

From Curves to Words and Back Again: Geometric Computation of Minimum-Area Homotopy

Hsien-Chih Chang¹, Brittany Terese Fasy^{2,3}[0000–0001–6714–7988],
Bradley McCoy³[0009–0009–1450–5978], David L. Millman, and
Carola Wenk⁴[0000–0001–9275–5336]

¹ Department of Computer Science, Dartmouth College, Hanover, USA.

² Department of Mathematical Sciences, Montana State University, Bozeman, USA.

³ School of Computing, Montana State University, Bozeman, USA.

⁴ Department of Computer Science, Tulane University, New Orleans, USA.

Abstract. Let γ be a generic closed curve in the plane. Samuel Blank, in his 1967 Ph.D. thesis, determined if γ is self-overlapping by *geometrically* constructing a combinatorial word from γ . More recently, Zipei Nie, in an unpublished manuscript, computed the minimum homotopy area of γ by constructing a combinatorial word *algebraically*. We provide a unified framework for working with both words and determine the settings under which Blank’s word and Nie’s word are equivalent. Using this equivalence, we give a new geometric proof for the correctness of Nie’s algorithm. Unlike previous work, our proof is constructive which allows us to naturally compute the actual homotopy that realizes the minimum area. Furthermore, we contribute to the theory of self-overlapping curves by providing the first polynomial-time algorithm to compute a self-overlapping decomposition of any closed curve γ with minimum area.

Keywords: curve representation, crossing sequence, homotopy area, self-overlapping curve, fundamental group, Dehn twist, change of basis, cancellation norm.

1 Introduction

A *closed curve* in the plane is a continuous map γ from the circle $\mathbb{S}^1$ to the plane $\mathbb{R}^2$. In the plane, any closed curve is homotopic to a point. A homotopy that sweeps out the minimum possible area is a *minimum homotopy*. Chambers and Wang [4] introduced the minimum homotopy area between two simple homotopic curves with common endpoints as a way to measure the similarity between the two curves. They suggest that homotopy area is more robust against noise than another popular similarity measure on curves called the *Fréchet distance*. However, their algorithm requires that each curve be simple, which is restrictive.

Fasy, Karakoç, and Wenk [12] proved that the problem of finding the minimum homotopy area is easy on a closed curve that is the boundary of an immersed disk. Such curves are called *self-overlapping* [10,15,18,23,24,26]. They

also established a tight connection between minimum-area homotopy and self-overlapping curves by showing that any generic closed curve can be decomposed at some vertices into self-overlapping subcurves such that the combined homotopy from the subcurves is minimum. This structural result gives an exponential-time algorithm for the minimum homotopy area problem by testing each decomposition in a brute-force manner.

Nie, in an unpublished manuscript [19], described a polynomial-time algorithm to determine the minimum homotopy area of any closed curve in the plane. Nie’s algorithm borrows tools from geometric group theory by representing the curve as a word in the fundamental group $\pi_1(\gamma)$, and connects minimum homotopy area to the *cancellation norms* [2,3,21] of the word, which can be computed using a dynamic program. However, the algorithm does not naturally compute an associated *minimum-area homotopy*.

Alternatively, one can interpret the words from the dynamic program *geometrically* as crossing sequences by traversing any subcurve cyclicly and recording the crossings along with their directions with a collection of nicely-drawn *cables* from each face to a point at infinity. Such geometric representation is known as the Blank words [1,22]. In fact, the first application of these combinatorial words given by Blank is an algorithm that determines if a curve is self-overlapping. Blank words are geometric in nature and thus the associated objects are polynomial in size. When attempting to interpret Nie’s dynamic program from the geometric view, one encounters the question of how to extend Blank’s definition of cables to *subcurves*, where the cables inherited from the original curve are no longer positioned well with respect to the subcurves. To our knowledge, no geometric interpretation of the dynamic program is known.

1.1 Our Contributions

We first show that Blank and Nie’s word constructions are, in fact, equivalent under the right assumptions (Section 3). Next, we extend the definition of Blank’s word to subcurves and arbitrary cable drawings (Section 4.1), and interpret the dynamic program by Nie geometrically (Section 4.2). Using the self-overlapping decomposition theorem by Fasy, Karakoç, and Wenk [12] we provide a correctness proof to the algorithm. Finally, we conclude with a new result that a minimum-area self-overlapping decomposition can be found in polynomial time. We emphasize that extending Blank words to allow arbitrary cables is in no way straightforward. In fact, many assumptions on the cables have to be made in order to connect self-overlapping curves and minimum-area homotopy; handling arbitrary cable systems, as seen in the dynamic program, requires further tools from geometric topology like *Dehn twists*.

2 Background

In this section, we introduce concepts and definitions that are used throughout the paper. We assume the readers are familiar with the basic terminology for curves and surfaces.

2.1 Curves and Graphs

A *closed curve* in the plane is a continuous map $\gamma : \mathbb{S}^1 \rightarrow \mathbb{R}^2$, and a *path* in the plane is a continuous map $\zeta : [0, 1] \rightarrow \mathbb{R}^2$. A path ζ is *closed* when $\zeta(0) = \zeta(1)$. In this work, we are presented with a *generic* curve; that is, one where there are a finite number of self-intersections, each of which is transverse and no three strands cross at the same point. See Figure 1 for an example. The image of a generic closed curve is naturally associated with a four-regular plane graph. The self-intersection points of a curve are *vertices*, the paths between vertices are *edges*, and the connected components of the complement of the curve are *faces*. Given a curve, choose an arbitrary starting point $\gamma(0) = \gamma(1)$ and orientation for γ .

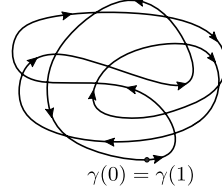


Fig.1: A generic plane curve induces a four-regular graph.

The *dual graph* γ^* is another (multi-)graph, whose vertices represent the faces of γ , and two vertices in γ^* are joined by an edge if there is an edge between the two corresponding faces in γ . The dual graph is another plane graph with an inherited embedding from γ .

Let T be a spanning tree of γ . Let E denote the set of edges in γ , the tree T partitions E into two subsets, T and $T^* := E \setminus T$. The edges in T^* define a spanning tree of γ^* called the *cotree*. The partition of the edges (T, T^*) is called the *tree-cotree pair*.

We call a rooted spanning cotree T^* of γ^* a *breadth-first search tree* (BFS-tree) if it can be generated from a breadth-first search rooted at the vertex in γ^* corresponding to the unbounded face in γ . Each bounded face f of γ is a vertex in a breadth-first search tree T^* , we associate f with the unique edge incident to f^* in the direction of the root. Thus, there is a correspondence between edges of T^* and faces of γ .

2.2 Homotopy and Isotopy

A *homotopy* between two closed curves γ_1 and γ_2 that share a point p_0 is a continuous map $H : [0, 1] \times \mathbb{S}^1 \rightarrow \mathbb{R}^2$ such that $H(0, \cdot) = \gamma_1$, $H(1, \cdot) = \gamma_2$, and $H(s, 0) = p_0 = H(s, 1)$. We define a homotopy between two paths similarly, where the two endpoints are fixed throughout the continuous morph. Notice that homotopy between two closed curves as *closed curves* and the homotopy between them as *closed paths* with an identical starting points are different. A homotopy between two injective paths ζ_1 and ζ_2 is an *isotopy* if every intermediate path $H(s, \cdot)$ is injective for all s . The notion of isotopy naturally extends to a collection of paths.

We can think of γ as a topological space and consider the *fundamental group* $\pi_1(\gamma)$. Elements of the fundamental group are called *words*, whose letters correspond to equivalence classes of homotopic closed paths in γ . The fundamental group of γ is a free group with basis consisting of the classes corresponding to the cotree edges of any tree-cotree pair of γ .

Let H be a homotopy between curves γ_1 and γ_2 . Let $\#H^{-1}(x): \mathbb{R}^2 \rightarrow \mathbb{Z}$ be the function that assigns to each $x \in \mathbb{R}^2$ the number of times the intermediate curves H sweep over x . The homotopy area of H is

$$\text{Area}(H) := \int_{\mathbb{R}^2} \#H^{-1}(x) dx.$$

The minimum area homotopy between γ_1 and γ_2 is the infimum of the homotopy area over all homotopies between γ_1 and γ_2 . We denote this by $\text{Area}_H(\gamma_1, \gamma_2) := \inf_H \text{Area}(H)$. When γ_2 is the constant curve at a specific point p_0 on γ_1 , define $\text{Area}_H(\gamma) := \text{Area}_H(\gamma, p_0)$. See Figure 2 for an example of a homotopy.

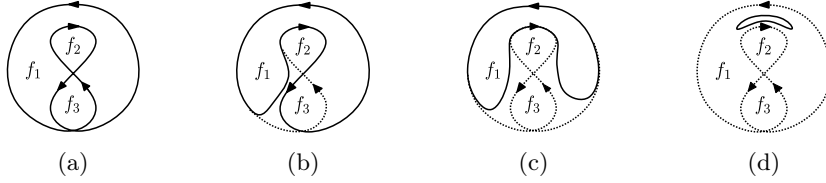


Fig. 2: (a) A generic closed curve in the plane. (b) We see a homotopy that sweeps over the face f_3 . (c) The homotopy sweeps f_3 again. (d) The homotopy avoids sweeping over the face f_2 . This is a minimum area homotopy for the curve, the area is $\text{Area}(f_1) + 2 \cdot \text{Area}(f_3)$.

For each $x \in \mathbb{R} \setminus \gamma$, the *winding number* of γ at x , denoted as $\text{wind}(x, \gamma)$, is the number of times γ “wraps around” x , with a *positive* sign if it is counterclockwise, and *negative* sign otherwise. The winding number is a constant on each face. The *winding area* of γ is defined to be the integral

$$\text{Area}_W(\gamma) := \int_{\mathbb{R}^2} |\text{wind}(x, \gamma)| dx = \sum_{\text{face } f} |\text{wind}(f, \gamma)| \cdot \text{Area}(f).$$

The *depth* of a face f is the minimal number of edges crossed by a path from f to the exterior face. The depth is a constant on each face. We say the depth of a curve is equal the maximum depth over all faces. We define the *depth area* to be

$$\text{Area}_D(\gamma) := \int_{\mathbb{R}^2} \text{depth}(x, \gamma) dx = \sum_{\text{face } f} \text{depth}(f) \cdot \text{Area}(f).$$

Chambers and Wang [4] showed that the winding area gives a lower bound for the minimum homotopy area. On the other hand, there is always a homotopy with area $\text{Area}_D(\gamma)$; one such homotopy can be constructed by smoothing the curve at each vertex into simple depth cycles [5], then contracting each simple cycle. Therefore we have

$$\text{Area}_W(\gamma) \leq \text{Area}_H(\gamma) \leq \text{Area}_D(\gamma). \quad (1)$$

2.3 Self-Overlapping Curves

A generic curve γ is *self-overlapping* if there is an immersion of the two disk $F : \mathbb{D}^2 \rightarrow \mathbb{R}^2$ such that $\gamma = F|_{\partial\mathbb{D}^2}$. We say a map F *extends* γ . The image $F(\mathbb{D}^2)$ is the *interior* of γ . There are several equivalent ways to define self-overlapping curves [10,24,23,15,18]. Properties of self-overlapping curves are well-studied [9]; in particular, any self-overlapping curve has rotation number 1, where the *rotation number* of a curve γ is the winding number of the derivative γ' about the origin [26]. Also, the minimum homotopy area of any self-overlapping curve is equal to its winding area: $\text{Area}_W(\gamma) = \text{Area}_H(\gamma)$ [12].

The study of self-overlapping curves traces back to Whitney [26] and Titus [24]. Polynomial-time algorithms for determining if a curve is self-overlapping have been given [1,23], as well as NP-hardness result for extensions to surfaces and higher-dimensional spaces [7].

For any curve, the *intersection sequence*⁵ $[\gamma]_V$ is a cyclic sequence of vertices $[v_0, v_1, \dots, v_{n-1}]$ with $v_n = v_0$, where each v_i is an intersection point of γ . Each vertex appears exactly twice in $[\gamma]_V$. Two vertices x and y are *linked* if the two appearances of x and y in $[\gamma]_V$ alternate in cyclic order: $\dots x \dots y \dots x \dots y \dots$.

A pair of symbols of the same vertex x induces two natural subcurves generated by *smoothing* the vertex x ; see Figure 3 for an example. (In this work, every smoothing is done in the way that respects the orientation and splits the curve into two subcurves.) A *vertex pairing* is a collection of pairwise unlinked vertex pairs in $[\gamma]_V$.

A *self-overlapping decomposition* Γ of γ is a vertex pairing such that the induced subcurves are self-overlapping; see Figure 3b and Figure 3d for examples. The subcurves that result from a vertex pairing are not necessary self-overlapping; see Figure 3c. For a self-overlapping decomposition Γ of γ , denote the set of induced subcurves by $\{\gamma_i\}_{i=1}^\ell$. Since each γ_i is self-overlapping, the minimum homotopy area is equal to its winding area. We define the *area of self-overlapping decomposition* to be

$$\text{Area}_\Gamma(\gamma) := \sum_{i=1}^\ell \text{Area}_W(\gamma_i) = \sum_{i=1}^\ell \text{Area}_H(\gamma_i).$$

Fasy, Karakoç, and Wenk [12,14] proved the following structural theorem.

Theorem 1 (Self-Overlapping Decomposition [12, Theorem 20]). *Any curve γ has a self-overlapping decomposition whose area is minimum over all null-homotopies of γ .*

3 From Curves to Words

In order to work with plane curves, one must choose a *representation*. An important class of representations for plane curves are the various *combinatorial words*.

⁵ also known as the unsigned Gauss code [5,13]

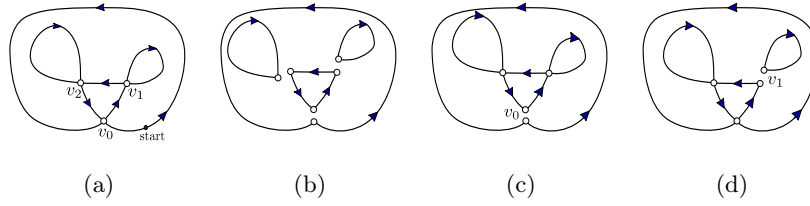


Fig. 3: (a) Curve γ with intersection sequence $\gamma_V = [v_0, v_1, v_1, v_2, v_2, v_0]$. (b) All vertices are paired. (c) One of the subcurves is not self-overlapping. (d) Both subcurves are self-overlapping.

One example is the *Gauss code* [13]. Determining whether a Gauss code corresponds to an actual plane curve is one of the earliest computational topology questions [8].

A plane curve (and its homotopic equivalents) can also be viewed as a word in the fundamental group $\pi_1(\gamma)$ of γ [1, 22, 19]. If we put a point p_i in each bounded face f_i , the curve γ is generated by the unique generators of each $\mathbb{R}^2 - \{p_i\}$. Nie [19] represents curves as words in the fundamental group to find the minimum area swept out by contracting a curve to a point. If the curve lies in a plane with punctures, one can define the *crossing sequence* of the curve with respect to a *system of arcs*, cutting the plane open into a simply-connected region. Blank [1] represents curves using a crossing sequences to determine if a curve is self-overlapping. While Blank constructed the words *geometrically* by drawing arcs and Nie defined the words *algebraically*, the dual view between the system of arcs and fundamental group suggests that the resemblance between Blank and Nie's constructions is not a coincidence.

In this section, we describe the construction by Blank; then, we interpret Blank's construction as a way of choosing the basis for the fundamental group under further restriction [22]. We prove that the Blank word is indeed unique when the restriction is enforced, providing clarification to Blank's original definition. In the full version of the paper, we give a complete description of Nie's word construction and prove that Nie's word and Blank's word are equivalent.

3.1 Blank's Word Construction

We now describe Blank's word construction [1, page 5]. Let γ be a generic closed curve in the plane, pick a point in the unbounded face of γ , call it the *basepoint* p_0 . From each bounded face f_i , pick a *representative point* p_i . Now connect each p_i to p_0 by a simple path in such a way that no two paths intersect each other. We call the collection of such simple paths a *cable system*, denoted as Π , and each individual path π_i from p_i to p_0 as a *cable*.

Orient each π_i from p_i to p_0 . Now traverse γ from an arbitrary *starting point* of γ and construct a cyclic word by writing down the indices of γ crossing the cables π_i in the order they appear on γ ; each index i has a *positive* sign if we

cross π_i from right to left and a *negative* sign if from left to right. We denote negative crossing with an overline $\bar{i}$. We call the resulting combinatorial word over the faces a *Blank word* of γ with respect to Π , denoted as $[\gamma]_B(\Pi)$. Figure 4 provides an example of Blank's construction.

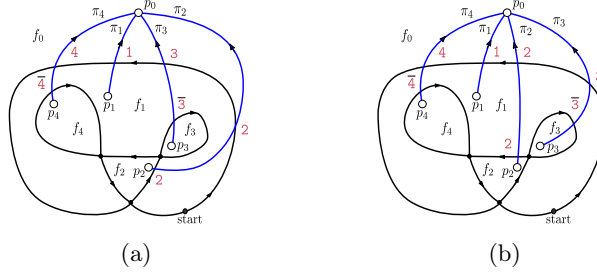


Fig. 4: (a) A curve γ with labeled faces and edges, Π_a is drawn in blue. The Blank word of γ corresponding to Π_a is $[\gamma]_B(\Pi_a) = [2314234]$. (b) The same curve with a different choice of cables Π_b . The corresponding Blank word is $[\gamma]_B(\Pi_b) = [3214324]$.

A word w is *reduced* if there are no two consecutive symbols in w that are identical and with opposite signs. We can enforce every Blank word to be reduced by imposing the following *shortest path assumption*: each cable has a minimum number of intersections with γ among all paths from p_i to p_0 . A simple proof [1,6] shows that if Π satisfies the shortest path assumption, the corresponding Blank word with respect to Π is reduced. However, the choice of the cable system, and how it affects the constructed Blank word, was never explicitly discussed in the original work (presumably because for the purpose of detecting self-overlapping curves, any cable system satisfying the shortest path assumption works). In general, reduced Blank words constructed from different cable systems for the same curve are not identical, see Figure 4a and Figure 4b for an example. In this paper, we show that if the two cable systems have the same *cable ordering*—the (cyclic) order of cables around point p_0 in the unbounded face—then their corresponding (reduced) Blank words are the same, under proper assumptions on the cable system.

Our first observation is that the Blank words are invariant under cable isotopy; therefore the cable system can be specified up to isotopy.

Lemma 1 (Isotopy Invariance).

The reduced Blank word is invariant under cable isotopy.

Proof. Let γ be a curve. Discretize the isotopy of the cables and consider all the possible *homotopy moves* [5] performed on γ and the cables involving up to two strands from γ and a cable, because isotopy disallows the crossing of two cables. No $1 \leftrightarrow 0$ move—the move that creates/destroys a self-loop—is possible as cables do not self-intersect. Any $2 \leftrightarrow 0$ move which creates/destroys a bigon is in between a cable and a strand from γ , which means the two intersections

must have opposite signs, and therefore the reduced Blank word does not change. Any $3 \rightarrow 3$ move which moves a strand across another intersection does not change the signs of the intersections, so while the order of strands crossing the cable changes, the order of cables crossed by γ remains the same. Thus the reduced Blank word stays the same.

We remark that we can perform an isotopy so that the Blank words are reduced even when the cables are not necessary shortest paths. In the rest of the paper, we sometimes assume Blank words to be reduced based on the context.

Manage the Cable Systems Next, we show that Blank words are well-defined once we fix the choice of basepoint p_0 and the cyclic cable ordering around p_0 , as long as the cables are drawn in a reasonable way. Fix a tree-cotree pair (T, T^*) of γ , where the root of the cotree is on p_0 . We say that a cable system Π is *managed* with respect to the cotree T^* if each path π_i has to be a path on T^* from the root p_0 to the leaf p_i . Given such a collection of cotree paths, one can slightly perturb them to ensure that all paths are simple and disjoint.⁶ See Figure 5 for examples. Not every cable system can be managed with respect to T^* .

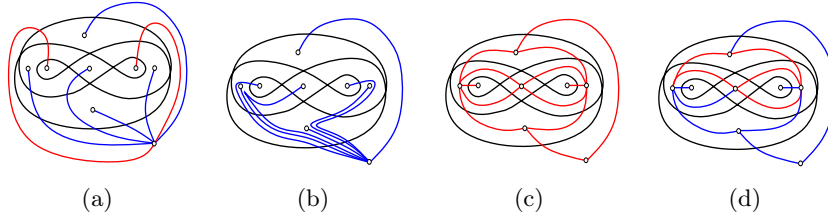


Fig. 5: (a) A cable system Π_1 on γ that is not managed. The red cables do not follow existing paths to the exterior face. (b) A managed cable system Π_2 on γ . (c) The dual γ^* in red. (d) The spanning tree T^* in γ^* generated by the managed cable system Π_2 .

In the full paper, we show that if two managed cable systems satisfying shortest path assumption with identical cable ordering around p_0 , their corresponding Blank words are the same. Note that managed cable systems require a fixed tree-cotree pair. We emphasize that the shortest path assumption is necessary; one can construct two (not necessarily shortest) cable systems having the same cable ordering but different corresponding reduced Blank words.

Lemma 2 (Blank Word is Unique). *Given a curve γ , if the basepoint p_0 and the cable ordering of a managed cable system Π satisfying the shortest path assumption is fixed, then the Blank word of γ is unique.*

Therefore, given any plane curve γ , the Blank word is well-defined (if exists), independent of the cable system after specifying a cyclic permutation of all the bounded faces of γ .

⁶ In other words, the cables are *weakly-simple* [25].

4 Foldings and Self-Overlapping Decompositions

In this section, we give a geometric proof of the correctness to Nie's dynamic program. To do so, we show that the minimum homotopy area of a curve can be computed from its Blank word using an algebraic quantity of the word called the *cancellation norm*, which is independent of the drawing of the cables. We then show a minimum-area self-overlapping decomposition can be found in polynomial time.

4.1 The Cancellation Norm and Blank Cuts

Given a (cyclic) word w , a *pairing* is a letter and its inverse $(\mathbf{f}, \bar{\mathbf{f}})$ in w . Two letter pairings, $(\mathbf{f}_1, \bar{\mathbf{f}}_1)$ and $(\mathbf{f}_2, \bar{\mathbf{f}}_2)$, are *linked* in a word if the letter pairs occur in alternating order in the word, $[\cdots \mathbf{f}_1 \cdots \mathbf{f}_2 \cdots \bar{\mathbf{f}}_1 \cdots \bar{\mathbf{f}}_2 \cdots]$. A *folding* of a word is a set of letter pairings such that no two pairings in the set are linked. For example, in the word $[231546\bar{5}4\bar{6}2\bar{3}]$ the set $\{(\bar{5}, 5), (\bar{3}, 3)\}$ is a folding while $\{(\bar{5}, 5), (6, \bar{6})\}$ is not.

The cancellation norm is defined in terms of pairings. The norm also applies in the more general setting where every letter has an associated nonnegative weight. A letter is *unpaired* in a folding if it does not participate in any pairing of the folding. For a word of length m , computing the cancellation norm takes $\mathcal{O}(m^3)$ time and $\mathcal{O}(m^2)$ space [2,21]. Recently, a more efficient algorithm for computing the cancellation norm appears in Bringmann *et al.* [3]; this algorithm uses fast matrix multiplications and runs in $\mathcal{O}(m^{2.8603})$ time.

The *weighted cancellation norm* of a word w is defined to be the minimum sum of weights of all the unpaired letters in w across all foldings of w [2,21]. If w is a word where each letter $\mathbf{f}_i$ corresponds to a face f_i of a curve, we define the weight of $\mathbf{f}_i$ to be $\text{Area}(f_i)$. The *area of a folding* is the sum of weights of all the unpaired symbols in a folding. The weighted cancellation norm becomes $\|w\| := \min_{\mathcal{F}} \sum_i \text{Area}(f_i)$ where $\mathcal{F}$ is the set of all foldings of w and i ranges over all unpaired letter in w .

A dynamic program, similar to the one for matrix chain multiplication, is applied on the word. Let $w = f_1 f_2 \cdots f_\ell$ where $\ell \geq 2$. Assume we have computed the cancellation norm of all subwords with length less than ℓ . Let $w' = f_1 f_2 \cdots f_{\ell-1}$. If f_ℓ is not the inverse of f_i for $1 \leq i \leq \ell-1$, then f_ℓ is unpaired and $\|w\| = \|w'\| + \text{Area}(f_\ell)$. Otherwise, f_ℓ participates in a folding and there exists at least one k where $1 \leq k \leq \ell-1$ and $f_k = f_\ell^{-1}$. Let $w_1 = f_1 \cdots f_{k-1}$ and $w_2 = f_{k+1} \cdots f_{\ell-1}$. Then, we find the k that minimizes $\|w_1\| + \|w_2\|$. We have

$$\|w\| = \min\{\|w'\| + \text{Area}(f_\ell), \min_k \{\|w_1\| + \|w_2\|\}\}$$

Nie shows that the weighted cancellation norm whose weights correspond to face areas is equal to the minimum homotopy area using the triangle inequality and geometric group theory. Our proof that follows is more geometric and leads to a natural homotopy that achieves the minimum area.

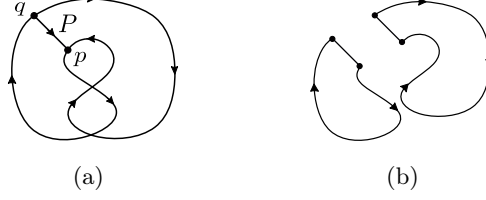


Fig. 6: (a) A curve with labeled path P . (b) The two induced subcurves from cutting along P .

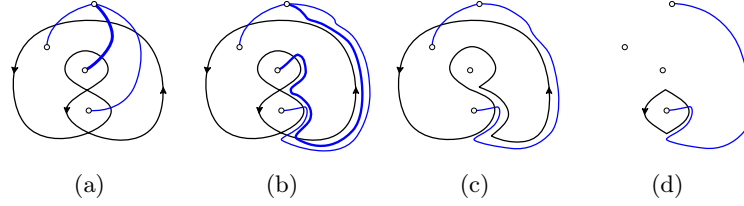


Fig. 7: (a) A curve with cables. (b) Isotopy the cables to not partially cut any faces. (c) One subcurve resulting from cutting along the middle cable. The curve is weakly simple and there are two cables in this face. (d) The other subcurve.

We now show how to interpret the cancellation norm geometrically. Let $(\mathbf{f}, \bar{\mathbf{f}})$ be a face pairing in a folding of the word $[\gamma]_B(\Pi)$ for some cable system Π . Denote the cable in Π ending at face f as π_f . Cable π_f intersects γ at two points corresponding to the pairing $(\mathbf{f}, \bar{\mathbf{f}})$, which we denote as p and q respectively. Let π'_f be the simple subpath of π_f so that $\pi'_f(0) = q$ and $\pi'_f(1) = p$. We call π'_f a *Blank cut* [1,10,17] (see Figure 6). Any face pairing defines a Blank cut, and the result of a Blank cut produces two curves each with fewer faces than the original curve: namely, γ_1 which is the restriction of γ from q to p following by the reverse of path π'_f , and γ_2 which is the restriction of γ from p to q followed by path π'_f .

In order to not partially cut any face, we require all Blank cuts to occur along the boundary of the face being cut. When cutting face f_i along path π_j , we reroute all cables crossing the interior of f_i , including π_j but excluding π_i , along the boundary of f_i through an *isotopy*, so that no cables intersect π_i . Lemma 1 ensures that the reduced Blank word remains unchanged. See Figure 7 for an example. Notice that different cables crossing f_i might be routed around different sides of f_i in order to avoid intersecting cable π_i and puncture p_i . This way, we ensure the face areas of the subcurves are in one-to-one correspondence with the symbols in the subwords induced by a folding.

Using the concept of Blank cut we can determine if a curve is self-overlapping. A subword σ of w is *positive* if $\sigma = \mathbf{f}_1 \mathbf{f}_2 \dots \mathbf{f}_k$, where each letter $\mathbf{f}_i$ is positive. A pairing $(\mathbf{f}, \bar{\mathbf{f}})$ is *positive* if one of the two subwords of the (cyclic) word w in between the two symbols $\mathbf{f}, \bar{\mathbf{f}}$ is positive; in other words, $w = [\mathbf{f} p \bar{\mathbf{f}} w']$ for some

positive word p and some word w' . A folding of w is called a *positive folding*⁷ if all pairings in w are positive, and the word constructed by replacing each positive pairing (including the positive word in-between) $\mathbf{f}p\bar{\mathbf{f}}$ in the folding with the empty string is still positive. Words that have positive foldings are called *positively foldable*. Blank established the characterization of self-overlapping curves through Blank cuts.

Theorem 2 (Self-Overlapping Detection [1]). *Curve γ is self-overlapping if and only if γ has rotation number 1 and $[\gamma]_B(\Pi)$ is positively foldable for any shortest Π .*

However, we face a difficulty when interpreting Nie’s dynamic program geometrically. In our proof we have to work with *subcurves* (and their extensions) of the original curve and the induced cable system. For example, after a Blank cut or a vertex decomposition, there might be multiple cables connecting to the same face creating multiple punctures per face, and cables might not be managed or follow shortest paths to the unbounded face (see Figure 7c and Figure 9b). In other words, the subword corresponding to a subcurve with respect to the induced cable system might not be a regular Blank word (remember that Blank word is only well-defined when the cable system is managed, all cables are shortest paths, and the cable ordering is fixed; see Section 2). To remedy this, we tame the cable system first by rerouting them into another cable system that is managed and satisfies the shortest path assumption, then merging all the cables ending at each face. We show that while such operations change the Blank word of the curve, the cancellation norm of the curve and the positive foldability does not change. We summarize the property needed below.

Lemma 3 (Cable Independence). *Let γ be any curve with two cable systems Π and Π' such that the weights of the cables in Π ending at any fixed face sum up to the ones of Π' . Then any folding F of $[\gamma](\Pi)$ can be turned into another folding F' of $[\gamma](\Pi')$, such that the area of the two foldings are identical. As a corollary, the minimum area of foldings (the cancellation norm) of $[\gamma](\Pi)$ and the existence of a positive folding of $[\gamma](\Pi)$ are independent of the choice of Π .*

In the full paper, we prove that for each folding there is a homotopy with equal area.

Lemma 4 (Folding to Homotopy). *Let γ be a curve and Π be a managed cable system satisfying the shortest path assumption, and let F be a folding of $[\gamma](\Pi)$. There exists a null-homotopy of γ with area equal to the area of F .*

⁷ Blank called these pairings *groupings*

4.2 Compute Min-Area Homotopy from Self-Overlapping Decomp.

A self-overlapping decomposition is a vertex decomposition where each subcurve is self-overlapping [12]. By Theorem 1, there exists a self-overlapping decomposition and an associated homotopy whose area is equal to the minimum homotopy area of the original curve.

In order to relate vertex decompositions and face decompositions, we define a word that includes both the faces and vertices. Given any curve γ and cable system Π , traverse γ and record both self-crossings and (signed) cable intersections; we call the resulting sequence of vertices and faces the *combined word* $[[\gamma]](\Pi)$. See Figure 8 for an example.

We now show that every self-overlapping decomposition (with respect to the vertex word of γ) determines a folding (of the face word of γ) using the combined word.

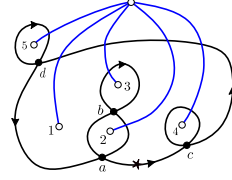


Fig. 8: A curve with combined word $[c4c4231d5da2b3ba]$.

Theorem 3 (S-O Decomp. to Folding). *Given a self-overlapping decomposition Γ and a cable system Π of γ , there exists a folding F of $[[\gamma]](\Pi)$ whose area is $\text{Area}_\Gamma(\gamma)$.*

Corollary 1 (Geometric Correctness). *The dynamic programming algorithm computes the minimum-area homotopy for any curve γ .*

Proof. By Theorem 1, there exists a self-overlapping decomposition with minimum homotopy area. By Theorem 3, some folding achieves a minimum area. Using Lemma 4, the minimum-area folding produces a minimum-area homotopy.

4.3 Min-Area Self-Overlapping Decomposition in Polynomial Time

Finally, we show that any *maximal* folding—where adding any extra pairs are linked—can be used to construct a self-overlapping decomposition.

Theorem 4 (Folding to S.O.D.). *Let γ be a curve and Π be a cable system. Given a maximal folding F of $[[\gamma]](\Pi)$, there is a self-overlapping decomposition of γ whose area is equal the area induced by the folding F .*

Proof. Begin with the combined word $[[\gamma]](\Pi)$. Decompose $[[\gamma]](\Pi)$ at the vertices given by the self-overlapping decomposition. Let $\Gamma = \{\gamma_1, \gamma_2, \dots, \gamma_s\}$ be the self-overlapping subcurves and $[[\gamma]](\Pi)_i$ be the corresponding subwords of $[[\gamma]](\Pi)$. If we remove the vertex symbols and turn each $[[\gamma]](\Pi)_i$ into a face word $[\gamma_i]'$, such word may not correspond to Blank words of the subcurves; indeed, when decomposing γ into subcurves by Γ , the subcurve along with the relevant cables may contain multiple cables per face and cables might not be managed or follow shortest paths. See Figure 9 for an example. However, we can first tame the cable system by choosing a new managed cable system Π^* where the cables

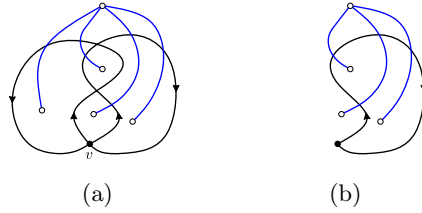


Fig. 9: We decompose the curve in (a) at vertex v into self-overlapping subcurves, the cable system on the induced subcurve in (b) has more than one marked point in a face and cables do not follow shortest paths.

follow shortest paths and has one cable per face (as in Section 3.1). Lemma 3 ensures that the cancellation norm and positive foldability of the subcurve remain unchanged. Denote the new face word of γ_i with respect to Π^* as $[\gamma_i] = [\gamma_i](\Pi^*)$.

Since each γ_i is a self-overlapping subcurve in Γ , we can find a positive folding F_i of $[\gamma_i]$ by Theorem 2, and the minimum homotopy area of γ_i is equal to the area of folding F_i . Now Lemma 3 implies that the subword $[\gamma_i]'$ from the original combined word also has a positive folding F_i' whose area is equal to the minimum homotopy area of γ_i . By combining all foldings F_i' of each face subword $[\gamma_i]'$, we create a folding F for $[\gamma](\Pi)$ (no pairings between different F_i' s can be linked). The area of folding F is equal to the sum of areas of foldings F_i' , which in turns is equal to $\sum_i \text{Area}_H(\gamma_i)$, that is, the homotopy area of self-overlapping decomposition $\text{Area}_\Gamma(\gamma)$. This proves the theorem.

The above theorem implies a polynomial-time algorithm to compute a self-overlapping decomposition with minimum area.

Corollary 2 (Polynomial Optimal Self-Overlapping Decomposition).

Let γ be a curve. A self-overlapping decomposition of γ with area equal to minimum homotopy area of γ can be found in polynomial time.

Proof. Apply the dynamic programming algorithm to compute the minimum-area folding F for $[\gamma](\Pi)$ with respect to some cable system Π . By Theorem 3 the area of F is equal to the minimum homotopy area of γ , and so does the corresponding self-overlapping decomposition given by Theorem 4.

Acknowledgements Brittany Terese Fasy and Bradley McCoy are supported by NSF grant DMS 1664858 and CCF 2046730. Carola Wenk is supported by NSF grant CCF 2107434 .

References

1. Samuel Joel Blank. *Extending Immersions and Regular Homotopies in Codimension 1*. PhD thesis, Brandeis University, May 1967.
2. Michael Brandenbursky, Światosław Gal, Jarek Kędra, and Michał Marcinkowski. The cancelation norm and the geometry of bi-invariant word metrics. *Glasg. Math. J.*, page 153–176, 2015.
3. Karl Bringmann, Fabrizio Grandoni, Barna Saha, and Virginia Vassilevska Williams. Truly subcubic algorithms for language edit distance and RNA folding via fast bounded-difference min-plus product. *SIAM J. Comput.*, 48(2):481–512, 2019.
4. Erin Chambers and Yusu Wang. Measuring similarity between curves on 2-manifolds via homotopy area. *29th ACM Symp. on Comput. Geom.*, pages 425–434, 2013.
5. Hsien-Chih Chang and Jeff Erickson. Untangling planar curves. *Discrete Comput. Geom.*, 58:889, 2017.
6. Dennis Frisch. Extending immersions into the sphere. 2010. URL: <http://arxiv.org/abs/1012.4923>.
7. David Eppstein and Elena Mumford. Self-overlapping curves revisited. *20th ACM-SIAM Symp. Discrete Algorithms*, pages 160–169, 2009.
8. Jeff Erickson. One-dimensional computational topology lecture notes. Lecture 7, 2020. URL: <https://mediaspace.illinois.edu/channel/CS+598+JGE+T1\textmdash%C2%A0Fall+2020/177766461/>.
9. Parker Evans. On Self-Overlapping Curves, Interior Boundaries, and Minimum Area Homotopies. Bachelor’s thesis, Tulane University, 2018.
10. Parker Evans, Brittany Terese Fasy, and Carola Wenk. Combinatorial properties of self-overlapping curves and interior boundaries. *36th ACM Symp. on Comput. Geom.*, 2020.
11. Benson Farb and Dan Margalit. *A Primer on Mapping Class Groups*. Princeton University Press, 2011.
12. Brittany Terese Fasy, Selçuk Karakoç, and Carola Wenk. On minimum area homotopies of normal curves in the plane. 2017. URL: <http://arxiv.org/abs/1707.02251>.
13. Carl Friedrich Gauss. Nachlass. I. Zur geometria situs. *Werke*, vol. 8, 271–281, 1900.
14. Selçuk Karakoç. *On Minimum Homotopy Areas*. PhD thesis, Tulane University, 2017.
15. Yijing Li and Jernej Barbič. Immersion of self-intersecting solids and surfaces. *ACM Trans. on Graph.*, 45:1–14, 2018.
16. Bruno Martelli. An introduction to geometric topology, 2016. [arXiv:arXiv:1610.02592](https://arxiv.org/abs/1610.02592).
17. Morris L. Marx. Extensions of normal immersions of S^1 into R . *Trans. Amer. Math. Soc.*, 187:309–326, 1974.
18. Uddipan Mukherjee. Self-overlapping curves: Analysis and applications. *Comput.-Aided Des.*, 40:227–232, 2014.
19. Zipei Nie. On the minimum area of null homotopies of curves traced twice. 2014. URL: <http://arxiv.org/abs/1412.0101>.
20. Zipei Nie. Private correspondence. 2016.
21. Ruth Nussinov and Ann B. Jacobson. Fast algorithm for predicting the secondary structure of single-stranded RNA. *Proc. Natl. Acad. Sci. USA*, 77(11):6309–6313, 1980.

- 22. Valentin Poénaru. Extension des immersions en codimension 1 (d'après Samuel Blank). *Séminaire N. Bourbaki (1966–1968)*, 10:473–505, 1968.
- 23. Peter Shor and Christopher Van Wyk. Detecting and decomposing self-overlapping curves. *Comput. Geom.: Theory and Applications*, 2:31–50, 1992.
- 24. Charles J. Titus. The combinatorial topology of analytic functions on the boundary of a disk. *Acta Math.*, pages 106(1–2):45–64, 1961.
- 25. Godfried Toussaint. On separating two simple polygons by a single translation. *Discrete Comput. Geom.*, 4(3):265–278, June 1989.
- 26. Hassler Whitney. On regular closed curves in the plane. *Compos. Math.*, 4:276–284, 1937.