BISECTION OF CIRCLE COLORINGS*

CHARLES H. GOLDBERG† AND DOUGLAS B. WEST‡

Abstract. Consider 2n beads of k colors arranged on a necklace, using 2a_i beads of color i. A bisection is a set of disjoint strings ("intervals") of beads whose union captures half the beads of each color. We prove that any arrangement with k colors has a bisection using at most \( \lceil k/2 \rceil \) intervals. In addition, if k is odd, an endpoint of one interval can be specified arbitrarily. The result is best possible. For fixed k, there is a polynomial-time algorithm to find such a bisection; it runs in \( O(n^{k-2}) \) for \( k \geq 3 \). We consider continuous and linear versions of the problem and use them to obtain applications in geometry, VLSI circuit design, and orthogonal functions.

AMS(MOS) classifications. 05A, 68C, 55

1. Introduction. Consider necklaces formed using two colors of beads. If an even number of each color is used, it is easy to show that a single interval can be chosen that will capture exactly half the beads of each color. Start with any interval capturing half the total number of beads. If it is short in color A and has too much of color B, then the complementary interval is imbalanced the other way. Sliding from an interval to its complement one bead at a time changes the imbalance by at most one bead at a time, so there must be some intermediate stage where the colors are in balance, i.e. where the interval contains half of each color.

A natural generalization of this problem arose in the study of VLSI circuit design. Bhatt and Leiserson [1] asked the corresponding question for 3-color necklaces, i.e. whether it is always possible to chose two intervals that together capture half the beads of each color. In this paper we prove that this is true, and in fact we obtain the best possible result for the general case of k colors. In order to do this, we prove a stronger result about continuously integrable functions on the circle. The continuous result has several geometric applications, and the discrete result has an application to "graph separators."

First we state the discrete result. Suppose 2n beads of k colors are placed around a circle, using 2a_i beads of color i; such a configuration is called a necklace or discrete coloring. A bisection of a coloring is a set of nonoverlapping intervals on the circle whose union contains exactly half of each color. The size of the bisection is the number of intervals used. We prove the following theorem.

Theorem 1. Every necklace with k colors of beads has a bisection using no more than \( \lceil k/2 \rceil \) intervals. No smaller size suffices for all arrangements of k colors. If k is odd, this size suffices even if one of the intervals is required to end between a specified pair of beads.

Of course, corresponding to this extremal problem, there is an optimization problem. Given a particular arrangement of beads, what is the bisection of smallest size? The complexity of this problem is open; it may be NP-complete. When the number of colors in the arrangement is bounded by k, the proof of Theorem 1 leads to an efficient algorithm to find a bisection of size at most \( \lceil k/2 \rceil \). Its running time is bounded by a polynomial in 2n, the number of beads in the arrangement. After time

* Received by the editors July 25, 1983, and in final form December 19, 1983. This work was presented at the SIAM Second Conference on the Applications of Discrete Mathematics, held at Massachusetts Institute of Technology, Cambridge, Massachusetts, June 27–29, 1983.
† Mathematics Department, Trenton State College, Trenton, New Jersey 08625.
‡ Mathematics Department, University of Illinois, Urbana, Illinois 61801.
\(O(n)\) for preprocessing, the algorithm runs in \(O(n^{k-2})\) if \(k \geq 3\). In particular, it is linear if \(k = 3\).

The direct proof given above for the 2-color case becomes increasingly ugly as the number of color increases, because no longer does the imbalance in the captured colors change by “at most one” when the parameters describing the choice of intervals change by one. We initially obtained the proof for \(k \leq 4\) by this direct method, but here we will present only a more elegant approach that works equally well for arbitrarily many colors. This proof uses a natural extension to a continuous problem. Paint the perimeter of the circle with various colors. The colors no longer need be restricted to disjoint intervals; instead, they may “mix.” We require only that each color have a continuously integrable density function (i.e., the cumulative distribution is continuous), and that the densities of the various colors sum to 1 at each point. (Note that some “densities” may actually be negative.) We call such a painting a circle coloring. Bisection is defined as before. The previous result holds again in the continuous case.

**Theorem 2.** Every circle coloring with \(k\) colors has a bisection using at most \([k/2]\) intervals. When \(k\) is odd, one endpoint of one interval may be chosen arbitrarily. No smaller size suffices for all circle colorings with \(k\) colors.

The proof uses methods from topology. An appropriate parameter space is defined to describe the possible placements of \([k/2]\) intervals, and the cumulative density functions map this space continuously into a \((k-1)\)-hyperplane. The coordinates of the hyperplane sum to \(\sum a_i\) and the “target point” \((a_1, \ldots, a_n)\) corresponds to capturing half of each color. Any inverse image of this point is a bisection. To show that this inverse image is nonempty, we want to show that the target point is “inside” the image of the boundary and apply continuity. In topological terms, “inside” is related to winding number; we show that if the boundary of the parameter space does not hit the target point, then it maps to a surface that has odd winding number relative to the target point. This odd parity follows by induction on the number of colors; we construct a problem with fewer colors in which the winding number of the boundary has the same parity. For example, a problem with fewer colors could be obtained by amalgamating the last two colors. That this all works depends heavily on a simple fact used in the 2-color argument mentioned at the beginning. The complement of a set of \([k/2]\) disjoint intervals on the circle is also a set of \([k/2]\) disjoint intervals on the circle, and it captures a complementary amount of each color. In particular, the complement of a bisection is also a bisection.

There are several applications of both the discrete and the continuous result. The continuous result can be viewed as a version applicable in projective space of the “intermediate value theorem” of single-variable calculus. In the four-color case, the continuous result implies that for any (directed) closed curve in \(\mathbb{R}^3\), there are two equal and opposite chords that together cut off half the length of the curve. In \(\mathbb{R}^2\), this can be strengthened to obtain a rectangle, or right angle subtended at any point using two other points on the curve, etc. These arguments lead to a proof of an old geometry conjecture for the special case of “Nice curves” in \(\mathbb{R}^2\). The conjecture is that any closed curve in \(\mathbb{R}^2\) contains four points that determine a square. The conditions and proof are similar to those of Jerrard [4] (see also [5]). The unproved case is that of nondifferentiable curves, for which these methods are less well suited.

The discrete result applies to “separators,” a construct used in VLSI design theory [6]. For present purposes, the following definition suffices. An \(f(n)\)-separator of a graph on \(n\) vertices recursively splits its nodes into two parts having sizes \([n/2]\) and \([n/2]\), with at most \(f(n)\) edges between them. Roughly speaking, if the nodes of a graph with an \(f(n)\)-separator are colored arbitrarily with up to \(k\) colors, where
If \( f(n) > cn \), then the discrete bisection result implies that the graph has an \( O(kf(n)) \)-separator that splits each of the colors in half when it separates the vertices. This uses a restatement of Theorem 1 for the problem of bisecting "opened necklaces." Suppose beads of \( k \) colors are arranged on a line segment. Then intervals for a bisection can be obtained by making at most \( k \) cuts. In this version it is not necessary to distinguish between odd and even \( k \), since the endpoint of the segment, which corresponds to the opening of the necklace, serves as the arbitrary prescribed cut if \( k \) is odd. Unfortunately, no proof for this cleaner formulation is known that does not use the circle coloring result.

2. Discrete from continuous. In this section we show that Theorem 2 implies Theorem 1. We begin by noting that some arrangements using \( k \) colors require the full \( \lfloor k/2 \rfloor \) intervals. (This example also provides the lower bound for the continuous problem.) Arrange the beads so that the beads of each color appear contiguously. Since no bisection can include all the beads of a color or exclude them all, there must be at least one "cut" (a switch between inclusion and exclusion) among the beads of each color. With at least \( k \) switches between beads chosen and beads omitted, the beads chosen must be separated into at least \( \lfloor k/2 \rfloor \) intervals.

Now, suppose that Theorem 2 holds. We can turn a necklace of \( 2n \) beads into a circle coloring by partitioning the circle into \( 2n \) equal segments ("units") and coloring them with \( k \) pure colors corresponding to the order of beads on the necklace. Theorem 2 guarantees a bisection with at most \( \lfloor k/2 \rfloor \) intervals, but the endpoints of the intervals (the cuts) need not occur at endpoints of the \( 2n \) units. Using induction on the number of these bad cuts, this can be corrected to obtain a bisection for the discrete problem. If there are no bad cuts, we are finished. Otherwise, suppose there is a cut inside a unit of color \( i \), which is used on \( 2a_i \) units around the circle. Since the continuous bisection captures altogether an integral amount \( a_i \) of color \( i \), there must be another bad cut in a unit of color \( i \). At any cut within color \( i \), the interval on one side contributes to the amount if color \( i \) captured, and the interval on the other side does not. Move the two bad cuts, by the same amount, so as to shrink the interval contributing to color \( i \) at one of them and expand the interval contributing to color \( i \) at the other, until one of the cuts reaches the endpoint of a unit or another bad cut. This produces a bisection with fewer bad cuts.

The same argument can be used to treat necklaces where the number of beads of a color need not be even. Apply the continuous result and transform the resulting bisection as above. Cuts will occur only at endpoints of units except that a color contributing an odd number of beads will have one unit with a cut at its midpoint. That cut can be moved a half unit in either direction to obtain the following result, where we broaden the definition of discrete bisection to mean a choice of beads capturing \( \lfloor 2a_i/2 \rfloor \) or \( \lfloor 2a_i/2 \rfloor \) beads of color \( i \).

**Theorem 3.** Every necklace with \( k \) colors of beads has a discrete bisection using no more than \( \lfloor k/2 \rfloor \) intervals. No smaller size suffices for all arrangements of \( k \) colors. If \( k \) is odd, one cut can be specified arbitrarily. Any pattern of \( \lfloor 2a_i/2 \rfloor \)'s and \( \lfloor 2a_i/2 \rfloor \)'s can be specified for the bead colors used an odd number of times.

3. The parameter space. In this section we specify the parameters used to describe a choice of intervals and obtain properties of the parameter space that will be important in the proof. For ease of discussion, we refer to any choice of intervals that together capture half the total length as a **snare**; if it consists of at most \( j \) disjoint intervals, we call it a \( j \)-**snare**. In the discrete case, length is measured by number of beads. In the continuous case, we measure length by integrating the densities, with the total amount...
of color $i$ being $2a_i$. In either case, the total length is $2 \sum a_i$; let $L = \sum a_i$. A snare captures the points contained in its intervals.

Given a coloring, specify an arbitrary point $P$ on the circle as a reference point. In the discrete case, $P$ must lie between two units. The intervals are determined by the cuts made between them, so we could specify the intervals by measuring the distances from $P$ to those cuts. However, it will be more convenient to organize the parameters into simplices.

Henceforth let $m = \lceil k/2 \rceil$, and restrict attention to $m$-snares. Since the complement of any bisection with $m$ intervals is also a bisection with $m$ intervals, the parameter space need only describe at least one $m$-snare from every complementary pair. In fact, our parameter space $B_k$ contains points describing all $m$-snares for which $P$ is not an interior point of a (captured) interval. Given this, the parameters describing such a snare are easily computed. Moving in a counterclockwise direction, let $y_0$ be the distance from $P$ to the first beginning of an interval in the snare. Continuing counterclockwise, let $x_i$ be the length of the $i$th interval in the snare, and let $y_i$ be the length of the gap between the $i$th and $(i+1)$th intervals, with $y_m$ the length between the $m$th and $P$. If the snare has less than $m$ intervals, this description still works, by setting leftover parameters to 0. Note that $\sum x_i + \sum y_i = L$.

As noted above, a bisection of size $m$ exists if and only if there is one where $P$ is not an interior point of an interval. Thus, the parameter space in which we look for bisections can be described as $B_k = X \times Y$, where $X$ and $Y$ are simplices whose coordinates sum to $L$, representing the choices for $\bar{x}$ and $\bar{y}$. However, for odd $k$ we claim that one cut can be made arbitrarily. We choose this to be the reference point $P$. By setting $y_0 = 0$, we obtain all snares in which the first interval starts at $P$. The snares in which the last interval ends at $P$ have $y_m = 0$ and possibly $y_0 > 0$; we ignore these since they are complements of those with $y_0 = 0$. Thus, when $k$ is odd we drop $y_0$, so that $X$ and $Y$ are both $m$-parameter, $(m-1)$-dimensional simplices. If $k$ is even, $X$ is $(m-1)$-dimensional and $Y$ is $m$-dimensional. In either case, $B_k$ is $(k-1)$-dimensional; it has $k + 1$ parameters, restricted by the two sum relations $\sum x_i + \sum y_i = L$.

Let $A_k$ be a $(k-1)$-dimensional hyperplane of $k$-tuples whose coordinates sum to $L$. Let $f : B_k \rightarrow A_k$ be the function whose value on a particular snare gives the amount of each color captured. In the discrete case, this counts the captured beads; in the continuous case, it integrates the density functions over the intervals of the snare. Given a coloring with total amount $2a_i$ of color $i$, the point of interest is $(a_1, \ldots, a_k)$; we want to insure that $\bar{a}$ is in the image of $f$. We say that two points $\bar{b}, \bar{b}'$ are antipodal in $A_k$ if $\bar{b} = 2\bar{a} - \bar{b}'$. Expressing this as $\bar{b} - \bar{a} = \bar{a} - \bar{b}'$, note that for antipodal points any excess for $\bar{b}$ in a color becomes a deficiency of the same size for $\bar{b}'$ in that color. In particular, points of $B_k$ that describe complementary snares map to antipodal points of $A_k$ under $f$.

For purposes of the proof, consider now the continuous case; we will return to the discrete case to discuss algorithms. Geometrically, a direct way to show that $\bar{a}$ is hit by $f$ would be to study the image of the boundary of $B_k$, and attempt to use the Borsuk–Ulam theorem regarding antipodal maps. Intuitively, one wants to show that $f(\partial B_k)$ “surrounds” $\bar{a}$, so that $\bar{a}$ must lie “inside” $f(\partial B_k)$, in which case continuity implies $\bar{a} \in f(\partial B_k)$.

We claim the points of $f(\partial B_k)$ come in antipodal pairs, with each pair the image of points in $\partial B_k$ describing complementary snares. If all the points of $\partial B_k$ could be paired up this way, then the Borsuk–Ulam theorem could be applied to get the desired result. Unfortunately, not all of $\partial B_k$ participates, so this theorem cannot be used directly. This problem arises because $(m-1)$-snares have many descriptions in the
Bisection of Circle Colorings 97

parameter space. Another map could be introduced to "collapse" the poorly-behaved part of the parameter space and then apply the Borsuk-Ulam theorem; this approach works but leads to very tedious and technical arguments, so we omit it. The inductive proof presented in the next section circumvents the difficulty, using arguments like those used to prove the Borsuk-Ulam theorem. Before embarking on that, we pause to obtain properties of the parameter space and the map $f$ that will be useful in the proof.

We need to study the boundary of $B_k$. The boundary points are those where at least one of the parameters in the simplices $X$ or $Y$ is 0. We will break the boundary into four pieces, a "front face" $B^+$, a "back face" $B^-$, and two other parts $C$ and $D$, depending on which of the parameters reaches an extreme value. In general, if $x_i = 0$, one of the intervals in the corresponding snare vanishes and two of the gaps merge. Similarly, if $y_i = 0$ for $1 \leq i < m$, one of the gaps vanishes and two of the intervals merge. In either case, the snare actually uses at most $m-1$ intervals. Conversely, describing a snare with less than $m$ intervals requires one of these parameters to be 0; all $(m-1)$-snares are described by points in $\partial B_k$. For even values of $k$, the parts $C$ and $D$ will contain the points that describe $(m-1)$-snares; for odd $k$ they will describe $(m-1)$-snares with a restricted endpoint, at $P$. $B^+$ and $B^-$ allow more freedom; for even $k$ their points describe $m$-snares with a restricted endpoint, and for odd $k$ they describe all $(m-1)$-snares.

Since $C$ and $D$ describe snares with less freedom, it comes as no surprise that their images have smaller dimension than the rest of the boundary. In fact, we will show that they do not contribute anything to image of the boundary; i.e. $f(\partial B_k) = f(B^+) \cup f(B^-)$. This justifies the terms "front face" and "back face". In relation to the preceding topological motivation, it is important that we can in this sense throw away $C$ and $D$, because these are the parts of $\partial B_k$ that do not participate in the pairing of points with antipodal images.

To define $B^+$, $B^-$, $C$, $D$, we consider two cases. First suppose $k$ is even, so that $Y$ has points $\tilde{y} = y_0, \ldots, y_m$. Let $B^+ = \{(\tilde{x}, \tilde{y}); y_0 = 0\}$, $B^- = \{(\tilde{x}, \tilde{y}); y_m = 0\}$, $C = \partial X \times Y$, and $D = X \times \partial Y - (B^+ \cup B^-)$. Since $B^+$ and $B^-$ consist of the points of $B_k$ with $y_0 = 0$ or $y_m = 0$, they describe all $m$-snares in which the first or last interval (which may be empty if $x_1 = 0$ or $x_m = 0$) begins or ends at $P$. $C$ and $D$ describe $(m-1)$-snares. Note that $\partial B_k = B^+ \cup B^- \cup C \cup D$.

Next, suppose that $k$ is odd, so that $Y = \{y_1, \ldots, y_m\}$. Let $B^+ = \{(\tilde{x}, \tilde{y}); x_1 = 0\}$, $B^- = \{(\tilde{x}, \tilde{y}); y_m = 0\}$, $C = (\partial X \times Y) - B^+$, and $D = (X \times \partial Y) - B^-$. This time $B^+$ and $B^-$ describe $(m-1)$-snares that do not or do capture $P$. $C$ and $D$ describe $(m-1)$-snares where the first or last interval (which may be empty if $x_1 = 0$ or $x_m = 0$) begins or ends at $P$. Again, $\partial B_k = B^+ \cup B^- \cup C \cup D$.

**Lemma 1.** Let $f : B_k \rightarrow A_k$ be a circle coloring, and assume that $\dim f(S) \leq \dim S$ for any $S \subset B_k$. Then the pieces $B^+$, $B^-$, $C$, $D$ of the boundary of $B_k$ satisfy the following properties.

(a) $B^+$ is naturally isomorphic to $B_{k-1}$.

(b) The points of $B^+$ and $B^-$ come in pairs with antipodal images.

(c) $f(\partial B_k) = f(B^+) \cup f(B^-)$, and in fact $f(C) \cup f(D) \subset f(B^+) \cup f(B^-)$.

(d) $\dim (f(C) \cup f(D)) \leq \dim A_k - 2$.

**Proof.** We consider the even and odd cases separately. Keep in mind that when $k$ is odd $\tilde{x}$ and $\tilde{y}$ have the same number of components, but when $k$ is even $\tilde{y}$ has one more component than $\tilde{x}$.

Assume $k$ is even. (a) is trivial, since deleting the fixed value $y_0 = 0$ makes $B^+$ precisely the parameter space $B_{k-1}$ for a problem with one less color. $B^-$ is a translate
of $B_{k-1}$. For (b), note that $f(\bar{x}; y_0, \cdots, y_{m-1}, 0) = 2\bar{a} - f(y_0, \cdots, y_{m-1}; 0, \bar{x})$, since these points from $B^-$ and $B^+$ describe complementary snares.

(c) and (d) rest on the fact that an $(m-1)$-snares has many descriptions in $B_k$; any one of the gaps or intervals could be the one that collapses. In particular, it has a description in both $B^+$ and $B^-$. For the image of a point in $C$ with $x_i = 0$,

$f(x_1, \cdots, x_{i-1}, 0, x_{i+1}, \cdots, x_m; \bar{y})$

$= f(0, x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_m; 0, y_0, \cdots, y_{i-2}, y_{i-1} + y_i, y_{i+1}, \cdots, y_m) \in f(B^+)$

$= f(x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_m, 0; y_0, y_1, \cdots, y_{i-2}, y_{i-1}, \cdots, y_{m-1}, 0) \in f(B^-)$

because all three parameter sets describe the same snare. Similarly, for the image of a point in $D$ with $y_i = 0$ for $1 \leq i < m$,

$f(\bar{x}; y_0, \cdots, y_{i-1}, 0, y_{i+1}, \cdots, y_m)$

$= f(0, x_1, \cdots, x_{i-1}, x_i + x_{i+1}, x_{i+2}, \cdots, x_m; 0, y_0, \cdots, y_{i-1}, y_{i+1}, \cdots, y_m) \in f(B^+)$

$= f(x_1, \cdots, x_{i-1}, x_i + x_{i+1}, x_{i+2}, \cdots, x_m, 0; y_0, y_1, \cdots, y_{i-1}, y_{i+1}, \cdots, y_{m-1}, 0) \in f(B^-)$.

Since we have given descriptions from both $f(B^+)$ and $f(B^-)$, this proves (c). Moreover, the description from $f(B^+)$ for each point of $f(C) \cup f(D)$ shows that the dimension of $f(C) \cup f(D)$ is at most $2m - 3 = k - 3$, because there are $2m - 1$ nonfixed parameters and two sum relations $\sum x_j = L = \sum y_i$, and the hypothesis of the lemma states that $f$ does not raise dimension. Since $\dim A_k = k - 1$, we have (d).

Now suppose $k$ is odd. (a) follows from the fact that $(0, x_2, \cdots, x_m; \bar{y}) \in B^+$ and $(x_2, \cdots, x_m; \bar{y}) \in B_{k-1}$ describe the same snare. For (b), $(0, x_2, \cdots, x_m; \bar{y}) \in B^+$ and $(\bar{y}; x_2, \cdots, x_m, 0) \in B^-$ describe complementary snares, whose images under $f$ are antipodal. To show the other claims, we proceed as in the even case. If some $x_i = 0$ with $i > 1$,

$f(x_1, \cdots, x_{i-1}, 0, x_{i+1}, \cdots, x_m; \bar{y})$

$= f(0, x_1, \cdots, x_{i-1}, x_i + x_{i+1}, x_{i+2}, \cdots, x_m; 0, y_1, \cdots, y_{i-2}, y_{i-1} + y_i, y_{i+1}, \cdots, y_m) \in f(B^+)$

$= f(x_1, \cdots, x_{i-1}, x_i + x_{i+1}, x_{i+2}, \cdots, x_m, 0; y_1, y_{i-1}, \cdots, y_i, y_{i+1}, \cdots, y_m, 0) \in f(B^-)$.

Similarly, if some $y_i = 0$ with $i < m$,

$f(\bar{x}; y_1, \cdots, y_{i-1}, 0, y_{i+1}, \cdots, y_m)$

$= f(0, x_1, \cdots, x_i, x_{i+1}, x_{i+2}, \cdots, x_m; 0, y_1, \cdots, y_{i-1}, y_{i+1}, \cdots, y_m) \in f(B^+)$

$= f(x_1, \cdots, x_i, x_{i+1}, x_{i+2}, \cdots, x_m, 0; y_1, y_{i-1}, \cdots, y_i, y_{i+1}, \cdots, y_m, 0) \in f(B^-)$.

This establishes (c). To show (d), note from the $f(B^+)$ description that in $f(B^+)$ \( \cap f(B^-) \) there are $2m - 2$ nonfixed parameters. Together with the same two sum relations, this yields $\dim f(C) \cup f(D) \leq 2m - 4 = k - 3 = \dim A_k - 2$. 

4. \( \lfloor k/2 \rfloor \) intervals suffice. In this section, we use induction on $k$ to show that an arbitrary circle coloring with $k$ colors and continuously integrable color densities has a bisection with at most \( \lfloor k/2 \rfloor \) intervals. We consider only bisections described by points in the parameter space $B_k$. In particular, none of the $m$ (possibly empty) specified intervals has the reference point $P$ in its interior, and $P$ is the starting point of an interval if $k$ is odd. The proof uses concepts and well-known results from topology. Due to the origin of the problem and the application of the results, we expect many of the readers of this paper to be discrete mathematicians. Therefore, we will try to
describe the relevant topological concepts in footnotes for interested readers; missing definitions or results can be found in [2] or [3].

If the boundary of $B_k$ contains a solution, there is nothing to prove, so assume $\partial f(\partial B_k)$. We would like to draw conclusions about the number of points in $f^{-1}(\bar{a})$, in particular that this number is odd and therefore nonzero. Unfortunately, $|f^{-1}(\bar{a})|$ does not behave well enough for arbitrary continuous functions $f$ (for example, it may become infinite), so instead we study a quantity that agrees with $|f^{-1}(\bar{a})|$ whenever $f$ is well-behaved (a local homeomorphism) at each point of $f^{-1}(\bar{a})$. This quantity is the degree of the map $f : (B_k, \partial B_k) \rightarrow (A_k, A_k - \bar{a})$, henceforth denoted $\deg(f)$. The degree of a continuous map from a set and boundary into another set with a point removed describes how many times the image covers the target point. In particular, if $f^{-1}(\bar{a})$ is empty, then $\deg(f) = 0$. Since $\deg(f)$ is a construct using homology, $\deg(f)$ does not change under continuous perturbations ("homotopies") of $f$, a fact that will be used when $f$ is poorly behaved. Later, we will relate $\deg(f)$ to the winding number of $f(B_k)$ around the target point, since that is the quantity to which we can apply the antipodal facts we derived about $\partial B_k$.

Since we seek only bisections corresponding to points in $B_k$, we will refer to such points as "solutions," and we will denote the problem of finding such solutions by the corresponding color function $f$. We will prove Theorem 2 by showing that $\deg(f)$ is odd (i.e., nonzero) when the problem has no solutions in $\partial B_k$. Theorem 2 follows readily; if a $k$-color problem has no solutions in $B_k$, then it has none in $\partial B_k$ and hence $\deg(f)$ is nonzero. This contradicts the fact that $\deg(f) = 0$ if $f^{-1}(\bar{a})$ is empty. We have reduced the theorem to the following topological lemma.

**Lemma 2.** Let $f : B_k \rightarrow A_k$ be a coloring bisection problem. If $f$ has no solutions in $\partial B_k$, then $\deg(f)$ is odd.

**Proof.** We prove this by induction on $k$. For $k = 1$, the parameter space $B_1$ consists of the single point $(L, L)$. Since the sum of the color densities must be 1 at each point, one color paints the circle uniformly. The snare corresponding to $(L, L)$ is the semicircle starting at $P$, so indeed it bisects the coloring. Recall that $\deg(f)$ is the degree of the map $f : (B_1, \partial B_1) \rightarrow (A_1, A_1 - \bar{a})$; in this case the second element of each pair is empty, and the first is a 1-point set. Any mapping of an $n$-point set to a 1-point set covers the target point $n$ times and has degree $n$, so here $\deg(f) = 1$.

For the induction step, we must show the following: given a $k$-color problem $f$ with no solutions in $\partial B_k$, there exists a $(k-1)$-color problem $f'$ with no solutions in $\partial B_{k-1}$ and with $\deg(f) \equiv \deg(f')(\mod 2)$.

We approach the computation of $\deg(f)$ by looking at the winding number of the boundary around the target point, because it turns out that this quantity equals

---

1 The precise definition of degree uses homology groups. Given a set $C$ and a specified subset $D$, these groups are denoted $H_i(C, D)$ and called "the $i$th homology group of $C$ modulo $D$." The elements of $H_i(C, D)$ are classes of formal sums of continuous functions mapping an $i$-dimensional ball or simplicial complex $\Delta'$ into $C$, with the requirement that the boundary of $\Delta'$ must map into $D$. Two of these formal sums belong to the same homology class if they can be continuously deformed into one another (i.e., are "homotopic"). For the domain and range of the color function $f$, the $(k-1)$st homology groups $H_{k-1}(B_k, \partial B_k)$ and $H_{k-1}(A_k, A_k - \bar{a})$ are infinite cyclic, i.e., isomorphic to $\mathbb{Z}$. The generator $g$ of $H_{k-1}(B_k, \partial B_k)$ is a standard homeomorphism of a $(k-1)$-simplex into $B_k$, since $B_k$ is $(k-1)$-dimensional. The generator $g'$ of $H_{k-1}(A_k, A_k - \bar{a})$ maps the $(k-1)$-simplex into a ball centered at $\bar{a}$; this covers $\bar{a}$ exactly once. The map $f$ induces a homomorphism $f_*$ between these groups. The degree of $f$ is defined to be the value of $d$ such that $f_* (g) = d \cdot g'$, where $f_*$ is the induced homomorphism on the top-dimensional homology group. In other words, $f$ maps $B_k$ into something that covers the target point $\deg(f)$ times. For example, if the entire image lies in the punctured set $A_k - \bar{a}$, so that $f^{-1}(\bar{a}) = \emptyset$, then $f_*(g)$ is in the homology class that is the 0 multiple of the generator, and $\deg(f) = 0$. 

---
deg (f). Let \( \partial f \) denote the restriction of \( f \) to \( \partial B_k \). Under the assumption that the boundary does not hit the target point, we have \( \partial f : B_k \to A_k - \hat{a} \). The **winding number** \( W(f; D, \hat{a}) \) is defined to be the degree of the map \( f|D : D \to A_k - \hat{a} \) (technically, the degree of the map \( f|D : (D, \emptyset) \to (A_k - \hat{a}, \emptyset) \)); we will compute \( W(\partial f, \hat{a}) \). An elegant result in topology states that \( W(\partial f, \hat{a}) = \deg (f) \), using the fact that \( H_i(C, D) \) and \( H_{i-1}(D, \emptyset) \) are isomorphic when \( C \) is contractible.

The winding number can be computed by following any fixed reference direction \( l^* \) from the target point and counting crossings with the image surface, a fact that is intuitively clear in two dimensions. Of course, the crossings must be counted with appropriate sign, depending on whether the ray enters or leaves the region of interest. Fixing the ray \( l^* \) from \( \hat{a} \) and counting the crossings with the proper sign yields the **intersection number** \( I(\partial f, l^*) \). Being a homology concept, \( I(\partial f, l^*) \) is invariant under continuous deformations of \( f \) or small changes in \( l^* \). The relevant topological fact is that \( W(\partial f, \hat{a}) = I(\partial f, l^*) \).

2 The winding number is a useful and sensible construct only when \( H_{d-2}(D, \emptyset) \) is infinite cyclic (isomorphic to \( \mathbb{Z} \)), which holds here for \( D = \partial B_k \). To compute the winding number, we need the induced homomorphism \( f_* \) between the \((k-2)\)nd homology groups, since the dimension of \( \partial B_k \) is \( k-2 \). \( H_{k-2}(\partial B_k, \emptyset) \) and \( H_{k-2}(A_k - \hat{a}, \emptyset) \) are both infinite cyclic. When the second argument to \( H_i(C, D) \) is the null set, the formal sums in each homology class must be such that all the image contributions arising from the boundary of the \( i \)-dimensional domain cancel. In other words, the generator \( g \) of \( H_{k-2}(\partial B_k, \emptyset) \) covers \( \partial B_k \) exactly once with \((k-2)\)-dimensional simplices whose boundaries cancel out. Similarly, the generator \( g' \) of \( H_{k-2}(A_k - \hat{a}, \emptyset) \) covers a sphere centered at \( \hat{a} \) exactly once. \( W(\partial f, \hat{a}) \) is the value of \( d \) such that \( f_*(g) = d \cdot g' \).

3 This result can be explained as follows. Given a set \( C \) and subset \( D \), there is a natural sequence of homology groups

\[
H_i(C, \emptyset) \to H_i(C, D) \to H_{i-1}(D, \emptyset) \to H_{i-1}(C, \emptyset),
\]

in which the arrows represent natural group homomorphisms. The outer homomorphisms are inclusion mappings. The central one is the “boundary mapping.” (To define the boundary mapping, consider a map \( \sigma : \Delta^i \to C \) belonging to one of the homology classes in \( H_i(C, D) \). Its restriction \( \partial \sigma \) maps \( \partial \Delta^i \) to \( D \). Since \( \partial \Delta^i \) is a union (formal sum) of \((i-1)\)-simplices, define \( \sigma_j \) by restricting \( \sigma \) to the \( j \)th simplex in \( \partial \Delta^i \). As a formal sum, it follows that \( \partial \sum \sigma_j = 0 \) (since \( \partial^2 = 0 \) by cancellation), so indeed \( \sum \sigma_j \) belongs to an element of \( H_{i-1}(D, \emptyset) \).) This sequence of homomorphisms is an exact sequence, in the sense of group homomorphisms (the image of the previous map is the kernel of the next). However, if \( C \) is contractible, i.e. has no holes, then the groups at the ends of the sequence are the trivial group; the only formal sum of continuous maps that can take the boundary of \( \Delta^i \) to the empty set is the 0 formal sum. Since the sequence is exact, the two groups in the middle are isomorphic. Applying this to \( (B_k, \partial B_k) \) and \( (A_k - \hat{a}, \emptyset) \), we find that \( H_{k-2}(\partial B_k, \emptyset) \) and \( H_{k-2}(A_k - \hat{a}, \emptyset) \) are also infinite cyclic. Considering the induced homomorphisms \( f_* : H_i(B_k, \partial B_k) \to H_i(A_k, A_k - \hat{a}) \) and \( f'_* : H_{i-1}(\partial B_k, \emptyset) \to H_{i-1}(A_k - \hat{a}, \emptyset) \) yields a commutative diagram, with two ways to get from \( H_i(B_k, \partial B_k) \) to \( H_{i-1}(A_k - \hat{a}, \emptyset) \). The diagram is commutative because, for a mapping \( \Delta^i \to B_k \) that is restricted to \( \partial \Delta^i \) and composed with \( f \), it doesn’t matter whether the restriction or the composition happens first. If \( g \) and \( g' \) are the generators, then following \( f_* \) by the \( A \)-isomorphism takes \( g \) to \( d \cdot g' \), where \( d = \deg (f) \). Following the \( B \)-isomorphism by \( f'_* \) takes \( g \) to \( w \cdot g' \), where \( w = W(\partial f, \hat{a}) \). Since the two routes yield the same result, \( \deg (f) = W(\partial f, \hat{a}) \).

4 This fact has a simpler explanation than that of the previous footnote. The intersection number is defined for any oriented \( i \)-dimensional surface intersecting an oriented \( (d - i) \)-dimensional surface in \( d \)-space. Given orientations for each, i.e. coordinate systems, the combined coordinate system at an intersection has a well-defined sign, corresponding to “left-handed” or “right-handed.” In particular, \( l^* \) is an oriented \( 1 \)-dimensional surface and \( f(\partial B_k) \) is an oriented \((k-2)\)-dimensional surface in \((k-1)\)-space, which means that \( f(\partial B_k) \) has a well-defined “top” and “bottom” (given that \( f \) does not raise dimension). As you travel \( l^* \) away from \( \hat{a} \), if you cross \( f(\partial B_k) \) from top to bottom count \(-1 \), and from bottom to top count \(+1 \). Tangency causes no problem, because you can count \( \pm \frac{1}{2} \) when you enter and \( \pm \frac{1}{2} \) when you leave, in the appropriate way. Now consider the definition of \( W(\partial f, \hat{a}) \) using degree and the description of the generator in \( H_{k-2}(A_k - \hat{a}, \emptyset) \). Note that \( W(\partial f, \hat{a}) \) is the number of times \( f(\partial B_k) \) covers each point when radially projected onto a reference sphere centered at \( \hat{a} \) (counted with appropriate sign in case of folds). This is precisely \( I(\partial f, l^*) \) for any \( l^* \).
$f(\partial B^-)$ has dimension at most $k-2$, by Lemma 1(d). Thus, almost every direction $l^*$ misses $f(C) \cup f(D) \cup f(\partial B^+) \cup f(\partial B^-)$, in which case $(\partial f)^-(l^*)$ consists only of points in $\text{int}(B^+) \cup \text{int}(B^-)$. If $x \in \text{int}(B^-)$ and $f(x) \in l^*$, then for the “antipodal point” $x^* \in B^-$ under the pairing established in Lemma 1(b), $f(x^*) = 2\bar{a} - f(x)$ belongs to $l^-$, the opposite ray from $\bar{a}$. Conversely, if $x \in \text{int}(B^+)$ and $f(x) \in l^*$, then $x^* \in B^-$ and $f(x^*) \in l^*$. Consequently, $I(f|B^+, l^*) = I(f|B^+, l^-)$. Let $l = l^+ \cup l^-$ be the full line through $\bar{a}$. Note that orienting $l$ requires us to change the orientation on $l^-$. We have proved

$$W(\partial f, \bar{a}) = I(\partial f, l^*) = I(f|B^+, l^*) + I(f|B^-, l^*) = I(f|B^+, l^*) + I(f|B^+, l^-)$$

$$= I(f|B^+, l^*) - I(f|B^+, l^-) = I(f|B^+, l) \quad \text{(congruence mod 2)}.$$  

All that remains is to obtain the desired $(k-1)$-colored problem. We do this by projecting along an appropriate line $l$ into a space with fewer colors. Let $A_{k-1} \subset A_k$ be those $k$-tuples whose last coordinate is 0. Intuitively, the most satisfying projection to use is the one that amalgamates the last two colors, although any direction will work as long as it is not parallel to $A_{k-1}$, does not hit $f(C) \cup f(D)$, and does not hit $f(\partial B^+)$. Within a hyperplane, such as $A_k$, a direction is specified by a $k$-tuple whose coordinates sum to 0. Linear projection along a direction simply adds a multiple of that $k$-tuple. Projection along the direction $(0, \ldots, 0, -1, 1)$ maps $b \in A_k$ to $(b_1, \ldots, b_{k-2}, b_{k-1} + b_k, 0) \in A_{k-1}$. If $f$ does not raise dimensions, there is a direction arbitrarily close to this that avoids $f(C) \cup f(D) \cup f(\partial B^+) \cup f(\partial B^-)$, and to which we can apply the chain of equalities and congruences we have built under that assumption.

Let $\pi$ be a projection along such a direction. If $\tilde{\rho}(x) = (\rho_1(x), \ldots, \rho_k(x))$ are the color densities at a point $x$ on the circle, then $\pi(\tilde{\rho}(x))$ are also continuously integrable color densities summing to 1, and this may be considered a $(k-1)$-color problem since the last coordinate is always 0. Note that the projection may make some color density negative, but that is allowed in the class of coloring problems we originally defined. To compute the color function $f'$ for the new problem, note that the color function is a sum of $m$ integrals of the color densities. Summation and integration commute with linear projection, so $f' = \pi \cdot f$ is the $(k-1)$-color problem with densities $\pi \cdot \tilde{\rho}(x): f$ counts the original colors, and $\pi$ redistributes them.

The parameter space for $f'$ is simply $B^+$, which by Lemma 1(a) is isomorphic to $B_{k-1}$. The target point in $f'$ is $\pi(\bar{a})$. In fact, the entire line $l$ maps under $\pi$ to $\pi(\bar{a})$. This leads us back to $\text{deg}(f')$, using the fact that the projection of $f$ crosses the projection of $l$ whenever $f$ crosses $l$, i.e. $I(f|B^+, l) = I(\pi f|B^+, \pi(l))$. Since we chose a direction that did not intersect $f(\partial B^+) \cup f(\partial B^-)$, the projected problem satisfies the hypothesis of the lemma; it has no solutions in the boundary of its parameter space. By induction, $\text{deg}(f')$ is odd). But now $\text{deg}(f)$ is also odd, since

$$\text{deg}(f) = W(\partial f, \bar{a}) = I(\partial f, l^*) = I(f|B^+, l)$$

$$= I(\pi f|B^+, \pi(l)) = \text{deg}(\pi f|B^+) = \text{deg}(f').$$  

The only remaining detail is the assumption we have made throughout this argument that $f$ does not increase dimension. As the Peano space-filling curves show, continuous functions can raise dimension. However, this lemma still holds for such functions, because we can apply it to a suitable simplicial approximation $\tilde{f}$, and simplicial functions
never raise dimension. The simplicial function is homotopic to \( f \), which means that it has the same degree, so \( \deg(\hat{f}) \) odd implies \( \deg(f) \) is also odd.

5. Bisection algorithm. Let us now return to the discrete case to discuss algorithms for finding bisections. When the number of colors in the necklace is fixed at \( k \), the guarantee that a bisection with at most \( m = \lfloor k/2 \rfloor \) intervals exists yields a straightforward algorithm to find the smallest bisection. This simple algorithm uses exhaustive search. An \( m \)-snare is determined completely by the location of the \( k \) “cuts” between captured and noncaptured beads, i.e. the endpoints of the intervals. (Note that if \( k \) is odd we have placed an arbitrary cut in advance, so \( k \) more determine the snare.) However, the \( k \) cuts cannot be placed with complete freedom, since the total number of beads included and excluded must both equal \( n \). The simplest way to implement this restriction is to search the entire parameter space \( B_k \). The snares represented many times are those with fewer intervals, and there are fewer of these. In other words, the number of points in \( B_k \) has the same order as the number of snares.

The number of integer points in a \( d \)-dimensional simplex whose \( d+1 \) variables sum to \( n \) is \( \binom{n+d}{d} \). Thus \( X \) has \( \binom{n+m}{m-1} \) integer points, and \( Y \) has that many or \( \binom{n+m}{m} \), depending on the parity of \( k \). In either case, the number of integer points in \( B_k \) is the product of polynomials in \( n \) of degrees \( \lfloor k/2 \rfloor - 1 \) and \( \lceil k/2 \rceil \), so it is \( O(n^{\lfloor k/2 \rfloor - 1}) \).

After \( O(n) \) operations for preprocessing, the image of any snare can be computed in constant time (i.e., \( O(k) \)), using the following idea due to Leiserson [7]. In one pass through the necklace, compute the cumulative distributions for the colors. Then, to compute the color amounts captured by a snare, sum the differences of these distributions between the endpoints of the \( m \) intervals in the snare. In the boundary, of course, we need only compute the images of the points in \( B^+ \). The search finds a bisection from every complementary pair of bisections with at most \( \lfloor k/2 \rfloor \) intervals (and with the specified cut if \( k \) is odd). It runs in \( O(n^{\lfloor k/2 \rfloor - 1}) \) time.

If \( k \geq 3 \), we can save one factor of \( n \) by using the topological ideas in the proof that \( \lceil k/2 \rceil \) intervals suffice. A “divide-and-conquer” search of the parameter space \( B_k \) runs in \( O(n^{\lfloor k/2 \rfloor - 2}) \) time. As with binary search, to which this reduces when \( k = 2 \), we split the parameter space into pieces, determine the piece in which to search for the desired point, and recurse. With preprocessing as above, the computation of the color function for any point in \( B_k \) takes \( O(k) \) time. Other operations required will also be polynomial in \( k \) but independent of \( n \), so the asymptotic running time of the algorithm will be determined by the number of function evaluations required.

“Divide-and-conquer” yields an \( O(n^{\lfloor k/2 \rfloor - 2}) \) algorithm to find a solution in \( B_k \), but it does not find all solutions. To ensure finding the smallest solution, we must test all snares with less than \( m \) intervals individually. As noted above, testing the points in \( B^+ \) suffices, but \( B^+ \) is a copy of \( B_{k-1} \), which has \( O(n^{\lfloor k/2 \rfloor - 2}) \) integer points, so testing all these points is cheap enough. Having checked \( \partial B_k \), we can assume there is no solution in the boundary.

To define a concept of winding number in this discrete situation, we use a simplicial map, because there is an easy way to compute intersection numbers for such maps and hence determine the winding number. At the integer points in \( B_k \), define \( f \) by the

\[ f(x) = \begin{cases} 
1 & \text{if } x \in C \\
-1 & \text{if } x \in N \\
0 & \text{otherwise}
\end{cases} \]

Given a color function \( f : B_k \to A_k \), the simplicial approximation we need must satisfy the antipodal properties obtained in Lemma 1 for \( B^+ \) and \( B^- \). To do this, merely choose the vertices for the simplicial decomposition of \( \partial B_k \) in antipodal pairs according to the pairing in Lemma 1(b) for \( f \). Let \( \hat{f} \) agree with \( f \) at these vertices, and let the simplicial decomposition respect the pairing. Define \( \hat{f} \) on the rest of \( B_k \) by linear interpolation from the vertices. This simplicial map has the antipodal property, and the argument succeeds.
discrete bead-counting color function. The natural unit regions in the grid formed by the integer points are not simplices. For \( k = 3 \) they are squares, and in general they are the product of \( ([k/2] - 1) \)- and \( [k/2] \)-dimensional simplices. Partition these further into simplices to view \( B_k \) as a simplicial complex of dimension \( k - 1 \), and extend \( f \) to the entire space by linear interpolation on these simplices. For example, when \( k = 3 \) the product of two segments partitions naturally into two triangles, and when \( k = 4 \) the product of a triangle with a segment partitions naturally into three tetrahedrons.

The computation of winding number uses the value of \( f \) at the integer boundary points, the extension of \( f \) by linear interpolation, and the idea that winding number equals intersection number. Let \( D \) be a simplicial subcomplex of \( B_k \), and suppose that \( f(\partial D) \) does not include the target point \( \bar{a} \). Then we want to compute \( W(f|_{\partial D}, \bar{a}) \) by counting the signed intersections of \( \partial D \) with some ray emanating from \( \bar{a} \). The boundary \( \partial D \) consists of simplices of dimension \( k - 2 \), since \( B_k \) itself has dimension \( k - 1 \). Since \( f \) is simplicial, the image of a boundary simplex \( \Delta \) is the simplex of dimension at most \( k - 2 \) whose vertices are images of the vertices of \( \Delta \). So, the problem of computing contributions to the intersection number reduces to deciding when a specified ray crosses a simplex determined by specified points, and in which "direction."

Let \( l^+ \) be an arbitrary direction vector in \( A_k \), let \( b_1, \ldots, b_{k-1} \) be the vertices of a simplex \( \Delta \) in \( \partial D \), and let \( c_i = f(b_i) - \bar{a} \). Then \( f(\Delta) \) intersects the ray \( \bar{a} + tl^+ \) \((t > 0)\) if and only if \( l^+ \) belongs to the convex cone determined by \( \{c_i\} \). Since \( f \) is simplicial, they can intersect only once, unless the ray is tangent. Membership in this convex cone is equivalent to the feasibility of \( a_0 t^+ = \sum a_i c_i \), with \( a_i \geq 0 \) and \( a_0 > 0 \). This homogeneous system of \( k \) equations in \( k \) unknowns can be solved by Gaussian elimination in \( O(k^3) \) operations. If the system has solutions, it is easy to test whether any satisfy the constraints on the \( a_i \). Let \( \bar{a} \) be such a solution, if it exists.

When the homogeneous system has a 1-dimensional solution space, the ray \( \bar{a} + tl^+ \) "crosses" this part of the boundary. To determine the contribution to the intersection number, determine whether the vectors \( c_i \), taken in order, form a right- or left-handed coordinate system. In other words, find the sign of the determinant composed of these vectors, again using \( O(k^3) \) operations. Let \( \varepsilon \) be this sign. Note that when \( f \) collapses dimension, \( \varepsilon = 0 \). The contribution to the intersection number from this piece of the boundary is

\[
\begin{align*}
0 & \quad \text{if the solution space is multidimensional. The ray is tangent here, and} \\
\varepsilon & \quad \text{if } \bar{a} \text{ has all } a_i > 0, \text{ so that the intersection occurred in the interior of the} \\
\varepsilon/2 & \quad \text{if } \bar{a} \text{ has exactly one } a_i = 0. \text{ One neighboring simplex shares the intersection point, and the contributions will support each other or cancel.} \\
\varepsilon/r(j) & \quad \text{if } \bar{a} \text{ has } j \text{ values such that } a_i = 0, \text{ and } r(j) \text{ simplices share that facet.}
\end{align*}
\]

Alternatively, the complications of nonunit contributions to the intersection number can be avoided by picking a better direction.

Due to the technicalities of simplicial subdivision for large \( k \), we first describe the algorithm in the 3-color case. Here the discrete parameter space \( B_3 \) is the product of two segments of length \( n \), simplicially subdivided. To "divide and conquer" the space, split each of the segments in half; this partitions the parameter space into four smaller squares. Together, the boundaries of these squares contain just under \( 6n \) integer points. Having applied the algorithm first to \( B_2 \), we know that none of these boundary points
describe bisections. To determine which region contains a bisection, we must compute
the winding number for \( f \) restricted to the boundary of each of the four regions.

Let \( l^* \) be an arbitrary direction in \( \mathbb{A}_3 \), say \((.5, -.25, -.25)'\). The boundary of each
region consists of 1-dimensional unit simplices, i.e. segments with endpoints, \( b_2 \). Under
the simplicial \( f \), each maps to a segment. The segment \( f(b_1), f(b_2) \) intersects the ray
\( \hat{a} + t\hat{l}^* \) if \( \hat{l}^* \) lies in the convex hull of \( \{c_1, c_2\} \), in which case the sign of \( \det (c_1, c_2) \)
determines the contribution to the intersection number. Clearly, these operations can
be done in constant time per integer point on the boundary as the segments of the
boundary are traversed.

Choosing one of the regions with a nonzero winding number, perform the same
search on the \( n/2 \) by \( n/2 \) grid. The number of function evaluations performed to find
a solution point is at most \( \sum 6n/2^i = 12n \). If \( O(n) \) storage is available, then the
contributions obtained from the boundary segments of the current region can be stored
when computed, so that that information need not be recomputed when they recur in
boundaries of smaller regions. This would save a factor of \( \frac{1}{3} \), because the number of
function evaluations (with the same amount of computation for each), would be
\( 4n + \sum 2n/2^i = 8n \). Note that the preliminary step of checking the boundary for smaller
bisections can be dropped, because the first full step of the main algorithm includes
that.

A 2-dimensional simplex can be cut into four half-size simplices. If \( y_1 + y_2 + y_3 \leq n \),
these pieces are those with a given \( y_i \leq n/2 \), and the piece where all \( y_i \leq n/2 \). Dividing
\( B_4 \), the product of a 2-simplex and a 1-simplex, partitions it into eight pieces; dividing
\( B_5 \), the product of two 2-simplices, partitions it into sixteen pieces. The boundaries
of these pieces still have approximately one lower-dimensional simplex for each vertex,
and can be traversed using a small number of paths in which only one vertex changes
between neighboring simplices. \( B_4 \) has approximately \( 4.5n^2 + 3(n/2)^2 \) of these boundary
vertices; \( B_5 \) has approximately \( 9n(n/2)^2 \). With divide-and-conquer, the total number
of function evaluations is \( \sum 6(n/2)^2 + o(n^2) = 8n^2 + o(n^2) \) for \( B_4 \) and \( \sum 4.5(n/2)^3 +
o(n^3) = 36n^3/7 + o(n^3) \) for \( B_5 \). Storage of the outer boundary as before permits a
reduction in the leading constant, but the factor saved decreases as \( k \) increases. For
\( k = 3, 4, 5 \), it is \( \frac{1}{3}, \frac{1}{6}, \frac{1}{12} \).

In general, a \( d \)-dimensional simplex can be cut into \( 2^d \) half-size \( d \)-dimensional
simplices. The number of vertices in the boundaries of the various pieces is \( O(n^{k-2}) \),
and using divide-and-conquer allows the entire algorithm to run in \( O(n^{k-2}) \).

6. Application to graph separators. Leighton pointed out an application of the
discrete result to separators in graphs. For present purposes, we define an \( f(n) \)-separator
of a graph to be a balanced binary tree with the vertices of the graph as leaves and
no more than \( f(n) \) edges of the graph between vertices in different subtrees of a
subtree with \( n \) leaves. By “balanced,” we mean half of the leaves in any subtree belong
to each of its subtrees. When the induced subgraph on the \( n \) leaf nodes of some partial
tree has its nodes split into its “left” and “right” sets, the induced edge cut has “few”
edges, i.e. bounded by \( f(n) \).

By applying the circle coloring result, it is possible to obtain a more refined
splitting of the nodes at the cost of some extra edges in the cut. In particular, suppose
the nodes come in \( k \) “types,” i.e. are labeled arbitrarily with \( k \) colors, and let \( |H| \) be
the number of vertices in \( H \). Then

Theorem 4. Suppose any induced subgraph \( H \) of a graph \( G \) has an \( f(|H|) \)-separator
and the nodes of \( G \) are labeled arbitrarily with \( k \) colors. If \( f(n) = O(n^j) \) for some \( j > 0 \),
then \( G \) has an \( O(kn^j) \)-separator that separates the vertices of each color as evenly as
possible at each level, in addition to separating the full vertex set as evenly as possible. If \( f(n) = O((\log n)^4) \), then \( G \) has an \( O((\log n)^{4+1}) \)-separator as described.

To explain this result more directly, we "unbend" the necklace result to an equivalent result about linear arrangements that eliminates the distinction between odd and even values of \( k \). A linear arrangement with \( k \) colors of beads has a bisection with at most \( k \) cuts. In reading from one end to the other, we switch from capturing beads to omitting beads each time we cross a cut. If the two ends are identified to become the reference point \( P \), this is precisely the necklace result. When \( k \) is odd, the first and last intervals are of opposite type, so that the corresponding intervals on the circle end at \( P \), and there are \( \lfloor k/2 \rfloor \) intervals used and omitted. When \( k \) is even, the first and last intervals on the line have the same type, so that in the circle version they merge at \( P \), and again there are \( \lfloor k/2 \rfloor \) intervals of each type. Thus "\( k \) cuts on the line segment" is a uniform way to state the result.

Any separator yields a particular ordering of the vertices of a graph, by reading the leaf nodes in order. Since the vertices have specified colors, this yields a linear arrangement of beads. Find a bisection with at most \( k \) cuts. Taking the intervals of captured beads yields a vertex partition of the original graph in which the vertices of each color are evenly split. To obtain an upper bound on the number of edges between the two parts, we take \( k \) times the maximum number of edges that can join vertices on opposite sides of a single cut. The endpoints of any such edge belong to opposite subtrees at some level. At most \( f(n/2^i) \) edges cross the cut at level \( i \). If \( f(n) = O(n^j) \) for some \( j \geq 0 \), then there are \( O(kn^j) \) edges across the vertex partition. If \( f(n) = O((\log n)^{4}) \), then there are \( O(k(\log n)^{4+1}) \) edges across the partition. To build the rest of the desired separator, consider each of the vertex parts separately, find the \( f(n/2) \)-separator on the subgraphs induced by those vertices, and recurse.

7. Continuous applications. Tom Trotter pointed out an application of the continuous result to a geometric problem in \( \mathbb{R}^3 \). Consider a curve \( \tilde{u}(t) = (u_1(t), u_2(t), u_3(t)), 0 \leq t \leq 2 \). An old problem \([4],[5]\) asks whether a curve in \( \mathbb{R}^2 \) always contains four points that determine a square. In \( \mathbb{R}^3 \) certainly one cannot hope for so much, but one can hope for a parallelogram. Actually, we can guarantee more. Every continuous curve in \( \mathbb{R}^3 \) contains four points that determine a parallelogram such that the opposite portions of the curve total half the length of the curve.

**Theorem 5.** If \( \tilde{u}(t) \) with \( 0 \leq t \leq 2 \) is a continuous curve in \( \mathbb{R}^3 \), then there are four points \( 0 \leq t_1 \leq t_2 \leq t_3 \leq t_4 \leq 2 \) such that \( \tilde{u}(t_2) - \tilde{u}(t_1) = \tilde{u}(t_3) - \tilde{u}(t_4) \) and \( t_2 - t_1 + t_4 - t_3 = 1 \).

**Proof.** Define a four-color circle coloring with density function \( \rho_i(t) = u_i(t) \) for \( i = 1, 2, 3 \) and \( \rho_4(t) = 1 - u_1(t) - u_2(t) - u_3(t) \). The derivatives may be discontinuous at isolated points, but they are continuously integrable and can be defined at discontinuities so that they sum to 1 at all points. The cumulative distributions yield the displacements of \( u_i(t) \) from \( u_i(0) \) and \( t \) minus the sum of those displacements. Integrating all the way from 0 to 2 yields the total amount of each color, which is 0 for the first three colors and 2 for the fourth, since the net displacement in any direction around the closed curve is 0.

Applying Theorem 2, let \( (t_1, t_2, t_3, t_4) \) be the endpoints of intervals in a bisection. Given the above "total amounts" of each color, we have

\[
0 = u_i(t_4) - u_i(t_3) + u_i(t_2) - u_i(t_1) \quad \text{for } i = 1, 2, 3,
\]

\[
1 = (t_4 - \sum u_i(t_4)) - (t_3 - \sum u_i(t_3)) + (t_2 - \sum u_i(t_2)) - (t_1 - \sum u_i(t_1)) \quad \text{for } i = 4,
\]

which yields \( \tilde{u}(t_4) - \tilde{u}(t_3) = \tilde{u}(t_1) - \tilde{u}(t_2) \) and \( t_4 - t_3 + t_2 - t_1 = 1 \). \( \Box \)
Narendra Karmarkar noticed another application, of which the following is a variant.

**Theorem 6.** Let $f_1, \ldots, f_{k-1}$ be $k-1$ continuously integrable functions on an interval $[a, b]$. Then there is a function orthogonal to each of $f_1, \ldots, f_{k-1}$ that takes on only the values $\pm 1$ and changes sign at most $k$ times in the interval $[a, b]$.

**Proof.** Introduce a $k$th function that is 1 minus the sum of the others. Apply the linear version of the continuous bisection result to obtain a bisection using at most $k$ cuts. Define the new function to be $+1$ on the intervals in the bisection and $-1$ on those in its complement. □

Karmarkar actually noted the corresponding result for periodic functions, using the version of the continuous result applicable to circle colorings.

**8. Conclusion.** In addition to the NP-completeness of the original problem, related problems remaining open include generalizations to higher dimensions and splits in other proportions. If intervals must be chosen to capture a fraction $\alpha$ of each color, it is no longer always possible to do it with $\lceil k/2 \rceil$ intervals. For example, consider the following arrangement with four colors of beads: let the beads of colors 1, 2, 3 appear contiguously, and put $\frac{1}{3}$ of the beads of color 4 between color 1 and color 2, between color 2 and color 3, and between color 3 and color 1. Restricted to two intervals, there must be a cut within each of colors 1, 2, 3, and the fourth cut appears somewhere. This means that one stretch of color 4 is entirely included, and one stretch is entirely omitted, so the fraction of color 4 captured must lie between $\frac{1}{3}$ and $\frac{2}{3}$. The question is, for what range of values of $\alpha$ will $\lceil k/2 \rceil$ intervals suffice? If $\alpha = 1/l$, a more difficult variation would be to find $l$ disjoint sets of intervals such that each set contains $1/l$ of each color. Here one might want to minimize the total number of intervals or the maximum in any set.

In moving to higher dimensions, the torus and 2-simplex may be considered. For the discrete or continuous torus, a simple analogue of choosing intervals would be choosing an even number of horizontal and vertical lines to create a “checkerboard,” capturing the units in the regions having a given parity in the checkerboard. However, it is no longer clear that a bisection must always exist. For colorings of a simplex, the bisection result for linear arrangements is the candidate for generalization. Given a triangulated simplex, again cuts could be made parallel to the boundaries to capture the material in the resulting regions of a given parity, but here there seems to be an example with two colors where such a bisection does not exist. There may be an analogue of the continuous result for nonrectilinear cuts in a simplex, for regions formed on the surface of a sphere by passing planes through the origin, etc.

**REFERENCES**


