Complexity of Octagonal and Rectangular Cartograms

T. Biedl and B. Genc *

1 Introduction

A cartogram is a type of map used to visualize data. In a map regions are displayed in their true shapes and with their exact relations with the adjacent regions. However, such a map can only be used to demonstrate the actual area values of the regions. Sometimes, we need to display other data on a map, such as population, pollution, electoral votes, production rates, etc. One efficient way to do so is to modify the map such that the area of each shape corresponds to the data to be displayed. A map with given relationships between regions for which each region has pre-specified area is called a *cartogram* (see [1] for details).

There are two major cartogram types: contiguous area cartograms [2, 3, 6, 7, 12], where the regions are deformed but stay connected, and non-contiguous area cartograms [8], where regions preserve their shapes but may lose adjacency relationships. *Rectangular cartograms*, where every region is a rectangle is a specific type of contigous area cartograms which tries to preserve both the adjacency relations and the shape, but this does not exist for all area values. Kreveld and Speckmann [13] introduced the first automated algorithms for such cartograms. Heilmann et al. proposed RecMap [5] to approximate familiar land covering map region shapes by rectangles. Rahman et al. studied slicing and good slicing graphs and their orthogonal drawings [9], which are similar to orthogonal cartograms.

It was left as an open problem whether testing the feasibility of a rectangular cartogram is NP-hard. In this paper, we make significant progress towards answering this question. We first study what we call *cartograms of orthogonal octagons* where every region is an orthogonal polygon with at most 8 sides. We also assume that the cartogram must be placed within a rectangle of fixed size (a *canvas*). We show that testing whether a cartogram of orthogonal octagons exists is NP-hard.

We then use a very similar reduction to prove NP-hardness of a problem where, all faces are rectangles, except for one face corresponding to the "sea" around islands and peninsulas (see the examples in [13]).

2 Definitions

Recall that a graph $G = (V, E)$ is called *planar* if it can be drawn in the plane without crossing. Such a drawing defines a cyclic order of incident edges around each vertex; the collection of these cyclic orders is called the *planar embedding*. A planar drawing of a planar graph defines connected regions of the plane called *faces*; the unbounded region is called the *outer-face*. A planar embedding of a graph defines uniquely the faces, except for the choice of the outer-face. A planar graph where both a planar embedding and an outer-face have been specified is called a *plane graph*.

Given a plane graph, we define the *dual graph* by defining a vertex for every face. For every edge in the primal graph incident to faces f_1 and f_2, we define a *dual edge* in the dual graph incident to the vertices of the faces f_1 and f_2.

An *orthogonal cartogram dual* is a plane graph with one special vertex C (the *canvas*) where every incidence between a vertex and an edge is labeled with one of {N,S,E,W} corresponding to four directions.

It may not be straightforward to see that any such plane graph indeed gives rise to a valid drawing, but this can be shown using the technique of converting an orthogonal representation into an orthogonal drawing proposed by Tamassia [11].

All that is needed to specify a cartogram is to demand an area of each face. Thus an *orthogonal cartogram* is an orthogonal cartogram dual G, together with a positive integer area for every vertex $v \neq C$ of G.

We will sometimes additionally demand that the whole drawing fits inside the canvas. Thus we may specify a $w \times h$ rectangle R and demand that the drawing fit inside it. In particular, $w \cdot h$ must be at least the area of all other vertices together, but it may be more, allowing for some "dead space" (also known as *the sea*) on the outer-face. Note that the aspect ratio of the rectangle for the canvas does not matter; if the drawing fits into any rectangle of area $w \cdot h$, then after suitable scaling it fits into all rectangles of area $w \cdot h$.

3 NP-hardness

We show now that testing whether an octagonal orthogonal cartogram can be realized is NP-hard. The proof is by reduction from Partition defined as follows. Assume that we are given a set A of positive integers $a_1 \dots a_n$ with $\sum_{i=1}^{n} a_i = 2S$ for some integer S. We want to find a subset I of A which satisfies $\sum_{a_i \in I} a_i = S$. It is known that this is NP-hard [4].

3.1 Construction

Given an instance of Partition $a_1, \dots, a_n$, we create the cartogram as follows. We have $2n + 5$ faces

*School of Computer Science, University of Waterloo, Waterloo, ON N2L 3G1. {biedl,bgenc}@uwaterloo.ca. Research supported by NSERC. The authors would like to thank Marc van Kreveld and Bettina Speckmann for helpful input.

$A_1, P_1, \dots, A_n, P_n$ and $M, B_1, \dots, B_4$, which all are rectangles except P_i, $i = 1, \dots, n$ is a Z-shaped octagon.

The adjacency relations between these faces are given in Figure 1, both as a drawing and by giving the cartogram dual. For easier visualization of the latter, we direct the edges and show the label only for the tail of each edge; we also split the canvas C into multiple vertices. Furthermore, we show A_i and P_i only for the special cases $i = 1, n$ and for one generic i; the generic i needs to be repeated $n - 3$ times.

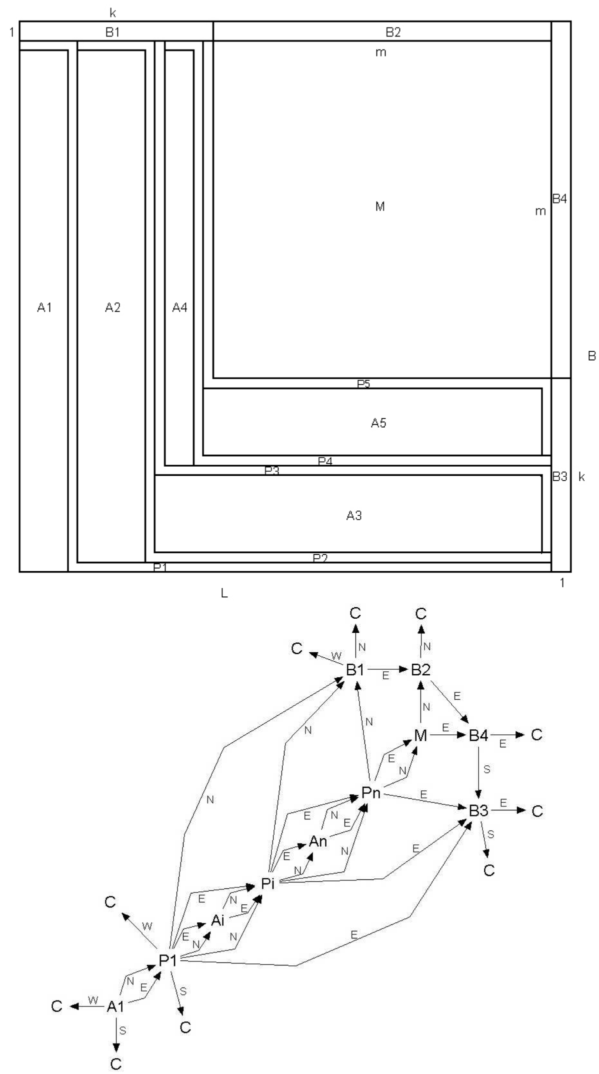

Figure 1: The drawing and the cartogram dual of the cartogram generated from the given PARTITION instance.

Now we explain the area requirements, depending on four parameters m, k, C and p. We will use $k = n$, $m = 2nS+4$, $C = 2n^2$ and $p = n+8$, but actually a wide range of parameters is possible and can be calculated from the proofs of the lemmas. The area requirements and purposes of faces are as follows:

- Each rectangle A_i corresponds to one number a_i of the PARTITION instance. We set area$(A_i) = C \cdot a_i$.
- Each Z-shaped octagon P_i is a buffer between rectangles A_i and A_{i+1} (or M); we set area$(P_i) = p$.
- M is a huge rectangle with area m^2 that splits the rest of the canvas into essentially two parts.
- $B_1, \dots, B_4$ builds a frame that forces M to be an $m \times m$-square. We set area(B_1) =area$(B_3) = k$, area$(B_2) = m$ and area$(B_4) = m + 1$.

Some easy calculations show that with our choice of parameters we have $2CS + np = (m + k)^2 - m^2$; this shows that the area of all regions together is $(m+k+1)^2$ and by using an $(m + k + 1) \times (m + k + 1)$ as canvas, there is no empty space left for a sea.

We now show our constructed cartogram is realizable iff the instance of PARTITION has a solution.

From cartogram to PARTITION Assume first that we have a realization of the cartogram. We need some intermediary lemmas.

Lemma 1 *The widths and heights of M, B_1, B_2, B_3 and B_4 are as labeled in Figure 1.*

Proof. Note that the left edges of B_3 and B_4 are collinear in any realization and touch the top and bottom of the canvas. Since they have a total area of $m + k + 1$ and the canvas has height $m + k + 1$, the widths of B_3 and B_4 have to be 1. Due to the individual area requirements, this fixes the height of B_3 to k and the height of B_4 to $m+1$. Similarly B_1 and B_2 will have a fixed height of 1 and B_1 will have a width of k while B_2 will have a width of m. This fixes the size of M to $m \times m$ in any realization. □

For the rest of the proof, L marks the line through the left side of the rectangle M and B is the line through the bottom side of M. See also Figure 2. For each A_i, we now have two possible layouts: A_i may be placed to the left of line L or below line B. See Figure 2. One can show that a rectangle A_i cannot be both below B and to the left of L, because there is not enough space for it. By our choice of parameters, one can immediately verify the following:

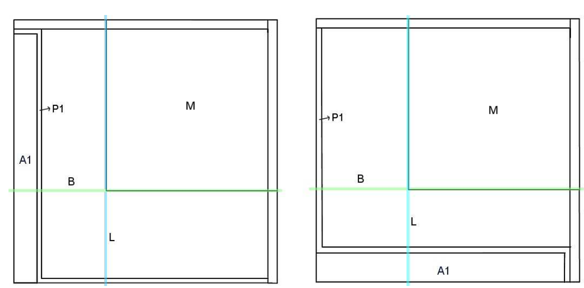

Figure 2: Two possible layouts for A_1.

Lemma 2 *The total area to the left of line L is less than $C(S+1)$. The total area below line B is less than $C(S+1)$.*

Lemma 3 *If I denotes the indices of rectangles A_i to the left of L, then $\sum_{i \in I} a_i = S$.*

Proof. By Lemma 2, we have $\sum_{i \in I} \text{area}(A_i) = \sum_{i \in I} Ca_i < C(S+1)$, so $\sum_{i \in I} a_i < S+1$, and hence $\sum_{i \in I} a_i \leq S$ since all numbers are integers. All rectangles with indices not in I are not to the left of L and hence must be below B. So $\sum_{i \notin I} \text{area}(A_i) < C(S+1)$ by Lemma 2, which similarly implies $\sum_{i \notin I} a_i \leq S$. Since $\sum_{i \in I} a_i + \sum_{i \notin I} a_i = 2S$, we must have equality for both sets. □

With this, a realization of the cartogram clearly gives a solution to the Partition instance.

From Partition to cartogram Now we work on the other direction. Assume that the Partition instance has a solution I, i.e., $\sum_{i \in I} a_i = S$. Principally, the idea to construct the cartogram is easy: Let the width and height of $M, B_1, \ldots, B_4$ be as indicated in Figure 1, and position each A_i, P_i pair in one of the two fashions shown in Figure 2, depending on whether $i \in I$.

The details are more complicated, because we need to choose the dimensions of A_i and P_i such that all regions fit exactly into the L-shaped region. We will show that such coordinates exist, by giving two layouts that don't quite work, and arguing that there exists a realization somewhere between them.

For the claims to come, we will need to introduce some notations. The *width* (*height*) of an L-shaped region X is the width (height) of its bounding box. Let $L(X)$ and $B(X)$ be the vertical and horizontal lines through the unique reflex vertex of X. We call the rectangle to the left of $L(X)$ the *left region*, and its width the *left width*. We call the region below B the *bottom region*, and its height the *bottom height*.

The first L-shaped region that we consider is what is left of the canvas after placing $M, B_1, \ldots, B_4$; we will call this L_0. It has width and height $m+k$ and left width and bottom height k. The other two L-shaped regions are the areas occupied by the two layouts that we are going to define.

We also need some notations for the realizations of P_i. Consider Figure 2 again. In either method of realizing P_i, it is the union of three rectangles that overlap at the corners. One of these spans the height of the available area; we call this the *left rectangle*. Another one spans the width of the available area; we call this the *bottom rectangle*. The third one is only adjacent to A_i and the outside; we call this the *end rectangle*.

Now we are ready to define the layouts precisely:

- Set $H = m+k$ and $W = m+k$; these keep track of the bounding box of the remaining L-region.
- For $i = 1, \ldots, n$:
  - Place A_i at the bottom left corner.
  - If $i \in I$, set the width of A_i to be $(\text{area}(A_i) + 2)/H$, and set the height accordingly. This leaves two units of area free above A_i; these will be covered by the end rectangle of P_i.
  - If $i \notin I$, set the height of A_i to be $(\text{area}(A_i) + 2)/W$, and set the width accordingly.
  - Choose the proper shape for P_i as in Figure 2.
  - The dimensions for P_i depend on the layout:
    - In the first layout, choose dimensions of P_i such that the left rectangle has area at least $n+4$ and the bottom rectangle has area at least 2. Recall that P_i has area $n+8$ and that the rectangles of P_i intersect, so there is enough area of P_i to do this.
    - In the second layout, the bottom rectangle has area at least $n+4$ and the left rectangle has area at least 2.
  - Update W and H by subtracting the union of A_i and P_i from the free region.

Let L_1 be the union of $A_1, P_1, \ldots, A_n, P_n$ in the first layout, and L_2 be the union of $A_1, P_1, \ldots, A_n, P_n$ in the second layout. Neither L_1 nor L_2 is a realization of the cartogram, since they don't fit into L_0. We will show that some layout between L_1 and L_2 does fit. To do so, we first show that L_1 is too wide and L_2 is too slim.

Lemma 4 *The left width of L_1 is not smaller than the left width of L_0, and the left width of L_2 is not bigger than the left width of L_0.*

Proof. These hold due to the dimensions we assigned to A_i and P_i pairs. □

Lemma 5 *There exists a realization of the cartogram such that all rectangles of all P_i's have area at least 2.*

Proof. L_1 has at least area $n+4$ in the left rectangle of each P_i, and at least 2 units area in the bottom rectangle, whereas for L_2 it was vice versa. We can now define intermediary layouts between L_1 and L_2, where we gradually shift area from the left rectangle of each P_i to the bottom rectangle. In the beginning we have L_1, where the left width is at least k (by Lemma 4), and at the end we have L_2, where the left width is at most k. At some point we have a layout L^* with left width k, and its height and width is $m+k$. L^* must have exactly the shape of L_0. Finally note that all rectangles of all P_i's in both L_1 and L_2 have area at least 2, and this holds for all intermediary drawings. □

Thus given a solution to Partition, we can obtain a cartogram, which proves our main result:

Theorem 6 *Testing whether an orthogonal octagonal cartogram is realizable is NP-hard.*

3.2 Rectangular cartograms with sea

The same construction leads to an NP-hardness result for rectangular cartograms if a sea is allowed. We replace octagon P_i with a rectangle R_i that connects to A_i and B_1. R_i has area 1. This leaves some empty space (sea) since R_i requires less area than P_i. The sea will take on the role of a "buffer". See Figure 3.

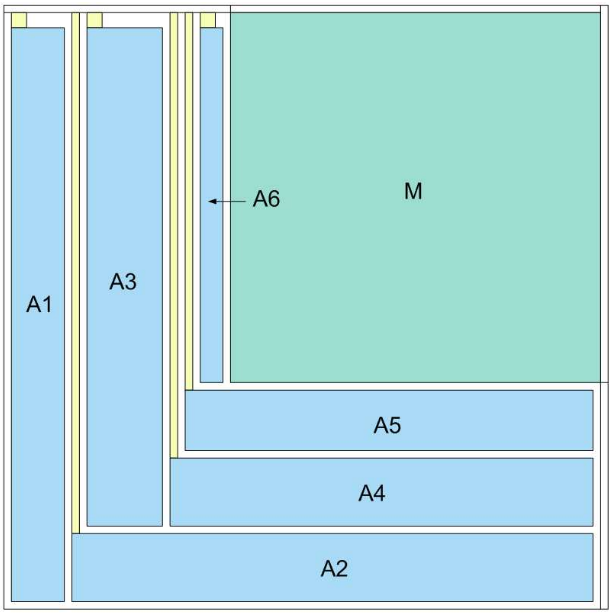

Figure 3: NP-hardness with only rectangular regions and a sea.

With exactly the same proof as before one shows that if this cartogram can be realized, then the PARTITION instance has a solution; note that nowhere in this part of the proof did we make use of the octagons P_i.

On the other hand, if PARTITION has a solution, then we create a cartogram as before. Now we can place R_i inside the end rectangle of P_i (if $i \in I$) or inside the left rectangle of P_i (if $i \notin I$); we know that these rectangles have area at least 2 and hence there is sufficient space for R_i. We thus obtain the following theorem.

Theorem 7 *Testing whether a rectangular cartogram with a sea is realizable is NP-hard.*

4 Conclusion and open problems

In this paper, we studied the complexity of realizing a rectangular cartogram that is bounded by a canvas. The main question (is this NP-hard?) remains open, but we showed that two closely related results are indeed NP-hard. In particular, the small (and realistic) step of adding a sea bounded by a canvas to a rectangular cartogram makes the problem NP-hard.

The most pressing open problem is to resolve the complexity of rectangular cartograms. Can we do away with the sea? Another very interesting problem is whether this problem is actually NP-complete, i.e., is it in NP? Also note that, PARTITION has a pseudo-polynomial time solution which means our proof does not ensure Strong NP-Hardness.

Finally, we are interested in exploring cartograms with orthogonal octagons (or k-gons for some small number of k) further. Note that k-gons are more flexible than rectangles, and thus more cartograms will be realizable. In particular, Speckmann et al. showed that all cartograms are realizable if all faces have a constant number of corners [10]. Also, a number of existing heuristics seem to rely on using k-gons for small k.

References

[1] Cartogram central. http://www.ncgia.ucsb.edu/projects/Cartogram_Central/, Web site at University of California at Santa Barbara, last accessed February 25, 2005.

[2] J. Dougenik, N. Chrisman, and D. Riemeyer. An algorithm to construct continuous area cartograms. *Professional Geographer*, 37:75–81, 1985.

[3] H. Edelsbrunner and R. Waupotitsch. A combinatorial approach to cartograms. In *SCG '95: Proceedings of the eleventh annual symposium on Computational geometry*, pages 98–108. ACM Press, 1995.

[4] M. R. Garey and D. S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. Freeman, 1979.

[5] R. Heilmann, D. A. Keim, C. Panse, and M. Sips. Recmap: Rectangular map approximations. In *INFOVIS '04: Proceedings of the IEEE Symposium on Information Visualization (INFOVIS'04)*, pages 33–40. IEEE Computer Society, 2004.

[6] D. H. House and C. J. Kocmoud. Continuous cartogram construction. In *VIS '98: Proceedings of the conference on Visualization '98*, pages 197–204. IEEE Computer Society Press, 1998.

[7] D. A. Keim, S. C. North, and C. Panse. Cartodraw: A fast algorithm for generating contiguous cartograms. *IEEE Transactions on Visualization and Computer Graphics*, 10(1):95–110, 2004.

[8] J. Olson. Noncontiguous area cartograms. *Professional Geographer*, 28:371–380, 1976.

[9] M. S. Rahman, K. Miura, and T. Nishizeki. Octagonal drawings of plane graphs with prescribed face areas. In *Graph-Theoretic Concepts in Computer Science: 30th International Workshop*, volume 3353 of *Lecture Notes in Computer Science*, pages 320–331, 2004.

[10] B. Speckmann. Personal communication.

[11] R. Tamassia. On embedding a graph in the grid with the minimum number of bends. *SIAM J. Computing*, 16(3):421–444, 1987.

[12] W. Tobler. Pseudo-cartograms. *The American Cartographer*, 13:43–50, 1986.

[13] M. v. Kreveld and B. Speckmann. On rectangular cartograms. In *European Symposium on Algorithms (ESA'04)*, volume 3221 of *Lecture Notes in Computer Science*, pages 724–735, 2004.