Reconstructing Points on a Circle from Labeled Distances

David Bremner* Erik D. Demaine† Perouz Taslakian‡ Godfried Toussaint‡

Abstract

In the *labeled beltway problem*, we are given a set of points with unknown coordinates, their clockwise ordering around the circumference of a circle, and a coloring of the edges between pairs of points. The goal is to embed the points around a unit circle while satisfying the constraint that two incident edges have the same geodesic length on the circle if and only if they have the same color. We give a polynomial-time algorithm to find such an embedding or to determine that no such embedding exists.

1 Introduction

In this paper we study a reconstruction problem where we are required to embed a set of points along the circumference of a circle that satisfy some constraints on the pairwise interpoint intervals.

Reconstruction problems have connections with many areas such as music theory, crystallography, and DNA sequencing [3, 8]. One variation, called the *turnpike problem*, dates back to the 1930's in the area of X-ray crystallography, where the objective was to reconstruct the coordinates of the atoms in a crystal. More recently, reconstruction problems have become important in the field of DNA *sequencing* — determining the pattern of the amino acids that constitute a strand of DNA. Given a DNA molecule, exposing it to a special kind of enzyme (called *restriction enzyme*) divides it into pieces of different lengths. The lengths of these fragments can be measured with standard techniques, and the challenge is to reconstruct the original ordering of these fragments in the DNA molecule. In molecular biology, this reconstruction problem is called the *partial digest* problem and is equivalent to the turnpike problem in crystallography.

In the context of music, the geodesic distances (along the circumference) between the points on a circle correspond to duration (time) intervals between onsets, in the case of rhythm, and pitch intervals for chords, scales, and melodies [9]. Questions concerning the existence and constructibility of melodies and rhythms from partial information is a well-studied problem in music theory, music perception, and music information retrieval. One popular method of encoding rhythms and melodies with only partial information is via rhythmic and melodic *contours*. The rhythmic contour is defined as the pattern of successive relative changes of durations in a rhythm. A pitch (or melodic contour) is defined in the same way [7]. Two types of contours have received a great deal of attention in the music literature: the *adjacent* contours that use the matrix of distances between pairs of consecutive points on the circle, and the *full* contours that use the entire matrix of distances between all pairs of points on the circle. An interesting composition problem in music theory is whether a given set of intervals (adjacent or full) admits a realization as a melody or rhythm [6]. Demaine et al. [1] give efficient algorithms for determining whether a rhythm may be reconstructed. In [2], Demaine et al. characterize all rhythms where each distance defined by pairs of points appears a unique number of times. Their characterization also provides an algorithm for generating such pointsets. In this paper we are concerned with a further reduction of information compared to that contained in the contours explored in the music literature to date. Rather than specifying whether interval lengths increase, decrease, or stay the same, we only use one bit of information for comparing two intervals: either they have the same duration or they do not. In what follows, we describe a reconstruction problem we call the *labeled beltway problem*, and provide a polynomial-time solution.

From a theoretical perspective, problems related to reconstructing sets from interpoint distances are, in general, computationally challenging. Even when the points are restricted to a line (turnpike or partial digest problem), the complexity of the problem remains unknown. One of the main difficulties of these problems lies in the fact that while a given pointset uniquely defines a multiset of distances, the inverse is not always true: there may be many pointsets defining a given multiset of distances; such pointsets are known as *homometric* sets. Lemke et al. [3] study the computational and combinatorial complexity of reconstruction problems. For a given set of $\binom{n}{2}$ distances, they give upper and lower bounds on the number of mutually noncongruent and homometric n-point sets that realize these distances in $\mathbb{R}^d$. They also show that the decision problem of whether a multiset of $\binom{n}{2}$ distances is realized by n points in $\mathbb{R}^d$ (for arbitrary dimensions d) is NP-complete.

Lemke et al. [3] call the version of the reconstruction

*University of New Brunswick, bremner@unb.ca
†Massachusetts Institute of Technology, edemaine@mit.edu
‡McGill University, {perouz,godfried}@cs.mcgill.ca

problem where the points lie on a circle the *beltway problem*. Here we consider a variation of the beltway problem, the *labeled beltway*, where the edges defined by the points along the circle are assigned labels. The problem can be stated as follows: given a set of points and their clockwise order around a circle, we want to find an embedding of the points on the circumference of a unit circle subject to some constraints. The constraints involve the *geodesic distance* between pairs of points, that is, the length of the shortest path between two points along the circumference of the circle. The distance constraints are given by a labeling of the edges defined by pairs of points such that, if two distinct edges sharing a common endpoint have the same label, then their geodesic distances must be equal. Moreover, if two distinct edges sharing a common endpoint have distinct labels, then their geodesic distances must be distinct. We give a polynomial-time solution to the labeled beltway problem by reducing it to solving a set of linear equations and nonequations.

Note that if, together with the ordering, we were given all the distances between pairs of points, the problem becomes easy in the plane, as well as on the circle. If the points are embedded on a circle, then we have two cases: if the lengths of the convex-hull edges sum to 1, then the given edge lengths measure the clockwise distance between pairs of points consecutive around the circumference, and we can embed. Otherwise, the longest edge of the convex hull equals the sum of the lengths of the remaining edges; in this case, invert the length of the longest edge around 1; all the edges now have clockwise distances and we can again embed.

2 Definition and Notation

Let $\{p_0, p_1, \ldots, p_{n-1}\}$ be a set of points embedded on a unit circumference circle such that $p_{(i+1) \bmod n}$ is the closest point to p_i in the clockwise traversal along the circumference starting from p_i for $i = 0, 1, \ldots, n-1$.[1] The *clockwise distance* between p_i and p_j, denoted by $\overrightarrow{d}(p_i, p_j)$, is the length of the clockwise traversal along the circumference of the circle from p_i to p_j. The *geodesic distance* between p_i and p_j, denoted by $d(p_i, p_j)$, is the length of the minimum clockwise and counterclockwise traversals along the circumference of the circle between p_i and p_j. We say that an edge p_ip_j is *oriented clockwise* if its geodesic distance is the length of the traversal from p_i to p_j along the circumference of the circle; it is *oriented counterclockwise* if its geodesic distance is the length of the traversal from p_j to p_i along the circumference. If the length of an edge is $\frac{1}{2}$, then the edge can be oriented either way; but otherwise, its orientation is forced. We specify the constraints on the geodesic distances between pairs of points by associating these distances with edges of the complete graph on the n points.

[1]Henceforth, all index arithmetic is modulo n.

3 Reduction to Linear Equations and Nonequations

We first state the constraints on the edges of the labeled complete graph more formally as follows: for every three distinct points p_i, p_j, p_k on the circle, if the edge p_ip_j has the same label as p_jp_k, then the geodesic distance between p_i and p_j is equal to the distance between p_j and p_k, that is, $d(p_i, p_j) = d(p_j, p_k)$ (*isometry constraints*); moreover, two edges with different labels must have different geodesic lengths (*anisometry constraints*).

We show how to reconstruct the points given isometry and anisometry constraints between pairs of edges that do not cross (that is, edges whose endpoints appear together in the cyclic order). In particular, such constraints capture all constraints, solving the problem. Our solution is based on representing the constraints by linear equations ($\sum_i a_i x_i = b_1$), linear inequalities ($\sum_i a_i x_i \le b_i$ and $\sum_i a_i x_i < b_i$), and linear nonequations ($\sum_i a_i x_i \ne b_i$) on n variables. We then solve this system by reducing to a sequence of linear programs with only linear equations and inequalities.

We parameterize the desired embedding of the n points on the circle by variables $x_0, x_1, \ldots, x_{n-1}$, where $x_i = \overrightarrow{d}(p_i, p_{i+1})$, the clockwise distance from p_i to p_{i+1}. These n variables determine an embedding up to rotation provided that they satisfy the following constraints (defining an open $(n-1)$-simplex):

$$\begin{aligned} \sum_{i=0}^{n} x_i &= 1, \\ x_i &> 0 \qquad \text{for } i = 0, 1, \ldots, n-1. \end{aligned}$$

The positivity constraints force the points to embed to distinct locations in the correct cyclic order.

The challenge in representing an isometry or anisometry constraint among geodesic distances is that a geodesic distance between two points p_i and p_j may be realized by either the clockwise distance $x_i + x_{i+1} + \cdots + x_{j-1}$ or the counterclockwise distance $x_j + x_{j+1} + \cdots + x_{i-1}$. If we knew the orientation of every edge, then we could write the isometry or anisometry constraint as a linear equation or nonequation. There are n choices for the orientations of the $n-1$ edges incident to a vertex p_i, depending on which wedge between consecutive edges contains the center of the circle. Considering all vertices, there are at most n^n choices for the orientations, each leading to a system of linear equations and nonequations. On the other hand, consider the n pairwise crossing edges $p_ip_{i+n/2}$; the orientation of each can be chosen independently. Hence, there are at least 2^n possible orientations.

For the edges that do not cross, the isometry and anisometry constraints can be reconstructed easily. Two edges p_ip_j and p_kp_l do not cross if their endpoints appear in the order p_i, p_j, p_k, p_l along the circumference. Fortunately, for constraints between such noncrossing edges, we can effectively determine the orientations of the edges to obtain a single system of linear equations and nonequations. To describe these constraints we first need the following lemma:

Lemma 1 *If the points p_i, p_j, p_k, p_l appear in clockwise order around the circle (with the possibility that $j = k$ or $l = i$ but not both) then we cannot have both edges p_ip_j and p_kp_l oriented counterclockwise.*

Proof. Suppose both edges are oriented counterclockwise. This implies that $\vec{d}(p_i, p_j) \geq \frac{1}{2}$ and $\vec{d}(p_k, p_l) \geq \frac{1}{2}$. Because of the ordering of the points around the circle, both p_k and p_l lie on the clockwise arc from p_j to p_i. Thus,

$$\vec{d}(p_j, p_i) = \vec{d}(p_j, p_k) + \vec{d}(p_k, p_l) + \vec{d}(p_l, p_i).$$

This equality cannot hold because $\vec{d}(p_j, p_i) \leq \frac{1}{2}$ while the right-hand-side is strictly greater than half. Thus, at least one of the edges must be oriented clockwise. □

We start with the isometry constraints, for which we need the following lemma:

Lemma 2 *If the points p_i, p_j, p_k, p_l appear in clockwise order around the circle (with the possibility that $j = k$ or $l = i$ but not both) such that $d(p_k, p_l) = d(p_i, p_j)$, then both edges p_ip_j and p_kp_l must be oriented clockwise.*

Proof. By Lemma 1 we know that we cannot orient both edges counterclockwise. Suppose only p_kp_l is oriented counterclockwise; this means that $d(p_k, p_l) = \vec{d}(p_l, p_k) \leq \frac{1}{2}$ and $d(p_i, p_j) = \vec{d}(p_i, p_j)$. Because of the ordering of the points p_i, p_j, p_k, p_l around the circle, we have

$$\vec{d}(p_l, p_k) = \vec{d}(p_l, p_i) + \vec{d}(p_i, p_j) + \vec{d}(p_j, p_k) \leq \frac{1}{2}.$$

Because at least one of the distances $\vec{d}(p_l, p_i)$ and $\vec{d}(p_j, p_k)$ is greater than zero, $\vec{d}(p_l, p_k) > \vec{d}(p_i, p_j)$ which implies that $d(p_k, p_l) > d(p_i, p_j)$; contradiction. Similarly, if only p_ip_j is oriented counterclockwise, then this would imply that $d(p_i, p_j) > d(p_k, p_l)$. Therefore, if $d(p_i, p_j) = d(p_k, p_l)$ then both edges must be oriented clockwise. □

Thus, if two edges p_ip_j and p_kp_l are noncrossing and their edge lengths must be equal, then by Lemma 2 we can assume that the edges are oriented clockwise and we force this by adding two linear inequalities:

$$\sum_{i \leq r < j} x_r \leq \frac{1}{2}, \quad \sum_{k \leq s < l} x_s \leq \frac{1}{2}.$$

Now that we have forced the edges to be oriented clockwise, we can force the equality of their lengths with a linear equation:

$$\sum_{i \leq r < j} x_r \quad = \quad \sum_{k \leq s < l} x_s.$$

For the anisometry constraints, we first show the following lemma:

Lemma 3 *If the points p_i, p_j, p_k, p_l appear in clockwise order around the circle (with the possibility that $j = k$ or $l = i$ but not both) such that at least one of the edges p_ip_j and p_kp_l must be oriented counterclockwise, then $\vec{d}(p_i, p_j) \neq \vec{d}(p_k, p_l)$.*

Proof. Assume that p_kp_l is oriented counterclockwise; then, $\vec{d}(p_l, p_k) \leq \frac{1}{2}$. By Lemma 1, p_ip_j must be oriented clockwise. Then, we have $\vec{d}(p_l, p_k) = \vec{d}(p_l, p_i) + \vec{d}(p_i, p_j) + \vec{d}(p_j, p_k) \leq \frac{1}{2}$. Because at least one of $\vec{d}(p_l, p_i)$ or $\vec{d}(p_j, p_k)$ is strictly positive, then $\vec{d}(p_k, p_l) > \vec{d}(p_i, p_j)$. Thus, the two edges have distinct clockwise lengths. □

Now consider an anisometry constraint between two noncrossing edges p_ip_j and p_kp_l. We represent this constraint by a linear nonequation:

$$\sum_{i \leq r < j} x_r \quad \neq \quad \sum_{k \leq s < l} x_s.$$

To show that the above nonequation holds independent of the orientation of each of the edges, we need to show that it holds if and only if the desired geodesic distances are distinct. First, if both edges can be oriented clockwise, then this constraint on the clockwise distances is equivalent to the distinctness of the geodesic lengths. Second, if one of the edges must be oriented counterclockwise, then by Lemma 2, the geodesic lengths must be different; and by Lemma 3, the linear nonequation must be satisfied.

4 Solving Systems of Linear Equations and Nonequations

The previous section yields a linear system of the following form:

$$Ax \leq \tfrac{1}{2} \tag{1}$$
$$Mx = f \tag{2}$$
$$Nx \neq g \tag{3}$$
$$x_i > 0 \tag{4}$$

Geometrically, (1), (2) and (4) define a (partly open) convex polyhedron P, while each row of (3) defines a forbidden hyperplane. To solve this system, we will use the following idea: initially we ignore the forbidden hyperplanes and find a relative interior point in the feasible region P. Then we check if this point lies on a forbidden hyperplane; if it does not lie on any forbidden hyperplane, we have a solution. Otherwise, if the point lies on h_i, then we recurse on one side of $h_i \cap P$. Recursing on one side of h_i ensures that we never pick a point on the same hyperplane more than once. Thus, in the worst case we will eventually run out of forbidden hyperplanes and one side of the last eliminated hyperplane will be a region with no forbidden points. We can now find a feasible point in this region.

After a suitable (and polynomial time computable) change of coordinates, we may assume that P has interior points. Let $h(a, \mu)$ denote the hyperplane $\{x \mid \langle a, x \rangle = \mu\}$.

Proposition 4 *If $h(a, \mu) \cap P$ has relative interior points, then for any sufficiently small $\varepsilon > 0$, $h(a, \mu+\varepsilon)$ and $h(a, \mu - \varepsilon)$ both have relative interior points.*

The idea is that given a point q in the relative interior of $P \cap h(a, \mu)$ (if such a q does not exist, then $h(a, \mu)$ is redundant and can be discarded), choose ε sufficiently small so that $h(a, \mu + \varepsilon) \cap P$ has relative interior, and does not collide with any possible parallel forbidden hyperplanes. If a solution exists, then there is one on the hyperplane $h(a, \mu + \varepsilon)$. To avoid calculation of ε, we add the inequality $a^T x > \mu$ to our system, replacing the nonequation $a^T x \neq \mu$. If this new system is infeasible, it means that $P \subset h(a, \mu)$, and the original problem is infeasible.

Our basic computational step is thus to find a relative interior point of a polyhedron defined by strict and nonstrict inequalities. This is equivalent to feasibility testing for systems of strict linear inequalities, which can be solved by linear programming.

The inequalities and (non)equations of our system describe the constraints for pairs of noncrossing edges. We have added a constant number of constraints per pair of noncrossing edges having the same or different labels. Because we have a total of $\binom{n}{2}$ edges, pairing them gives us $\Theta(n^4)$ constraints. The linear programs required can be solved in time polynomial in n and the number of bits required of the output. In our case, $\Theta(n)$ bits suffice to disambiguate all the distinct distances, at which point the solution will be correct; so we can solve the linear programs in time polynomial in n. The number of times we solve such a system is at most equal to the number of forbidden hyperplanes, which is a polynomial function of n. Thus we have a polynomial-time solution to the labeled beltway problem.

5 Open Problem

A natural variation on the labeled beltway problem treats the coloring as global instead of local: instead of just constraining the equality or inequality of the geodesic lengths of *incident* edges, it constrains the relative lengths of all edge pairs. This problem seems substantially more difficult because it becomes possible to construct specific numbers (less than 1) with $\Theta(n)$ bits. After some effort, it is not clear to us whether this global version of the labeled beltway problem is polynomially solvable or NP-complete.

Acknowledgments

This work was initiated at the 21st Bellairs Winter Workshop on Computational Geometry held January 27–February 3, 2005 at Bellairs Research Institute. We thank the other participants of that workshop for helpful discussions and contributing to a fun and creative atmosphere. Perouz also thanks Victor Campos for useful and entertaining discussions on the topic of this paper.

References

[1] E. D. Demaine, J. Erickson, D. Krizanc, H. Meijer, P. Morin, M. Overmars, and S. Whitesides. Realizing partitions respecting full and partial order information. *Journal of Discrete Algorithms*, 6:51–58, 2008.

[2] E. D. Demaine, F. Gomez-Martin, H. Meijer, D. Rappaport, P. Taslakian, G. T. Toussaint, T. Winograd, and D. R. Wood. The distance geometry of music. *Computational Geometry: Theory and Application*, 2008. To appear.

[3] P. Lemke, S. S. Skiena, and W. D. Smith. Reconstructing sets from interpoint distances. Technical Report 37, DIMACS Technical Report, 2002.

[4] R. D. Morris. New directions in the theory and analysis of musical contour. *Music Theory Spectrum*, 15(2):205–228, 1993.

[5] I. Quinn. Fuzzy extensions to the theory of contour. *Music Theory Spectrum*, 19(2):232–263, 1997.

[6] J. Rahn. Possible and impossible melodies: Some formal aspects of contour. *Journal of Music Theory*, 36(2):259–279, 1994.

[7] I. Shmulevich. A note on the pitch contour similarity index. *Journal of New Music Research*, 33(1):17–18, March 2004.

[8] S. S. Skiena and G. Sundaram. A partial digest approach to restriction site mapping. *Bulletin of Mathematical Biology*, 56:275–294, 1994.

[9] G. T. Toussaint. A mathematical analysis of African, Brazilian, and Cuban *clave* rhythms. In *Proceedings of BRIDGES: Mathematical Connections in Art, Music and Science*, pages 157–168, Towson University, Towson, Maryland, U.S.A., July 27-29 2002.