ELSEVIER

Computational Geometry 7 (1997) 343–360

Computational Geometry
Theory and Applications

A combinatorial approach to cartograms ☆

Herbert Edelsbrunner *, Roman Waupotitsch

Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

Communicated by E. Welzl; received 1 September 1995; revised 9 January 1996

Abstract

A homeomorphism from $\mathbb{R}^2$ to itself distorts metric quantities, such as distance and area. We describe an algorithm that constructs homeomorphisms with prescribed area distortion. Such homeomorphisms can be used to generate cartograms, which are geographic maps purposely distorted so their area distributions reflects a variable different from area, as for example population density. The algorithm generates the homeomorphism through a sequence of local piecewise linear homeomorphic changes. Sample results produced by the preliminary implementation of the method are included. © 1997 Elsevier Science B.V.

Keywords: Combinatorial topology; Algorithms; Simplicial maps; Homeomorphisms; Singular value decomposition; Arrays; Trees; Range search

1. Introduction

Cartograms and homeomorphisms

The starting point for the research reported in this paper is a problem in cartography that has to do with deformations of geographic maps [2]. A *cartogram* is a geographic map that is purposely distorted so its spatial properties represent quantities not directly associated with position on the globe [3]. Of particular interest are distortions that do not sacrifice continuity, that is, cartograms generated by applying a homeomorphism to the plane carrying the original geographic map.

A more general and more mathematical view of the problem studied in this paper is the generation of a homeomorphism $\mathbb{R}^2 \to \mathbb{R}^2$ whose effect on area distortion can be prescribed.

Continuous versus combinatorial methodology

Previous approaches to producing cartograms are based on continuous methods dealing with force fields and differential equations. The algorithmic ideas are usually limited to iterations approximating

☆ This research is partially supported by the National Science Foundation, under grant ASC-9200301.

* Corresponding author. Supported through the Alan T. Waterman award, grant CCR-9118874.

SSDI 0925-7721(96)00006-5

a continuous solution [3,10]. The approach in this paper is different and based on combinatorial methods in geometry [6], topology [8] and algorithms [1]. We use simplicial complexes tiling $\mathbb{R}^2$ in the construction of piecewise linear homeomorphisms. The final deformation is the composition of a sequence of local homeomorphic deformations. Each local deformation maintains area except in a small region specified within the tiling.

We also study the question of quality of the homeomorphism expressed in terms of local distance distortions. The assessment of quality opens the way to an objective comparison of different homeomorphisms or cartograms. Possibly more importantly, it can be used to improve quality through area invariant deformations.

The approach to constructing homeomorphic deformations described in this paper has been implemented. Applications of the resulting software to a specific problem instance can be found in Section 7. We describe two data structures representing the homeomorphism, one based on arrays and the other on trees. They differ in storage and time requirements. The final picture, the image under the homeomorphism, is constructed by searching through either one of these data structures.

Outline of the paper

Section 2 presents definitions of the topology and geometry concepts needed for our method. Section 3 describes the main steps of the homeomorphism construction. Section 4 develops a data structure for constructing and evaluating a homeomorphism based on arrays. Section 5 studies an alternative data structure based on trees. Section 6 discusses the assessment of quality of the thus constructed homeomorphisms. Section 7 mentions a few applications and exhibits sample results produced with out software. Finally, Section 8 offers a few concluding remarks.

2. Definitions and concepts

We use periodic tilings to impose a discrete structure on the real plane, $\mathbb{R}^2$. Finite portions of these infinite tilings are algorithmically manipulated. It is convenient to tile with triangles so barycentric coordinates can be used to extend vertex maps to continuous piecewise linear maps in an unambiguous manner. For most of the definitions we follow the notation in [8], but see also [4,9]. Some of the definitions are more general than required for the technical discussions in this paper. We hope the generality motivates the reader to imagine generalizations of our method, for example to dimensions beyond 2 and to continuous maps that are not homeomorphisms.

Simplicial complexes

A *k-simplex*, σ, is the convex hull of $k+1$ affinely independent points. The *dimension* of σ is $\dim \sigma = k$. In $\mathbb{R}^2$, at most 3 points can be affinely independent, so we have only 0-simplices (points or vertices), 1-simplices (line segments or edges) and 2-simplices (triangles). Each subset of the $k+1$ points defines also a simplex called a *face* of σ. The face is *proper* if the subset is proper. It is convenient to call the empty set a (-1)-simplex and to consider it an (improper) face of every simplex.

For a vector e, the simplex $\sigma + e = \{p + e \mid p \in \sigma\}$ is a *translate* of σ, and for a scalar λ, $\lambda\sigma = \{\lambda p \mid p \in \sigma\}$ is a *scaled copy* of σ. An edge with endpoints u and v is usually denoted uv. The *cone* of u over A, $uA = \bigcup_{a\in A} ua$, is defined if $ua \cap ub = u$ for all $a \neq b$ in A. For example, if u is a point not in the affine hull of a k-simplex σ then $u\sigma$ is a $(k+1)$-simplex. By convention, $u\emptyset = u$.

A *simplicial complex* is a collection K of simplices for which $\sigma \in K$ and τ a face of σ implies $\tau \in K$, and $\sigma_1, \sigma_2 \in K$ implies $\sigma_1 \cap \sigma_2$ is a face of both. We allow countably infinite collections but require each vertex belong to only finitely many simplices. A simplicial complex $L \subseteq K$ is a *subcomplex* of K. Not every subset $M \subseteq K$ is also a subcomplex, but its *closure*, $\operatorname{Cl} M = \{\tau \in K \mid \tau$ a face of $\sigma \in M\}$ is a subcomplex of K. The *open star* of $\tau \in K$ is $\operatorname{St}\tau = \{\sigma \in K \mid \tau$ a face of $\sigma\}$. In contrast to the *closed star*, $\operatorname{Cl}\operatorname{St}\tau$, the open star $\operatorname{St}\tau$ is in general not a subcomplex of K.

The *vertex set* of K is $\operatorname{vert} K = \{\sigma \in K \mid \dim\sigma = 0\}$. The *underlying space* of K is $|K| = \bigcup_{\sigma\in K} \sigma$. It is often cumbersome to distinguish a complex from its underlying space, and occasionally we will prefer ambiguous over awkward language. A *triangulation* of $|K|$ is a simplicial complex L with $|L| = |K|$. A *subdivision* of K is a triangulation L of $|K|$ so each simplex of L is contained in a simplex of K. A subdivision of K can be generated by *starring* from a new vertex, u, added to K: each simplex $\sigma \in K$ that contains u is replaced by cones of u over all faces of σ disjoint from u. A *stellar subdivision* is the result of repeated starring.

Simplicial maps and homeomorphisms

Each point $x \in |K|$ belongs to the (relative) interior of exactly one simplex $\sigma \in K$. If σ is the convex hull of vertices $a_0, a_1, \ldots, a_k$, then there are unique positive real numbers $\alpha_0, \alpha_1, \ldots, \alpha_k$ with

$$\sum_{i=0}^{k} \alpha_i a_i = x \quad \text{and} \quad \sum_{i=0}^{k} \alpha_i = 1.$$

The *barycentric coordinate* of x with respect to a vertex u of K is $q_u(x) = \alpha_i$ if $u = a_i$ and $q_u(x) = 0$ if u is not a vertex of σ. It is important to note that $q_u : |K| \to [0,1]$ is continuous for each u, and that q_u is nonzero only for points x contained in simplices of the open star, $\operatorname{St} u$. We remark that barycentric coordinates have been used before in cartography to construct so-called rubber-sheet maps solving a problem different from the construction of cartograms [11].

Let K and L be two simplicial complexes. A *vertex map* $f : \operatorname{vert} K \to \operatorname{vert} L$ has the property that the vertices of a simplex in K are mapped to (not necessarily distinct) vertices of a simplex in L. The map f can be extended in an unambiguous way to a *simplicial map* $g : |K| \to |L|$ defined by

$$g(x) = \sum_{u\in\operatorname{vert} K} q_u(x) \cdot f(u).$$

g is continuous because each q_u is continuous. Furthermore, if f is a bijection, then g is a *homeomorphism*, that is, g is bijective and g and g^{-1} are both continuous. In this case, we call g a *simplicial homeomorphism*. K and L are *isomorphic* if such a homeomorphism exists.

The composition of two simplicial maps is again a simplicial map, and the composition of two homeomorphisms is again a homeomorphism. We will make extensive use of this composition property and generate homeomorphisms from possibly many components, each a simple piecewise linear homeomorphism.

Tilings

A simplicial complex K *tiles* $\mathbb{R}^2$ if $|K| = \mathbb{R}^2$. In the specialized literature, K would be referred to as a simplicial face-to-face tiling [9], but since we consider only this type we call K simply a *tiling*. A *symmetry* of K is an Euclidean motion (translation, rotation, reflection) that maps each simplex of K to a simplex of K. A symmetry is a particularly simple homeomorphism from $|K|$ to itself. The set of all symmetries forms the *symmetry group*. The set of translations in this group forms the *translation subgroup*. K is *periodic* if the translation subgroup contains translations in 2 linearly independent directions.

As an example, consider a tiling T of $\mathbb{R}^2$ consisting of equilateral triangles and their faces, see Fig. 1. T is unique up to similarity, and for specificity we define T so its vertices have coordinates $(i - j/2, j\sqrt{3}/2)$, $i, j \in \mathbb{Z}$. There is a one-to-one correspondence between the symmetries of T and the injective maps from the 3 vertices of a particular triangle to the vertices of any triangle in T. The symmetry group is therefore isomorphic to $S_3 \times \mathbb{Z} \times \mathbb{Z}$, where S_3 is the group of permutations of 3 elements. The translation subgroup is isomorphic to $\mathbb{Z} \times \mathbb{Z}$, so T is periodic.

3. Construction

Homeomorphisms in this paper are composed of instances of simple units. The most important such units are the inflations and their inverses, the deflations. They can be defined using closed stars, domes and corridors, which are subcomplexes of the periodic tiling T defined above. Stellar subdivisions are used in the creation of inflations. We begin by defining and discussing inflations on the lowest level of resolution.

Domes and corridors

The closed stars of any two vertices in T are isomorphic; each is defined by 6 triangles, see Fig. 1. A *dome* is the closure of a set of 4 triangles in contiguous positions in a star $\operatorname{St} u$. u is the *center vertex* of the dome. The middle 2 triangles form the *diamond* of the dome. A closed star contains 6 different domes. The boundary of a closed star and a dome are isomorphic, each being a cycle of 6 edges and vertices. This will be important for our construction, which maps the interior of a dome to the interior of a star.

A dome has one of 6 orientations specified by the 2 rows of triangles that need to be altered when the dome expands to a star. To formalize this observation, consider a dome D with center vertex u,

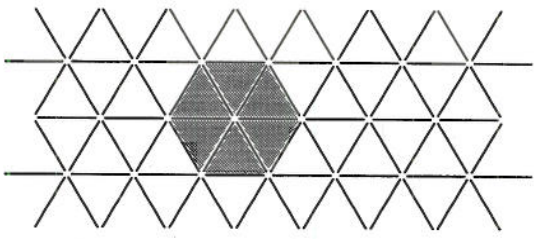

Fig. 1. Every vertex in T belongs to 6 triangles defining its star.

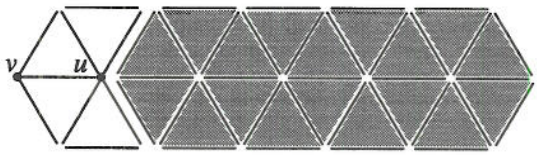

Fig. 2. The corridor of a dome is an infinite subcomplex of T. It shares 2 edges and 3 vertices with the dome.

and let uv be the edge decomposing D into 2 triangles on each side. v is the *start vertex*, $e = u - v$ the *direction vector*, and $C = \{\sigma + ie \mid \sigma \in D,\ i \geqslant 1\}$ is the *corridor* of D, see Fig. 2.

Inflations and deflations

Let $D \subseteq T$ be a dome, u the center vertex, and C the corridor. We construct 2 stellar subdivisions T' and T'' of T and a simplicial homeomorphism $\iota : |T'| \to |T''|$ equal to the identity except inside D and C. The 2 subdivisions are obtained through sequences of starring operations inside D and C, see Figs. 3 and 4. Observe that T' and T'' are isomorphic, so there is a unique bijective vertex map $f : \operatorname{vert} T' \to \operatorname{vert} T''$ equal to the identity everywhere except inside D and C. Its extension $\iota : |T'| \to |T''|$ is a simplicial homeomorphism. Since $|T'| = |T''| = |T| = \mathbb{R}^2$, the homeomorphism $\iota : |T| \to |T|$ is piecewise linear albeit not defined by a vertex map for T.

We call ι an *inflation*; its inverse, $\delta = \iota^{-1}$, is called a *deflation*. ι is completely determined by the choice of the dome, D. Intuitively, it moves every vertex in the interior of $D \cup C$ by one position along e. The first vertex in this sequence is freed up and used to expand the dome to a star. The triangles in C are deformed but area remains the same. Triangles outside D and C are left untouched. ι increases area only inside D. More precisely, the diamond of D is increased to twice the area, and the area of the 2 end triangles remains unchanged. The corresponding deflation halves the area of the

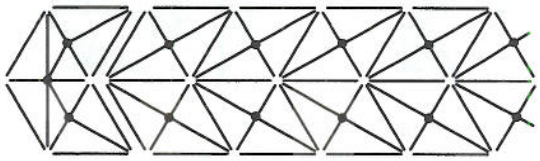

Fig. 3. To obtain T' from T subdivide the 4 triangles of the dome D into 10 triangles using starring operations adding the midpoints of the 3 interior edges of D. The corridor is decomposed into convex quadrilaterals. Each quadrilateral consists of 2 triangles and is further subdivided by a starring operation adding the midpoint of the dividing edge.

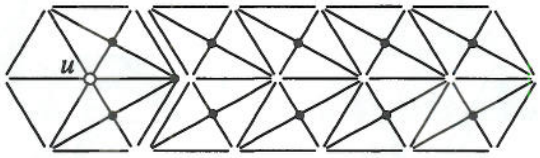

Fig. 4. To obtain T'' from T subdivide the 6 triangles of St u into 10 triangles using starring operations adding the midpoints of 2 interior edges. The part of the corridor outside the star is decomposed into convex quadrilaterals similar to but different from before. Each quadrilateral consists of 2 triangles and is further subdivided by a starring operation adding the midpoint of the dividing edge.

quadrilateral inside St u whose image is the diamond of D. The center vertex, u, is removed by the deflation.

Composition

The stellar subdivision T' of T is really the superposition of T and the preimage of T under ι. Similarly, T'' is the superposition of T and the image of T under ι. It is interesting to look at the preimage and image of T under the composition of several inflations. Let us consider an example of 3 inflations which distribute the impact of local growth in 3 different directions. Fig. 5 shows the 3 domes that define the inflations $\iota_1, \iota_2, \iota_3$ with center vertices 1, 2, 3. We execute the inflations in this sequence and thus consider $h = \iota_3 \circ \iota_2 \circ \iota_1$. The expansion effect is strongest for the central triangle shared by all 3 domes, see Fig. 5. Composition is not commutative, which can be seen from the asymmetry of the image. Fig. 6 shows the hexagon decomposed by triangles of T and their preimages under h. Symmetrically, Figs. 5 and 6 show the preimage and image of the homeomorphism $\delta_1 \circ \delta_2 \circ \delta_3 = h^{-1}$ composed of 3 deflations.

Instead of defining h as the composition of 3 inflations, we can also design our own image of the 7 triangles decomposing the union of domes. For example, the central region in the right part of Fig. 5 could be drawn with straight edges thus producing a uniform expansion to 7 times the original area. There are indeed plenty of possibilities resulting in different homeomorphic deformations. Similarly, we can create other operations, or *aggregates*, for special deformation tasks.

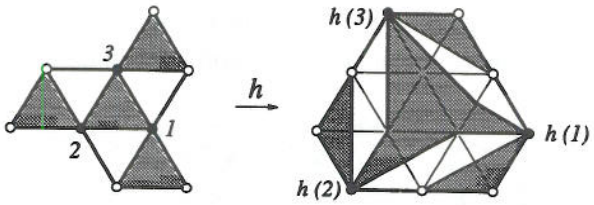

Fig. 5. The 3 inflations are defined by overlapping domes covering 7 triangles of the tiling. Any 2 domes overlap in 2 triangles, and all 3 domes overlap in 1 triangle. The image of the union of domes is a hexagon decomposed into 7 regions each corresponding to one of the 7 triangles.

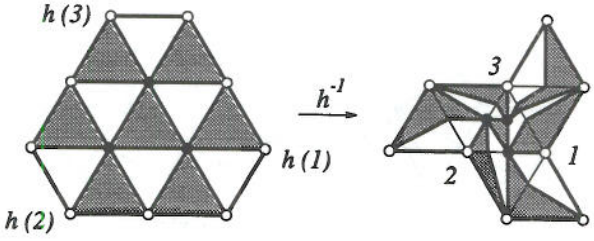

Fig. 6. In T, the image of the union of 3 domes is decomposed into $13 = 7 + 3 \cdot 2$ triangles. The preimages of the 13 triangles form a pattern of 13 regions inside the union of domes.

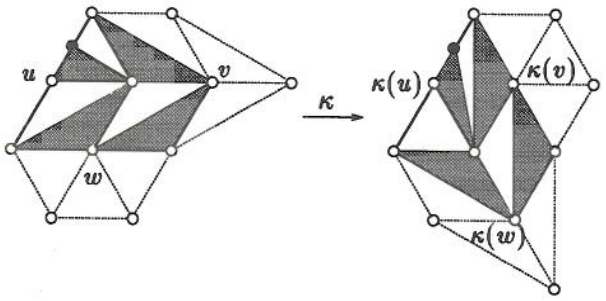

Fig. 7. The knee bends the corridor of an inflation by 60 degrees.

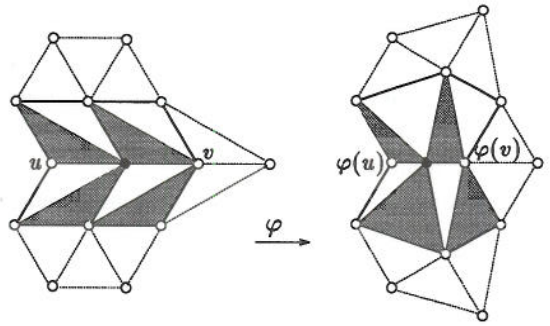

Fig. 8. The fork diverts the impact of an inflation into 2 directions.

Knees and forks

Two particular aggregates are now described. They can be used to redirect the directional effect of an inflation or a deflation. The *knee* bends a corridor by an angle of 60 degrees, see Fig. 7. A knee, κ, is similar to a deflation, δ, followed by an inflation, ι, with dome inside the corridor of δ and direction vector 60 degrees different from that of δ. The difference between κ and $\iota \circ \delta$ is the local deformation in the knee area, which is less for κ.

The *fork* is shown in Fig. 8. It splits a corridor into two. If applied together with an inflation, it halts the deformation along the corridor and propagates half the amount to each of the new corridors. A fork, φ, is similar to a deflation, δ, followed by two half inflations, ι_0 and ι_1. Half inflations will be defined in the next paragraph considering coarse resolution levels. The local deformation in the forking region is again less for φ than it is for $\iota_1 \circ \iota_0 \circ \delta$. The part of the tiling affected by a knee or a fork can be described as the union of 2 or 3 corridors. This is important in the discussion of data structures used in the construction of homeomorphisms.

Coarse resolution levels

For $\varrho \in \mathbb{R}$ define $\varrho T = \{\varrho\sigma \mid \sigma \in T\}$, where $\varrho\sigma = \{\varrho p \mid p \in \sigma\}$ is a scaled copy of σ. Clearly, $1T = T$, and $\varrho T = (-\varrho)T$ because central reflection through the origin belongs to the symmetry group of T. For every positive integer k, T is a subdivision of kT. Consider some fixed $k \geqslant 2$ and an inflation ι defined within kT, see Fig. 9. ι doubles the area of $2k^2$ triangles in T, which is the same amount of growth as achieved by k^2 inflations within T. The number of rows in the corridor of ι is $2k$,

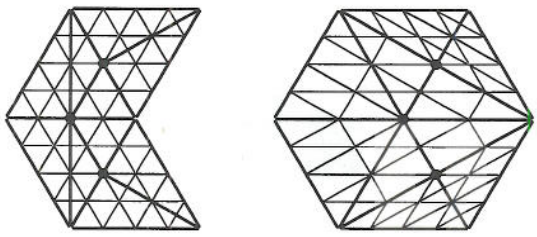

Fig. 9. An inflation in $4T$. The triangles in the diamond of the dome are expanded uniformly.

which is considerably less than the total of $2k^2$ rows in the corridors of the k^2 inflations on the lowest level of resolution. This suggests a hierarchical approach to constructing homeomorphisms as the composition of simple deformations on different levels of resolution. This idea is further developed in Section 4 where only powers of 2 are allowed as values of k. In a sense, we write the homeomorphism in binary notation.

A possibility suggested by revisiting inflations on a coarse resolution level is the idea of a *fractional* inflation: the center vertex, u, is moved by less than one position in the direction $e = u - v$. Indeed, in kT there are $k - 1$ vertices of T strictly between u and $u + e$. Any one of these can be chosen as the image of u. The effect of a fractional inflation is a milder expansion of the diamond region and a milder shift along the corridor. Of course, in principle the image of u can be chosen anywhere along the axis of the inflation and does not even have to be a vertex of T. The only limitation to the freedom of design is the increase in complexity and difficulty of an efficient algorithmic realization.

4. Array-based data structure

We describe a hierarchical data structure keeping track of the homeomorphism as it evolves. Each level of the hierarchy is a 2-dimensional array representing $T_\ell = kT$, with $k = 2^\ell$ and ℓ a non-negative integer. The array also represents the inflations defined within T_ℓ. See also Section 5 where trees are used to represent the hierarchy. Levels are indexed from bottom up, so the lowest level is $T_0 = T$. We begin by describing a single level and then discuss how levels are connected.

Array of queues

The data structure for the tiling T_ℓ is a 2-dimensional array $\mathrm{A}_\ell[i, j]$, $i, j \in \mathbb{Z}$. For now we assume T_ℓ and A_ℓ are infinite in both directions. Each element $\mathrm{A}_\ell[i, j]$ is a vertex of T_ℓ, and the neighboring vertices are given by index pairs $[i-1, j-1]$, $[i-1, j]$, $[i, j-1]$, $[i, j+1]$, $[i+1, j]$, $[i+1, j+1]$, see Fig. 10. The coordinates of the vertex $\mathrm{A}_\ell[i, j]$ in T_ℓ are

$$\begin{pmatrix} 2^\ell & -2^{\ell-1} \\ 0 & 2^{\ell-1}\sqrt{3} \end{pmatrix} \cdot \begin{pmatrix} i \\ j \end{pmatrix}. \tag{1}$$

It is convenient to implicitly assume the transformation and to call i and j the coordinates of the vertex.

An inflation, ι, is specified by a dome, D, determined by its center vertex, u, and its direction vector, $e = e_\iota$. There are only 6 possibilities for e, namely $(-1,-1)$, $(-1,0)$, $(0,-1)$, $(0,1)$, $(1,0)$ and $(1,1)$. To record the changes caused by ι, each vertex $u_\lambda = u + \lambda e$, $\lambda \geqslant 0$ and $\lambda \in \mathbb{Z}$, stores e to

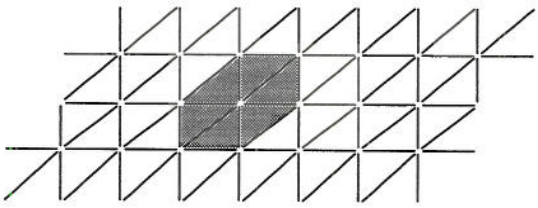

Fig. 10. The regular tiling with squares decomposed into 2 triangles each by mutually parallel diagonals is isomorphic to T.

indicate its motion under ι. We have $\iota(u_\lambda) = u_{\lambda+1}$ for all $\lambda \geqslant 0$ and $\iota^{-1}(u_\lambda) = u_{\lambda-1}$ for all $\lambda \geqslant 1$. The preimage of u_0 is $u_0 - \frac{1}{2}e$, which is not a vertex in A_ℓ. To indicate that ι *creates* u_0 we mark the label e that is stored at u_0.

In the general situation we deal with a sequence of inflations on level ℓ. Let $h = \iota_n \circ \cdots \circ \iota_2 \circ \iota_1$ be their composition, and call the index the *time stamp*. A single vertex $v = \mathrm{A}_\ell[i, j]$ may be affected by several of these inflations, and it stores them in chronological order in a queue of its own, Q_v. The entry $e_t = e_{\iota_t}$ in Q_v indicates the image and preimage of v under ι_t. The most common operation is determining the image of a point by tracing the point through the sequence of simple homeomorphisms. To speed up this operation, we equip e_t with a next-pointer, $\mathrm{n}_{v,t}$, to the entry $e_{t'}$ in $\mathrm{Q}_{\iota_t(v)}$, with minimal $t' > t$. The next-pointer is undefined if there is no such t'. Similarly, we equip e_t with a previous-pointer, $\mathrm{p}_{v,t}$, to the entry $e_{t''}$ in $\mathrm{Q}_{\iota_t^{-1}(v)}$ with maximal $t'' < t$. The previous-pointer is undefined if there is no such t'' or $\iota_t^{-1}(v)$ is not a vertex in A_ℓ. In the latter case, e_t is marked indicating v is created by ι_t.

Tracing points

The image of a point is computed by tracing the path it takes under the sequence of inflations. The queue entries provide sufficient information to trace out this path. The image of a line segment is computed by tracing its endpoints and cutting the line segment into smaller pieces if an inflation affects different parts of the line segment differently. Consider first a vertex v in A_ℓ. The path leading from v to its image under h can be traced by following next-pointers. The path starts with the first entry in Q_v and ends when n is not defined. Similarly, the path from v to its preimage under h can be traced by following previous-pointers. The path starts with the last entry in Q_v and ends when p is undefined.

Every point $x \in \mathbb{R}^2$ can be specified by its nonzero barycentric coordinates with respect to vertices of T_ℓ. In all cases, the image of x is obtained with the same coordinates from the images of the vertices: if $x = q_u u + q_v v + q_w w$ then

$$\iota(x) = q_u \iota(u) + q_v \iota(v) + q_w \iota(w).$$

In many cases, the triangle in T_ℓ that contains $\iota(x)$ is $\iota(uvw)$. Otherwise, it is one of at most 3 triangles overlapping $\iota(uvw)$. Similarly,

$$\iota^{-1}(x) = q_u \iota^{-1}(u) + q_v \iota^{-1}(v) + q_w \iota^{-1}(w),$$

except in cases where the preimage of uvw is not a geometric triangle. For each inflation there are 4 such triangles, namely the ones that partially overlap the image of the dome diamond, see Fig. 4. Such a triangle is cut into half by adding the midpoint of an edge. After computing the new barycentric coordinates of x, considering this midpoint a new vertex, we can find $\iota^{-1}(x)$ as before.

Only minor modifications are necessary to adapt the data structure and path tracing strategy to handle deflations and aggregates. We need pointers along all corridors making up an aggregate. The other changes have to do with the fact that deflations and aggregates do not necessarily map vertices to vertices. This is the case for one vertex in a deflation and can be the case for most of the vertices affected by an aggregate. As a result, an undefined next-pointer for v does not imply that no deformation applies to v, but merely suggests that v needs to be treated like an arbitrary point of $\mathbb{R}^2$ from then on. To

distinguish these two possibilities we add a field to the queue entries that identifies the corresponding deformations.

Hierarchy of tilings and arrays

We call the infinite sequence $T_0, T_1, \ldots$ a *tiling hierarchy* and we use the corresponding sequence of arrays $A_0, A_1, \ldots$ to represent the homeomorphism defined by inflations within the T_i. T_ℓ is a subdivision of $T_{\ell+1}$. More precisely, each triangle in $T_{\ell+1}$ is subdivided into 4 triangles in T_ℓ. It follows the vertex $A_{\ell+1}[i,j]$ is the same as $A_\ell[2i, 2j]$. Only vertices with even coordinates in A_ℓ exist in $A_{\ell+1}$.

By simple index manipulation (multiplication and division by 2) we can go from one array to the next. This does not mean it is easy to compose inflations on different levels in any arbitrary sequence. Indeed, we need to avoid pointers from lower to higher levels. An inflation on a high level can otherwise require a pointer each from a large number of triangle on lower levels. The effort to establish these pointers would defeat the whole purpose of using inflations on high levels to speed up the computations. For this reason we require that the inflations on level $\ell + 1$ precede the inflations on level ℓ, for all ℓ. The gain in efficiency seems to justify the associated loss in flexibility. After all, it is natural to first get the rough features of the homeomorphism right before working on its details.

5. Tree-based data structure

As an alternative to arrays, we can base the representation of the tiling hierarchy on trees. The storage improves to constant per inflation, but tracing a path slows down and is no longer constant time per step. The basic idea is to store corridors with a constant number of descriptors. It is thus suitable for inflations, deflations, knees and forks alike.

Two-key search

Suppose we are given a sequence of integer items, $\phi_1, \phi_2, \ldots, \phi_n$. We seek a data structure for the ϕ_i so the following kind of query can be answered quickly: given a pair of integers (t, x), find the smallest $i > t$ with $\phi_i < x$, see Fig. 11. We store the items in the leaves of a binary tree T [1, Chapter III]. T is balanced and the leaves are sorted by item index from left to right. Each interior

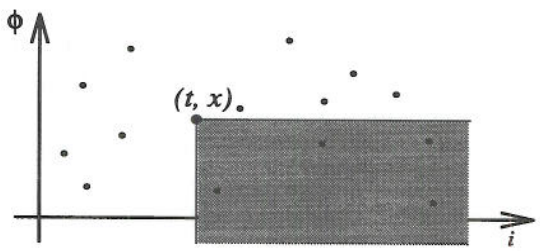

Fig. 11. Plot each item ϕ_i as a point with coordinates (i, ϕ_i) and convert the query into the open orthant below and to the right of the point (t, x). The answer is the leftmost point (i, ϕ_i) inside the orthant.

node $\nu \in \mathrm{T}$ stores the range of indices in its subtree, $[a_\nu, b_\nu]$, and the index, $m(\nu) \in [a_\nu, b_\nu]$, of the smallest item.

To answer a query (t, x), traverse T starting at its root. At an interior node ν we distinguish several cases:

Case 1. $\phi_{m(\nu)} \geqslant x$. Return the empty answer.
Case 2. $\phi_{m(\nu)} < x$.
  Case 2.1. $b_\nu \leqslant t$. Return the empty answer.
  Case 2.2. $t < b_\nu$. Compute the answer for the left subtree and return it if non-empty. Otherwise, compute and return the answer for the right subtree.

The time required to find the answer is proportional to the height of the tree. To see this note that there are 3 types of nodes ν characterized by $a_\nu \leqslant b_\nu \leqslant t$, $a_\nu \leqslant t < b_\nu$ and $t < a_\nu \leqslant b_\nu$. The nodes of second type lie on a single path traversed in Case 2.2 of the algorithm. A node of first type takes only constant time, Case 2.1, and is visited only if it is the root or the child of a node of second type. A node of third type either takes only constant time for one test, Case 1, or it for sure returns a non-empty answer. Also, it is visited only if it is the right child of a node whose left child returns the empty answer. After the first time a node of third type returns a non-empty answer, this case cannot happen any more. If follows that a query is answered by traversing at most 2 paths in T.

Storing corridors

Consider an inflation, ι, on the finest level of resolution, in the tiling $T = T_0$. There are 6 possible direction vectors, and we have a data structure for each possibility. The *axis* of ι is the oriented line $l : x = u + \lambda e$, $\lambda \in \mathbb{R}$, determined by the center vertex, u, and the direction vector, e. Observe the axis splits the dome and its corridor along the middle. The two-key item associated with ι consists of its time stamp or index, i, and the position ϕ_i of the start vertex, $v = u - e$, along l. All inflations with the same axis l are stored in a tree T_l as described above. Similarly, all other deformation types are represented by corridors, and the ones with axis l are stored in T_l, together with the inflation corridors.

For a vertex $x \in T$ on l and a time t, the next deformation affecting x can be computed by searching T_l as explained. Since x belongs to 6 axes, one per direction vector, we need to search 6 trees and take the earliest deformation among the 6 answers.

For each direction vector, there is a countably infinite sequence of parallel axes. Only finitely many are associated with deformations performed by the algorithm, and these are stored in a dictionary, which can be represented as a sorted tree [1]. Each entry in the dictionary corresponds to an oriented line, l, and stores the corresponding tree T_l. To find the next deformation for a vertex $x \in T$ we thus search all 6 dictionaries and one tree inside each dictionary.

The same idea also works for tilings T_ℓ on higher levels of resolution. Note, however, that an axis l in T_ℓ affects vertices on $2^\ell - 1$ lines in T parallel to l. Only the middle line, l itself, also belongs to T_ℓ. If a vertex v lies on a line in T but not in T_ℓ, we need to search 2 axes for this direction vector, namely the 2 axes in T_ℓ next to v on both sides.

The tree-based data structure consists of 6 dictionaries per level of the tiling hierarchy. Each entry stores a non-empty tree T. Each deformation is stored in a leaf each of 6 trees T. The total amount of storage is therefore proportional to n, the number of unit deformations. The cost per step in tracing a

path is $O(\log n)$ time per level of resolution. There are at most $1 + \log_2(N+1)$ levels if $N \geqslant 2$ is an upper bound on the absolute size of the coordinates of all relevant vertices in T.

Intervals of axes

It is interesting to observe that in the tree-based data structure deformations on different levels of resolution can be mixed freely. Indeed, with a minor modification of the data structure it is possible to perform deformations in tilings kT for any positive integer k. The modification substitutes two segment trees [7] for the list of dictionaries per direction vector. They are used to discriminate between deformations affecting an axis in T and deformations that have no effect on this axis. More specifically, the two segment trees store two intervals (of parallel axes) for each corridor. If the deformation is defined in kT, the two intervals consist of the $k-1$ lines on one side of the axis and the k lines on the other side, including the axis itself. The reason the $2k-1$ lines are stored in two intervals, rather than one, is that the starting points at which the deformation affects a line are aligned differently on the two sides of the axis.

A vertex $x \in T$ determines an axis and the two segment trees are used to identify the set of intervals containing this axis. This set is given as the disjoint union of roughly $2\log_2 n$ smaller sets, each represented by a tree T as described above. To answer a query we first search the two segment trees and then search logarithmically many trees T, which in total takes time $O(\log^2 n)$. The entire data structure consists of 12 segment trees, two per direction vector, and the amount of storage it requires is $O(n \log n)$.

6. Measuring quality

The homeomorphism solving a cartogram problem is by no means unique. It is therefore essential to develop a computable notion of quality that permits the comparison of different homeomorphisms. We consider two such notions: the worst and the average distance distortion. The discussion focuses on inflations, and all results can be extended to include other types of deformations.

Linear pieces

An inflation consists of 5 linear (or rather affine) pieces. To describe them, let ι be an inflation with dome D, center vertex u, direction vector e, and corridor C. The diamond of D consists of 2 triangles, σ_ℓ and σ_r, to the left and right of the axis $l: x = u + \lambda e$. The two sides of C (together with the 2 non-diamond triangles of D) to the left and the right of l are denoted C_ℓ and C_r. Outside $D \cup C$, ι is the identity map. Inside $D \cup C$ it consists of a linear map each for σ_ℓ, σ_r, C_ℓ and C_r. In each case, the map is of the form $\iota(x) = Ax + b$, and A and b are readily computed from the vertex coordinates of a single triangle and its image. For example, assuming $e = (1,1)$, the 4 matrices A for σ_ℓ, σ_r, C_ℓ, C_r are

$$\begin{pmatrix} 2 & 0 \\ 1 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix}, \quad \begin{pmatrix} 2 & -1 \\ 1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 \\ -1 & 2 \end{pmatrix}; \tag{2}$$

the corresponding translation vectors b are not important for the discussion in this section. The growth factor for area is the determinant of A, and indeed it is 2 for σ_ℓ and σ_r, and 1 for C_ℓ and C_r.

Singular values

In the above example, A captures the distortion of a piece in the triangulated square tiling. We are interested in the distortion of the plane represented by the regular triangle tiling, so we need to consider BA, where B is the matrix in (1). To simplify the discussion we ignore this important detail.

The effect of A on distance between points is measured by the two *Lipschitz numbers*,

$$L_A = \max\left\{\frac{\|Ax\|}{\|x\|}\,\middle|\, 0 \neq x \in \mathbb{R}^2\right\} \quad \text{and} \quad \ell_A = \min\left\{\frac{\|Ax\|}{\|x\|}\,\middle|\, 0 \neq x \in \mathbb{R}^2\right\},$$

where $\|y\| = |y0|$ is the Euclidean distance of y from the origin $0 \in \mathbb{R}^2$. The distance distortion can be studied by considering the image of the unit circle, $\mathbb{S}^1 = \{x \in \mathbb{R}^2 \mid \|x\| = 1\}$. $A(\mathbb{S}^1)$ is an ellipse determined by the singular value decomposition of A, see [5, Chapter 2.5]. This is a decomposition of the form

$$A = U \cdot \begin{pmatrix} s_1 & 0 \\ 0 & s_2 \end{pmatrix} \cdot V, \tag{3}$$

where U and V are orthonormal matrices (rotations in $\mathbb{R}^2$). The diagonal entries, $s_1 \geqslant s_2$, are the *singular values* of A. s_1 and s_2 are the lengths of the ellipse axes, see Fig. 12. It follows that $L_A = s_1$ and $\ell_A = s_2$. To compute singular values we can use their relationship with the eigenvalues of $A^{\mathrm{T}}A$. Specifically, the singular values are square-roots of the roots of the characteristic polynomial

$$\det\left(A^{\mathrm{T}}A - \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}\right),$$

see [5, Chapter 8.5]. As an example, consider the characteristic polynomial for the first matrix A in (2), which is

$$(5 - \lambda)(1 - \lambda) + 1.$$

Its roots are $\lambda = 3 \pm \sqrt{5}$, and the Lipschitz numbers are $L_A = s_1 = \sqrt{3 + \sqrt{5}} \approx 2.288$ and $\ell_A = s_2 = \sqrt{3 - \sqrt{5}} \approx 0.874$. Note, however, that this only measures the distortion in the decomposed

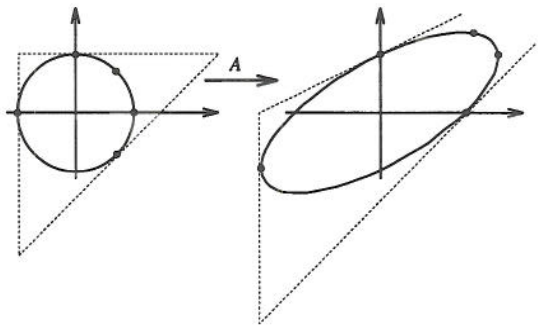

Fig. 12. The matrix expanding the upper triangle of a diamond maps the unit circle to the ellipse with axes of length approximately 2.288 and 0.874.

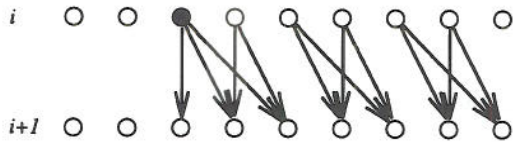

Fig. 13. A row of triangles in T at time i and $i+1$. The shaded node corresponds to the triangle in the dome diamond of ι_i. It expands to twice its original area and it is the only triangle whose image overlaps 3 triangles in T.

square tiling, and the distortion in the corresponding regular tiling T is less. It is given by the Lipschitz numbers

$$L_C = \sqrt{\frac{8+\sqrt{28}}{3}} \approx 2.105 \quad \text{and} \quad \ell_C = \sqrt{\frac{8-\sqrt{28}}{3}} \approx 0.950$$

of $C = B \cdot A \cdot B^{-1}$, where B is the matrix in (1).

Dag of matrices

In general, we study a piecewise linear homeomorphism $h = \iota_n \circ \cdots \circ \iota_2 \circ \iota_1$. The deformation of each linear piece is given by a product of matrices as in (2). To determine the singular values of such a product there is little else we can do than first compute the product matrix and then its singular values. All matrix entries are small integers, $-2, -1, 0, 1, 2$, so we can compute products exactly. We can improve the overall running time by exploiting that the matrix chains overlap in possibly large portions.

Specifically, we define a directed acyclic graph, G, whose nodes are triangles in T with time stamps in $\{0, 1, \ldots, n\}$. The source nodes are triangles with time stamp 0 and the sink nodes have time stamp n. There is an edge from (σ, i) to $(\sigma', i+1)$ if $\iota_i(\sigma)$ and σ' overlap. Each edge is labeled with the corresponding matrix, A. Paths from source to sink nodes are chains of matrices. Most edges are labeled with the identity matrix, and we may choose to ignore them by identifying their 2 triangles, which are the same anyway only with different time stamps. By definition of inflation, each node has at most 2 incoming and at most 3 outgoing edges, see Figs. 13 and 3. The dag G is implicit in the data structures described in Sections 4 and 5, which provide images of triangles indirectly through the images of their vertices. The composition of the n inflations can be computed by traversing the entire G from sources to sinks. A breadth-first traversal avoids recomputing products of partial matrix chains [1, Chapter VI.23]. A triangle (σ, i) is the last node of a number of initial matrix chains, each associated with a piece of σ. The pieces subdivide σ. If (σ, i) has 2 (or 3) outgoing edges then this does not imply each of its pieces is split into 2 (or 3). Indeed, many paths do not correspond to any non-empty linear piece of h. Computing matrix products along such paths can be avoided by keeping track of the pieces subdividing each σ.

Distance distortion

The extreme distance distortions are captured by the Lipschitz numbers L_h and ℓ_h, which are the maximum and minimum of

$$\left\{ \frac{|h(x)h(y)|}{|xy|} \;\middle|\; x \neq y \in \mathbb{R}^2 \right\}.$$

Both can be computed from the matrix products at the sink nodes of G. Another relevant quality measure is the *condition number*,

$$C_h = \max_{\zeta} \left\{ \frac{L_{h|\zeta}}{\ell_{h|\zeta}} - 1 \right\},$$

defined over all maximal regions $\zeta \subseteq \mathbb{R}^2$ so $h|\zeta$, the restriction of h to ζ, is linear. Large ratios occur in transition regions between small and large area distortions. Small condition numbers indicate good quality homeomorphisms. More tolerant to large condition numbers in small regions than C_h is the *average condition number*,

$$\overline{C}_h = \sum_{\zeta} \text{area}(\zeta) \cdot C_{h|\zeta}.$$

By definition, the condition number $C_{h|\zeta}$ vanishes in undeformed regions of $\mathbb{R}^2$, so their contribution to $\overline{C}_h$ is zero.

7. Implementation and results

The performance of our approach in terms of quality of generated cartograms is illustrated with examples redrawing the states of the USA. The cartogram problem asks for deforming each state so its area represents the number of electoral votes in the 1992 presidential election.

Our current implementation is based on trees representing the tiling hierarchy. We show the images of two homeomorphisms. The first is produced by our implementation of a matching strategy. The current implementation is less than perfect and tends to produce homeomorphisms of inferior quality. We believe the implementation can be improved to achieve good quality over a broad range of cartogram problems. The second homeomorphism is produced with a more conventional iteration scheme. While the latter approach gives a pleasant and accurate homeomorphism in this specific cartogram problem, this cannot be said for some other problems we tried.

Matching strategy

One idea for finding simple deformations that compose to a desired homeomorphism is to transport area from regions of surplus to regions of need. An inflation-deflation pair with matching corridors transports area from the deflation to the inflation diamond. If no common corridor is available, we use knees and forks to design more complicated transportation channels.

The way surplus regions are matched with regions of need has a profound influence on the quality of the generated homeomorphisms and the running time of the algorithm. At the same time, this is currently the least developed part of our method, that is, the matching algorithm we use is heuristic. We plan noncrossing transportation channels, if possible. To limit the number of crossings among the channels, we first pick a direction and choose all initial inflation-deflation pairs with corridors along this direction. Channels containing knees and forks transport the remaining surplus to regions of need. With this strategy, it is possibly to inflate or deflate different portions of the same district or state in a uniform manner.

To simplify the matching problem, we replace the deflation operation described in Section 3 by a similar but different aggregate. It restricts the area reduced by the deformation to within a diamond of two triangles, just as in the case of an inflation. We use deflation factors ranging from 0.5 through $1 - 1/(\varrho + 1)$ and inflation factors ranging from $1 + 1/\varrho$ through 2.0 on resolution level $\varrho \geqslant 1$. The use of fractional inflations and deflations has a positive effect on the quality of the resulting homeomorphism.

It was found that visually unpleasant side effects of deformations decrease with finer tilings. Experience seems to suggest that the triangle edges should be shorter than the "smallest feature size" of the geographic map by a factor between 10 and 20. Finer tilings do not necessarily imply an increase in the number of inflations and deflations because many operations are executed on coarse resolution levels.

Sample deformations

Fig. 14 shows the geographically accurate decomposition of the USA into states, not including Alaska and Hawaii. The homeomorphisms shown in Figs. 15 and 16 deform the continental USA so the area of the states reflects the number of electoral votes in the 1992 presidential elections. Shading in Fig. 16 indicates states with a majority of votes for the elected president, Bill Clinton. Fig. 15 shows the deformation of the USA at an intermediate stage of the matching algorithm, with average cartographic error of about 50%. In [3], the *error* for a district or state ζ is defined as

$$\max\left\{\frac{a(\zeta)}{t(\zeta)}, \frac{t(\zeta)}{a(\zeta)}\right\} - 1,$$

where $a(\zeta)$ is the actual area (after deformation) and $t(\zeta)$ is the targeted area in the cartogram. We use the same definition of error to allow for direct comparisons with results in [3].

The cartogram in Fig. 16 has average cartographic error of less than 1% and the maximum error for any state is 2.33%. In the computation leading to this cartogram we use a heuristic built on top of the methods described in earlier sections of this paper. The heuristic iterates and changes the area of each state by some small amount towards the desired area. In each step, a set of inflations or deflations are applied to the diamonds inside the state to achieve the desired change.

Fig. 14. The states of the USA approximating actual shape and size.

Fig. 15. The deformed United States at an intermediate stage of the computations.

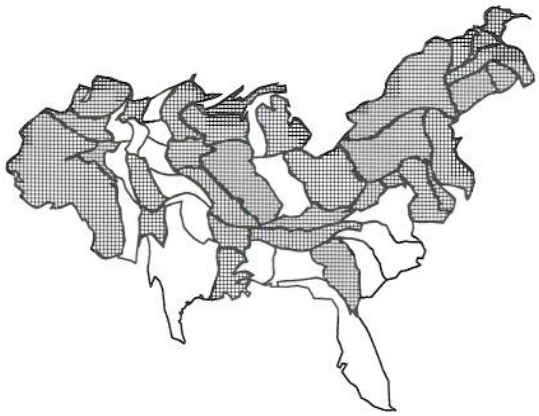

Fig. 16. A cartogram of the United States based on the electoral votes in the 1992 presidential elections.

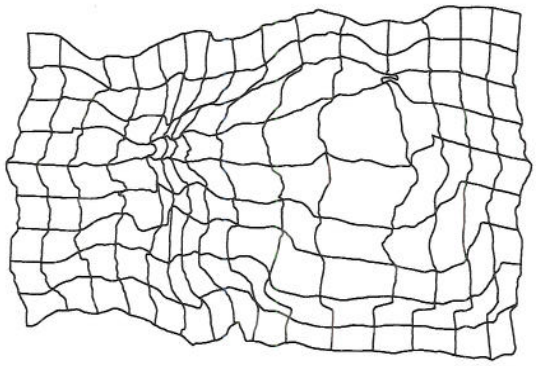

Fig. 17. The image of a regular square grid under the homeomorphism of Fig. 16.

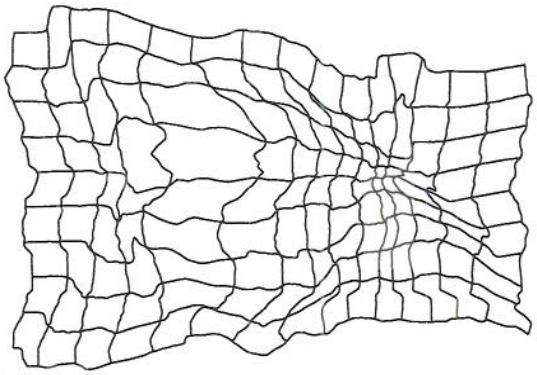

Fig. 18. The preimage of a regular square grid under the homeomorphism of Fig. 16.

The deformation resulting from the application of a homeomorphism is intuitively visualized through the image and the preimage of the regular square grid, see Figs. 17 and 18.

8. Remarks

This paper describes a combinatorial approach to homeomorphisms generating cartograms. Even though combinatorial methods are introduced to describe homeomorphisms and to manipulate them, the algorithm which takes advantage of the discretization of the problem still needs improvement to produce accurate as well as visually pleasing cartograms. In addition there is hope that a post-processing step based on quality measurement mechanisms suggested in Section 6 can eliminate some of the unpleasant effects introduced by the matching strategy.

It is premature to judge the direct practical relevance of the algorithm described in this paper. Besides the quality issue, the applicability of the algorithm will largely depend on the number of

basic operations used in the construction, since experience shows that a truthful mapping of region boundaries can result in a very large number of line segments.

Acknowledgements

The authors thank Jack Snoeyink for bringing the cartogram problem to their attention, and Michael McAllister for providing pointers to the literature on cartograms.

References

[1] T.H. Cormen, Ch.E. Leiserson and R.L. Rivest, Introduction to Algorithms (MIT Press, Cambridge, MA, 1990).
[2] D. Dorling, Map design for census mapping, Geographic J. 30 (1993) 167–179.
[3] J.A. Dougenik, N.R. Chrisman and D.R. Niemeyer, An algorithm to construct continuous area cartograms, Professional Geographer 37 (1985) 75–81.
[4] P.J. Giblin, Graphs, Surfaces and Homology (Chapman and Hall, London, 1981) second edition.
[5] G.H. Golub and Ch.F. van Loan, Matrix Computations (Johns Hopkins Univ. Press, Baltimore, 1989) second edition.
[6] P.M. Gruber and J.M. Wills, eds., Handbook of Convex Geometry, Volumes A and B (Elsevier, North-Holland, Amsterdam, 1993).
[7] K. Mehlhorn, Data Structures and Algorithms 3: Multi-dimensional Searching and Computational Geometry (Springer, Heidelberg, 1984).
[8] J.R. Munkres, Elements of Algebraic Topology (Addison-Wesley, Redwood City, CA, 1984).
[9] E. Schulte, Tilings, in: P.M. Gruber and J.M. Wills, eds., Handbook of Convex Geometry, (Elsevier, North-Holland, Amsterdam, 1993).
[10] W.R. Tobler, A continuous transformation useful for districting, Ann. New York Acad. Sci. 219 (1973) 215–220.
[11] M.S. White Jr and P. Griffin, Piecewise linear rubber-sheet map transformation, American Cartographer 12 (1985) 123–131.