is of sufficient (font-)size to be legible;
— different texts do not overlap.

These (and in addition a lot of esthethical) criteria are described by [Imhof] in an attempt to characterize good quality map labeling (having mostly manual map making in mind). Now there is an increasing need for large, especially technical maps, for which legibility is much more important than beauty.
The application which brought the problem to our attention is the design of ground-water quality maps by the municipal authorities of the city of München. They have a net of drillholes spread over the city. The map has to contain the sites of these holes and for every hole a block of measuring results such as the concentration of certain chemicals.
The growing importance of such technical maps induces a need for the computerization of map making, the need for fully automated algorithms. Typically labels in technical maps are axis-parallel rectangles of identical sizes. By rescaling one of the axes we can assume that the rectangles are squares.

An adequate formalization is the following:

Problem MAP LABELING

Given n distinct points in the plane. What is the supremum σ_{opt} of all reals σ such that for each point there is a closed square, with side length σ, satisfying the following two properties?

(1) The point is one of its corners.
(2) All squares are pairwise disjoint.

We call σ_{opt} the *optimal size*. A set of non-intersecting squares is called a *valid labeling.*

In Figure 1 a valid labeling for a sample problem is depicted, in Figure 2 a

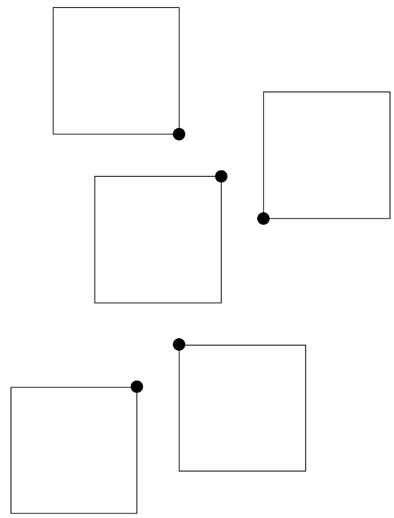

Figure 1: A valid labeling

valid labeling of optimal size for the same problem is shown. By a reduction from 3-SAT [Formann and Wagner] showed that the corresponding decision problem is $\mathcal{NP}$-complete. The main result of that paper is an approximation algorithm that finds a valid labeling of at least half the optimal size. In addition it is shown that, provided that $\mathcal{P} \neq \mathcal{NP}$, no polynomial time approximation algorithm with a quality guarantee better than 1/2 exists.
The running time of the approximation algorithm is in $\mathcal{O}(n \log n)$. Beside the application mentioned above, our algorithm is also used by the PTT Research Labs of the Netherlands to produce maps for mobile radio networks. With a very similar algorithmic approach we were able to solve the so-called METAFONT labeling problem posed by [Knuth and Raghunathan].
In this paper, we show that our algorithm is optimal also with respect to the running time by providing an $\Omega(n \log n)$ lower bound for any approximation algorithm for the MAP LABELING problem that guarantees that the size of the valid labeling constructed is at most a certain factor off the optimal size.
This lower bound is shown for the *algebraic computation tree model* (see e. g. [Ben-Or]). The central lower bound theorem of [Ben-Or] states

Theorem 1.1
Let $W \subseteq R^n$ be any set, N be the number of connected components of W, and $C(W)$ the maximal number of steps (arithmetical operations, comparisons, taking square roots) of an algebraic computation tree that solves the membership problem for W and is optimal with respect to the worst case behaviour (length of the longest path from the root to a YES/NO-leaf). Then

$$C(W) \geq c_1 \log N - c_2 n$$

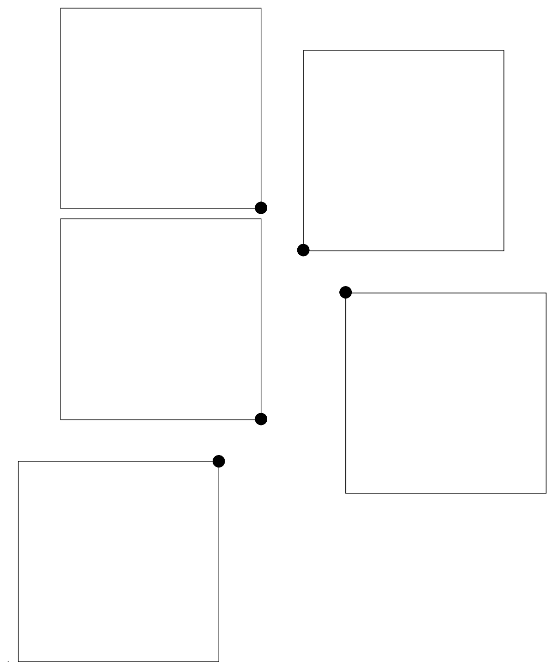

Figure 2: A valid labeling of optimal size for the example shown in Figure 1

where c_1 and c_2 are positive constants.

2 The Reduction

A precise formulation of the approximate MAP LABELING problem with approximation guarantee 1/2 as a kind of decision problem is

Given a set of n distinct points in the plane that have to be labeled and a size σ.
Say YES, if there is a valid labeling of size 2σ.
Say NO, if there is no valid labeling of size σ.
Say anything, otherwise.

Obviously, an approximate MAP LABELING algorithm with the claimed approximation quality can be used to solve this partial decision problem by saying YES, if it produces a labeling of size at least σ and NO, if it produces a labeling of size smaller than σ.
We now construct a linear-time reduction to this decision problem from the approximate ϵ-CLOSENESS problem which is precisely the following:

Given a set of distinct points on the real line.
Say NO, if no two of them are within 1.8ϵ of each other.
Say YES, if two of them are within ϵ of each other.
Say anything, otherwise.

Lemma 2.1
If there is an algebraic decision tree for approximate MAP LABELING that requires $S(n)$ steps in the worst case, then there is one for approximate ϵ-CLOSENESS that requires $S(4n) + cn$ steps, where c is a positive constant.

Proof:
Take the n distinct real numbers a_1 to a_n and transform them into the following set P of points of four different types in the plane as an input for the approximate MAP LABELING algorithm, where $\delta = \epsilon/10$:

$$\begin{aligned} P &= \{l_i, m_i, r_i, s_i | i = 1, \ldots, n\} \\ \text{where} & \\ l_i &= (a_i - \delta, 0) \\ m_i &= (a_i - \delta, 2\delta) \\ r_i &= (a_i + \delta, \delta) \\ u_i &= (a_i + \delta, 3\delta) \end{aligned}$$

Observe that the construction transforms the numbers a_1 to a_n into a set of $4n$ *distinct* points in the plane. See Figure 3 for an example.

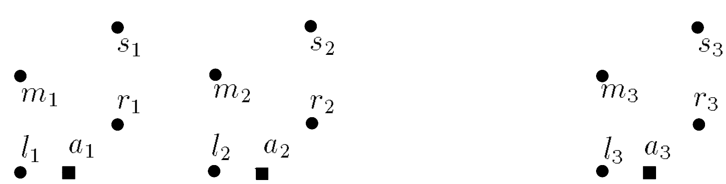

Figure 3: A set of reals and its transform

Now we decide the approximate MAP LABELING problem on these $4n$ points for the size $\sigma = \epsilon/2 - \delta = 0.4\epsilon$. We have:
If no two of the a_i are within 1.8ϵ of each other, then there is a valid labeling for P of size 0.8ϵ. This can be constructed by choosing at every point the square that bounds away from the other three points with the same index; i. e. for every l_i we take the south-west-square, for every r_i the south-east-square and so on. Thus no

two squares labeling points with the same index conflict. The same holds for points with different indices since we know that no two of them are within 1.6ϵ of each other. Thus, we are sure that that the approximate MAP LABELING algorithm says YES.
If, on the other hand, two of the a_i's, say a_j and a_k with $a_j < a_k$, are within ϵ of each other, there cannot be a valid labeling of size 0.4ϵ. This can be seen as follows: Assume that there is such a valid labeling. Any label of size 0.4ϵ of a point of type r_i has to be placed in the south-east of its point, since in any of the other three possible positions it would hit another point. Having noticed this we can argue similarly, that every such label of a point of type l_i has to be placed in the south-west of its point. Now we distinguish two cases (see Figures 4 and 5 for examples):

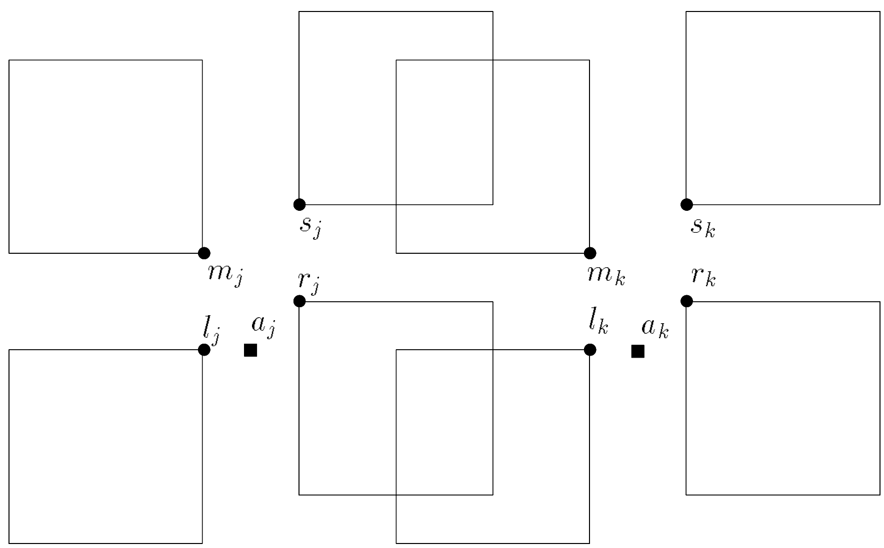

Figure 4: Case 1: $a_k - a_j \geq \delta$

If the distance of the two a_i's is at least δ, the labels for r_j and l_k are in conflict. If not, the labels for l_j and l_k overlap. This is a contradiction to the assumed *valid* labeling. Thus, we are sure that the approximate MAP LABELING algorithm says NO.

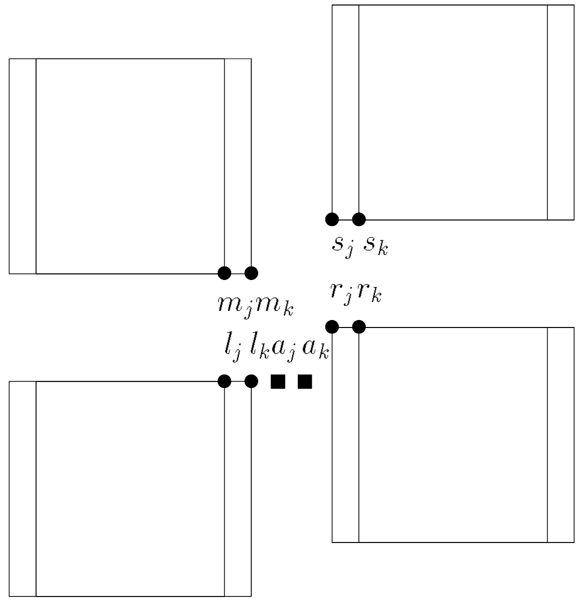

Figure 5: Case 2: $a_k - a_j < \delta$

Overall we have an approximate ϵ-CLOSENESS algorithm as claimed, by performing the transformation to $4n$ points in linear time, applying the approximate MAP LABELING algorithm and simply answering NO if its answer is YES and vice-versa. ◻

3 A lower bound for approximate ϵ-CLOSENESS

[Ben-Or] proves the $\Omega(n \log n)$ lower bound for the ϵ-CLOSENESS problem by showing that the set W_ϵ has $n!$ connected components, where

$$W_\epsilon = \{(a_1, \ldots, a_n) | \text{ no two of the } a_i\text{'s are within } \epsilon \text{ of each other}\}.$$

Using the fact that $\log(n!) \in \Omega(n \log n)$ Theorem 1 then yields the claimed result. In case of our approximation problem we are not able to say exactly what is the set, for which we decide the membership problem. The reason for this is that there is a set of inputs for which the algorithm is allowed to say NO or YES, whatever it likes. Of course, in the computation model we discuss here, it will always say the same for a fixed input.

In fact, it decides the membership problem for a set W_{approx}, with

$$W_\epsilon \subseteq W_{approx} \subseteq W_{1.8\epsilon}.$$

Of course, for both the sets W_ϵ and $W_{1.8\epsilon}$ we know that they have $n!$ connected components. We will now demonstrate that this also holds for any set lying between these two with respect to the $\subseteq$-relation:
Observe that, in general, it is not true, that a set B lying between two other sets A and C, consisting of k connected components each, has that many components itself (see Figure 6 for a counterexample).

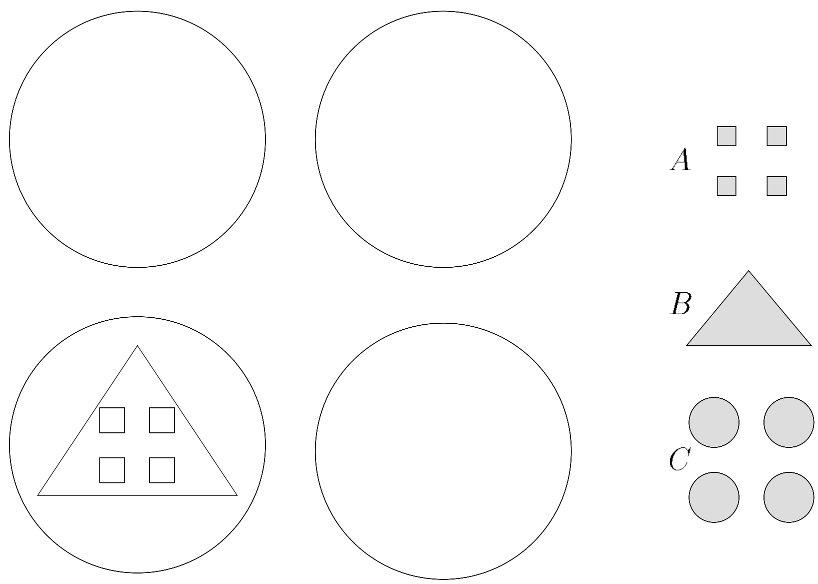

Figure 6: A counterexample for $k = 4$

In order to show this property in our case we have to take a closer look at the topological properties of W_ϵ:
As observed by [Fredman and Weide], W_ϵ consists of the $n!$ disjoint and noncontiguous regions

$$W_\epsilon^\pi = \{(a_1, \ldots, a_n) | a_{\pi(1)} < a_{\pi(2)} < \ldots < a_{\pi(n)} \text{ and } |a_j - a_i| > \epsilon \text{ for all } i \neq j\}$$

where π is a permutation of the integers 1 to n. Analogously, $W_{1.8\epsilon}$ is the disjoint union of the sets $W_{1.8\epsilon}^\pi$. Clearly, for every fixed permutation π we have $W_{1.8\epsilon}^\pi \subseteq W_\epsilon^\pi$. Thus every set lying between $W_{1.8\epsilon}$ and W_ϵ consists of $n!$ disjoint and noncontiguous subsets W_{approx}^π, with

$$W_{1.8\epsilon}^\pi \subseteq W_{approx}^\pi \subseteq W_\epsilon^\pi.$$

These subsets are not necessarily connected, but we know for sure that W_{approx} consists of at least $n!$ connected components (see Figure 7).

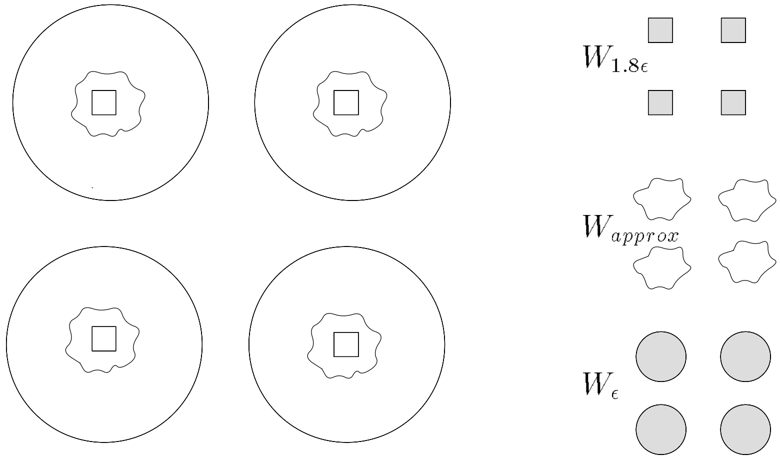

Figure 7: The rough topology of $W_{1.8\epsilon} \subseteq W_{approx} \subseteq W_{\epsilon}$

So, by [Ben-Or]'s main theorem we have shown

Theorem 3.1
The running time of any algorithm solving the approximate ϵ-CLOSENESS problem (in the algebraic computation tree model) is in $\Omega(n \log n)$.

From this we achieve by Lemma 2:

Corollary 3.2
The running time of any algorithm solving the approximate MAP LABELING problem (in the algebraic computation tree model) is in $\Omega(n \log n)$.

Concluding we would like to point out that the constant 1.8 in the formulation of approximate ϵ-CLOSENESS was chosen just for technical reasons. The proof works for any other constant larger than one. One could also choose functions of n instead of the constants if they are in the right order. All of these generalizations transform to the approximate MAP LABELING problem.

References

[Ben-Or] M. Ben-Or, *Lower bounds for algebraic computation trees*, Proceedings of the 15th Annual ACM Symposium on the Theory of Computing (1976) 80–86

[Formann and Wagner] M. Formann, F. Wagner, *A Packing Problem with Applications to Lettering of Maps*, Proceedings of the 7th ACM Symposium on Computational Geometry (1991) 281–288

[Fredman and Weide] M. L. Fredman, B. Weide, *On the complexity of computing the measure of* $\cup[a_i, b_i]$, Communications of the ACM (1978) 540–544

[Imhof] E. Imhof, *Positioning Names on Maps*, The American Cartographer **2** (1975) 128-144

[Knuth and Raghunathan] D. E. Knuth and A. Raghunathan, *The Problem of Compatible Representatives*, SIAM Journal on Discrete Mathematics **5** (1992) 422–427