THE FOLDED RIBBON THEOREM. A CONTRIBUTION TO THE STUDY OF IMMERSED CIRCLES

BY

GEORGE K. FRANCIS[1]

1. **Introduction.** A regular loop in the oriented plane is a closed curve that can be parametrized by a regular map from the oriented circle. The curve has a continuous nonvanishing vector[2].

A regular homotopy is a continuous one-parameter family of regular loops, for which the tangent fields also vary continuously with the time parameter. As such, a regular homotopy can be represented by a map from an annulus. If this map is differentiable of class C^1, its Jacobian will never vanish with rank less than one. If the Jacobian does not vanish at all, it is called a monotone regular homotopy, or monotopy for short. The sign of the Jacobian gives the orientation of the monotopy as an immersion of the annulus.

The tangent winding number of a regular loop is the topological degree of the tangent field. Two regular loops of like tangent winding number are regularly homotopic. It is the purpose of this paper to construct a regular homotopy between two such loops, which is the succession of two oppositely oriented monotopies. Such a homotopy can be thought of as a folded immersion of an annulus, whose single fold occurs along an intermediate regular loop.

A regular loop is normal if it has but a finite number of simple intersection points. A simple intersection point, or node, has two preimages on the circle, and the two tangents of the loop at this point are independent. The normal loops form a dense-open (generic) subspace of the space of regular loops under the conventional C^1-topology.

It is first shown that a regular loop is monotopic to a normal loop. The combinatorial theory of C. J. Titus is employed to keep track of a succession of signed detours that eventually modify this normal loop to a canonical normal loop proper to the tangent winding number class of the original loop. The detours do not change the tangent winding number.

Next is shown how two canonical normal loops of like tangent winding number are to be detoured to a common normal loop. An ancillary construction shows how a detour of a loop that terminates a monotopy modifies the monotopy to terminate

Received by the editors October 30, 1967.

(1) The contents of this paper form a part of the author's dissertation submitted as a partial requirement for the Ph.D. degree at the University of Michigan under the direction of Professor Charles J. Titus.

(2) The definitions will be made precise and the assertions will be proved presently.

on the detoured loop, provided the detour and the orientation of the monotopy have the same sign.

In summary then, given two regular loops of like tangent winding number, each is monotopic to a normal loop. For each of these two normal loops there is a succession of detours ending in the common loop. If in each succession the detours are all compatible with the respective monotopy, then both original regular loops are monotopic to this common loop. Attempting to avoid detours incompatible with the normalizing monotopy complicates matters. Instead, observing that the arc of a loop that is replaced by a detour itself constitutes a detour of the new loop, (of opposite sign), a simple operator calculus on detours allows the reassembly of all the detours into two new successions, each now being compatible with the respective monotopies. Combining these two monotopies back to back yields the required folded monotopy.

The author wishes to express his appreciation to Professor Charles J. Titus of the University of Michigan for suggesting the possibility of the Folded Ribbon Theorem and for his encouragement during the discovery of its proof. I am particularly indebted to him for suggesting the trick that filled the last gap, the case of tangent winding number zero.

2. **Preliminaries.** For two real numbers $a<b$, let $[a, b]$ denote the closed interval from a to b. Let $\langle a, b\rangle$ denote the open interval, to distinguish this from the ordered pair (a, b). A continuous map

$$g\colon [a, b] \to R^2\colon x \to g(x)$$

from the interval to the oriented Euclidean plane will be called a *curve*. It is a *loop* if $g(a)=g(b)$. A *smooth curve* has a continuous first derivative $g'(x)$; for a *smooth loop*, also $g'(a)=g'(b)$. The curve is said to be *piecewise smooth* if it is smooth at all but a finite number of parameter values. At such a value x, the curve is required to have a left and a right derivative $g'(x-0)$ and $g'(x+0)$. A (piecewise) smooth curve is (*piecewise*) *regular* if the (left and right) derivative does not vanish anywhere. The continuous transverse vector field along g, obtained by rotating each tangent $g'(x)$ through ninety degrees clockwise, is denoted by $g^{\perp}(x)$. A regular loop is generally considered as an immersion of the based, oriented circle S^1, parametrized between 0 and 2π. As such, the interval $[0, 2\pi]$ will also be considered as the principal domain of a periodic, point valued function. This way, explicit discussion of the situation at the end points of this domain, corresponding equally to the base point of the circle, can be dispensed with.

The *image* $[g]$ of a (piecewise) regular curve g has a *node* at x if there is another parameter value $x^*\neq x$, so that $g^{-1}g(x)=\{x, x^*\}$, g is smooth at x and x^*, and the tangents $g'(x)$ and $g'(x^*)$ are independent. There is a *corner* at x if $g^{-1}g(x)=\{x\}$, and $g'(x-0)$, $g'(x+0)$ are independent. A piecewise regular curve is *piecewise normal* if it has a finite number of nodes and corners, and for the remaining points $g^{-1}g(x)=\{x\}$, $g'(x-0)=g'(x+0)=g'(x)$. The adjective "piecewise" is dropped in the absence of corners; the adjective "simple" is added in the absence of nodes. Occasional use will be made also of loops that map the base point of S^1 to a corner of $[g]$. On a few occasions, a normal curve $g[a, b]$ is permitted to start or end with a transverse self intersection, called a *tee*.

For various purposes, regular loops are endowed with a right transverse field $\hat{g}: S^1 \to R^2\backslash\{0\}$. The determinant, written $\det(\hat{g}(x), g'(x))$, is positive for all x. If no such field, called a *fringe*, is specified, the field $g^{\perp}$ can be used. The three data, g, $\hat{g}$, and g' are collected in a triple of column vectors, called a *frame*, written $\partial g = [g, \hat{g}, g']$.

Let the closed oriented disc D^1 be provided with polar coordinates (t, x); the radial parameter t ranges over the extended real line $[-\infty, +\infty]$. The cylinder $[a, b]\times S^1$ identifies with the subannulus of D^1, given by $a \leqq t \leqq b$. Because the Cartesian plane $R^2(t, x)$ covers the open punctured disc with periodicity 2π in x, maps from circles and annuli shall also be considered as maps originating in $R^2(t, x)$, periodic in the x parameter. The two principal vectors $\partial F/\partial t$ and $\partial F/\partial x$, of a map $F: R^2 \to R^2$, define the Jacobian determinant $J_F: R^2 \to R^1$ and the frame $\partial F = [F; \partial F/\partial t, \partial F/\partial x]$. On a closed domain, a differentiable function is assumed to be extendable to a small open neighborhood.

A *differentiable homotopy* between two regular loops g_i, $i = \pm 1$, is a C^1-map

$$G: [-1, +1]\times S^1 \to R^2: G(i, x) = g_i(x), \qquad i = \pm 1.$$

If G is C^0 in t, but C^1 in x, with $\partial G(t, x)/\partial x \neq 0$, then G is a *regular homotopy*, [10]. It is a positively oriented monotone differentiable homotopy, or *monotopy* for short, from g_{-1} to g_{+1} if

(M1) $J_G(t, x) > 0$, all t and x,

(M2) $\partial G(i, x) = \partial g_i(x)$ for all x, $i = \pm 1$.

A homotopy G is a *negative monotopy* from g_{-1} to g_{+1} if the map $G^*(t, x) = G(-t, x)$ is a positive monotopy from g_{+1} to g_{-1}. Observe that for a negative monotopy $\partial G(i, x)/\partial t = -\hat{g}_i(x)$, $i = \pm 1$. This asymmetry in the definition is prompted merely by a desire to keep the fringe of a regular loop to its right. By the inverse function theorem, monotopies are locally *univalent* (one-to-one) maps of an annulus. A positive monotopy is a special form of a differentiable increasing homotopy as defined by Titus in [7]. It is also an a-boundary as defined by Marx in [3]. A C^1-regular map, globally univalent on its domain, is an *embedding* (diffeomorphism).

The *tangent winding number*, TWN(g), of a regular loop g is the topological degree of the map

$$S^1 \to S^1: x \to g'(x)/|g'(x)|.$$

Because degree is a homotopy invariant, it follows that a regular homotopy preserves the TWN. The converse of this is the Whitney-Graustein Theorem [10].

This paper presents a constructive proof, independent of Graustein's, of a stronger converse:

THEOREM. *For two regular loops* g_i, $i = \pm 1$, *of like TWN, there always is a regular loop* g_0 *and two monotopies* H_i *from* g_i *to* g_0 *of like sign equal to* sgn $(TWN \pm \frac{1}{2})$. *The composition of the first with reverse of the second monotopy is called a folded monotopy.*

REMARK 1. Note that TWN$=0$ belongs to both cases. For TWN$=1$ two concentric circles are monotopic. Not so for two circles of disjoint interiors; yet each is monotopic to a circle surrounding both of them. It suffices to establish the theorem for TWN$\geqq 0$. For, let TWN$(g_i) \leqq 0$, set $h_i(x) = g_i(-x)$ and $\hat{h}_i(x) = -\hat{g}_i(-x)$. Then TWN$(h_i) = -TWN(g_i)$. If H_i are the two monotopies furnished for h_i, then $G_i(t, x) = H_i(t, -x)$ serve for the g_i.

REMARK 2. General position arguments and covering homotopy theory apply, under more generous differentiability conditions, to establish a monotopy between two regularly homotopic C^3-immersions of a finite dimensional compact, boundaryless C^∞-manifold into a C^∞-manifold of at least two dimensions higher. In codimension one, it follows under these circumstances, from the work of Poenaru [5] that a regular homotopy can be approximated by one that is a succession of oppositely oriented monotopies. Such a so-called pseudo-immersion of the cylinder over the source possesses an indefinite number of folds. Attempts to reduce the number of folds of a given pseudo-immersion abstractly have failed so far even in the simplest context dealt with here.

REMARK 3. The methods and constructions developed in this paper are meant to be quite elementary. The geometric constructions in §§3 and 6 could serve as an introduction to Poenaru [5]. Parts of §4 also extend to higher dimensions. §5 reviews and extends the combinatorial theory of normal immersions of the circle developed by Titus [6], [8]. No analogue of the intersection sequence for a normal immersion of a higher dimensional manifold is presently available. Results in this study have found application in Verhey [9] and Marx [3]. A simple application is found at the end of §7.

3. **Monotopies.** The goal of this section is to demonstrate the following

PROPOSITION 1. *A regular loop is monotopic to a normal loop arbitrarily close by.*

This is an application of Whitney's result [10], that the normal loops constitute a generic subclass in the space of regular loops. Since the domain of a regular loop is compact,

$$\|g_1 - g_0\| = \max_{x \in S^1} \{|g_1(x) - g_0(x)|, |g_1'(x) - g_0'(x)|, |\hat{g}_1(x) - \hat{g}_0(x)|\}$$

provides the space of C^1-regular loops with a topology. A subset that is both dense and open in this topology is called *generic*.

Define a regular loop g_1 to be (*right*) *parallel* to a regular loop g_0 if

$$\underset{x\in S^1, i=0,1}{\text{minimum}} \det(g_1(x)-g_0(x), g_i'(x)) > 0.$$

The loop g_1 is *left* parallel to g_0, if g_0 is right parallel to g_1. This condition will be shown to suffice for two loops to be monotopic. Clearly, it is not a necessary condition. The property is also C^1-stable, in the sense that a loop $\bar{g}_1$ sufficiently close to g_1 is still parallel to g_0.

The continuous fringe $\hat{g}_0$ will be used to thicken g_0 to a one-parameter family g_r of loops parallel to g_0, for r sufficiently small and positive. Because the normal loops are dense, there is a normal approximation to one of these nearby loops, which is still parallel (and so, monotopic) to g_0.

LEMMA 1. *If g_1 is right parallel to g_0, then the map*

$$H(t, x) = (1-t)g_0(x)+tg_1(x)$$

is a C^1-map with positive Jacobian for $0 \leqq t \leqq 1$.

Proof. $J_H(t, x)=(1-t)\det(g_1(x)-g_0(x), g_0'(x))+t\det(g_1(x)-g_0(x), g_1'(x))$. For a fixed x, $J_H(t, x)$ is a linear function in t between two positive values. ∎

Suppose $f\colon R^1 \to R^1$ is continuous. Consider the following one-parameter averaging operators on f:

$$A_tf(x) = A_f(t, x) = \frac{1}{2}\int_{x-t}^{x+t} f(s)\, ds,$$

$$M_tf(x) = M_f(t, x) = \tfrac{1}{2}(f(x+t)+f(x-t)),$$

$$\Delta_tf(x) = \Delta_f(t, x) = \tfrac{1}{2}(f(x+t)-f(x-t)).$$

Then A_f is continuously differentiable in both x and t with

$$\partial A_f/\partial t = M_f \quad\text{and}\quad \partial A_f/\partial x = \Delta_f.$$

If f is periodic (or has compact support) then for all t and x

$$\min_x f \leqq M_f \leqq \max_x f,$$

and if both t and $f(x)$ are nonnegative then

$$t\min_x f \leqq A_tf \leqq t\max_x f.$$

At $t=0$, $A_0f=0=\Delta_0f$ and $M_0f=f$, all x. Hence, by uniform continuity, $|\Delta_tf(x)|$ is uniformly small for small $|t|$.

Under the hypothesis of Lemma 1, H can be considered as a C^1-immersion of the vertical strip: $0\leqq t\leqq 1$, in the (t, x)-plane, periodic in x. Moreover, it is possible to lift the transverse fields $\hat{g}_i$, $i=0, 1$, to periodic vector fields w_i along the verticals $t=i$, $i=0, 1$, in the domain of H. If K now is a C^1-automorphism of this vertical strip (of the same period in x), with the property that $\partial K(i, x)/\partial t=w_i(x)$, then the composition $H \circ K$ is the required monotopy from g_0 to g_1. The problem of constructing a suitable K reduces further to the following:

LEMMA 2. *Let*

$$w(x) = \begin{bmatrix} u(x) \\ v(x) \end{bmatrix}$$

be a periodic vector field along and transverse to the ordinate axis $t=0$ in the (t, x)-plane. There is a C^1-map F: $[0, 1] \times R^1 \to [0, 1] \times R^1$ of the same period in x, with positive Jacobian, and satisfying the boundary conditions:

$$\partial F(0, x) = \begin{bmatrix} 0 & u(x) & 0 \\ x & v(x) & 1 \end{bmatrix} \quad \textit{and} \quad \partial F(1, x) = \begin{bmatrix} 1 & 1 & 0 \\ x & 0 & 1 \end{bmatrix}.$$

Proof. For a small and positive parameter m, let z: $[0, 1] \to [0, 1]$ be a C^2-smooth bump function with the properties: $z(0)=0$, $z'(0)=1$, increasing monotonically to its sole maximum at $z(\frac{1}{2})=m$, then decreasing monotonically through its only inflection point $z''(\frac{3}{4})=0$, until $z(1)=z'(1)=0$, and all the while from $\frac{1}{2}$ to 1,

$$-3m \leqq z'(t) \leqq 0.$$

The frame of the function F, given as follows

$$\partial F(t, x) = \begin{bmatrix} t-z(t)+A_{z(t)}u(x) & 1-z'(t)+z'(t)M_{z(t)}u(x) & \Delta_{z(t)}u(x) \\ x+A_{z(t)}v(x) & z'(t)M_{z(t)}v(x) & 1+\Delta_{z(t)}v(x) \end{bmatrix}$$

reduces at the extremes to

$$\partial F(0, x) = \begin{bmatrix} 0 & u(x) & 0 \\ x & v(x) & 1 \end{bmatrix} \quad \text{and} \quad \partial F(1, x) = \begin{bmatrix} 1 & 1 & 0 \\ x & 0 & 1 \end{bmatrix}.$$

From the transversality of the periodic field w it follows that $0 < u_\# = \min u(x) \leqq \max u(x) = u^\#$, and that there is a positive $v^\#$ with $|v(x)| \leqq v^\#$. Consider the entry $N(t, x) = 1 - z' + z'M_z u$ in the Jacobian of F. Over the first half interval of t, $z' \geqq 0$ and $N \geqq 1 - z' + z'u_\#$. This lower bound is, for $u_\# \neq 1$, a monotone function between its extremal values 1 and $u_\#$. Hence N is positive, bounded below by $\min\{1, u_\#\}$. (This bound also works for $u_\# = 1$.) Over the second half time interval, $-3m \leqq z' \leqq 0$, hence $N \geqq 1 - 0 - 3mu^\#$. Consequently, if the initial choice of m (and z) insured that $m < 1/6u^\#$, then over the entire time $N \geqq \min\{\frac{1}{2}, u_\#\}$.

The upper right entry of the Jacobian can be bounded $|\Delta_z u| < \min\{\frac{1}{2}, u_\#\}/9v^\#$ by restricting the initial choice of m, say by m_1. Because $|M_z v| \leqq v^\#$, this leads to the inequality

$$|z'M_z v \Delta_z u| \leqq \min\{\tfrac{1}{2}, u_\#\}/3.$$

Finally, pick some m_2 small enough, so that $1 + \Delta_z v \geqq \frac{1}{2}$. Collecting all three conditions, had m originally been chosen as $\min\{m_1, m_2, 1/6u^\#\}$, then

$$J_F = (1+\Delta_z v)N - z'M_z v\Delta_z u \geqq \min\{\tfrac{1}{2}, u_\#\}/6 > 0. \quad \blacksquare$$

LEMMA 3. *If g_1 is parallel to g_0 and loop $\bar{g}_i$ is sufficiently close to g_i, $i=0, 1$, then $\bar{g}_1$ is parallel to $\bar{g}_0$.*

Proof. Set $h_i=\bar{g}_i-g_i$. Computation shows that $\det(\bar{g}_1-\bar{g}_0, \bar{g}_i')$ expands to the sum of four determinants:

$$\det(g_1-g_0, g_i')+\det(g_1-g_0, h_i')+\det(h_1-h_0, g_i')+\det(h_1-h_0, h_i').$$

In each of the latter three, at least one factor is bounded in absolute value while its other factor goes to 0 with $\|h_i\| \to 0$. ∎

LEMMA 4. *The map $G(t, x)=g(x)+A_t\hat{g}(x)$ is regular with positive Jacobian for t sufficiently small, and $\partial G(0, x)=\partial g(x)$. Further, each loop $g_r(x)=G(r, x)$ is parallel to g_0, for r small enough and positive.*

Proof. $J_G=\det(M_t\hat{g}, g'+\Delta_t\hat{g})$. Let $\min\det(\hat{g}, g')=k>0$. Then $J_G(0, x)\geqq k$ for all x. Being continuous in t, $J_G(t, x)\geqq k/2$ for $|t|<r_1$, some appropriately small and positive r_1. Moreover, it is possible to specify an r_2 so that for $2|\xi-x|<r_2$,

$$\det(\hat{g}(\xi), g'(x)) > k/2.$$

Because

$$\det(A_t\hat{g}(x), g'(x)) = \frac{1}{2}\int_{x-t}^{x+t} \det(\hat{g}(\xi), g'(x))\, d\xi,$$

it is bounded below by $tk/2$, for $0<t\leqq\min\{r_1, r_2\}$. Next, let $|\Delta_t\hat{g}|<k/(3\max|\hat{g}|)$ for $|t|<r_3$.

Finally, for $0<t\leqq r=\min\{r_1, r_2, r_3\}$,

$$\begin{aligned}\det(g_t-g, g_t') &= \det(A_t\hat{g}, g'+\Delta_t\hat{g})\\ &\geqq \det(A_t\hat{g}, g')-|A_t\hat{g}|\,|\Delta_t\hat{g}|\\ &\geqq rk/2-r\max|\hat{g}|(k/3\max|\hat{g}|) = rk/6 > 0,\end{aligned}$$

and $\det(g_t-g, g')\geqq rk/2>0$. Consequently, g_t is parallel to g_0 for all $0<t\leqq r$. ∎

LEMMA 5. *Let $\mathscr{F}$ be a class of loops dense in the class of regular loops. Then for each regular loop g there is a loop $\bar{g}$ in $\mathscr{F}$ which is close and monotopic to g.*

Proof. As in Lemma 4, find $0<r$ sufficiently small so that $g_r=g+A_r\hat{g}$ is parallel to g and $\|g-g_r\|<e/2$. By Lemma 3 and the assumed density of $\mathscr{F}$, there is a loop $\bar{g}$ in so close to g_r that it is still parallel to g, and yet $\|\bar{g}-g\|<e$. ∎

Proof of Proposition 1. As already mentioned, the normal loops are dense among regular loops. Apply Lemma 5.

4. **Detours and modifications.** The purpose of this section is to make rigorous the visually intuitive idea of taking the normal loop $[g]$ in hand, and pulling a piece of it into a new and more favorable position in the target. If sufficient care is taken to move the piece consistently to the "right of itself", such a "deformation" defines a modification of any monotopy terminating with $[g]$, to one that terminates

with the new position. The essential part of the proof of the main theorem is a catalogue of such deformations that eliminate all but the absolutely necessary self-intersections. The proof of the theorem for such canonical loops is relatively easy.

Let $[a, b]$ be a parameter interval of a regular loop g on which it is univalent. A *simple detour* d, of g, with *support* $[a, b]$, is a C^1-diffeomorphism (with a continuous fringe $\hat{d}$ ad libitum) satisfying the following three properties:

(D1) The target set $[d] \cup g[a, b]$ is a closed Jordan loop.

(D2) $\partial d(a) = \partial g(a)$ and $\partial d(b) = \partial g(b)$.

(D3) There is assigned to the pair (d, g) a *signature*

$$\begin{aligned}\operatorname{sgn}(d, g) &= \lim_{x \to a+0} \operatorname{sgn}\det(d'(x), g'(x)),\\ &= \lim_{x \to b-0} \operatorname{sgn}\det(g'(x), d'(x)).\end{aligned}$$

PROPOSITION 2. *Let G: $[-1, 0] \times S^1 \to R^2$ be a positive monotopy with $\partial G(0, x) = \partial g(x)$, and d a positive detour of g. If dg represents the loop $dg(x) = d(x)$ for $x \in$ supp (d), and $dg(x) = g(x)$ for $x \in S^1 \backslash$supp (d), then there is a modified monotopy dG such that $\partial dG(0, x) = \partial dg(x)$.*

The idea of the proof is to replace G restricted to a diffeo-disc in the domain, on which G is univalent and whose boundary contains the support of the detour, by a diffeomorphism that now has $[d]$ on the boundary of its image. The principal step is a strong Schoenflies result that fills in a disc, while respecting the given boundary data up to first order.

LEMMA 6. *The class of real analytic regular loops is dense in the space of C^1-regular loops. The class of C^1-embeddings of the circle into the plane is open in the space of regular loops.*

Proof. Both of these are particular examples of well-known folk theorems. For primitive, self-contained proofs see author's thesis [1]. ▮

LEMMA 7. *Let f be a simple fringed loop. There is a C^1-diffeomorphism F: $D^1 \to R^2$ with $\partial F|$Bdy $D^1 = \partial f$.*

Proof. Direct application of the Riemann Mapping Theorem does not guarantee concordance with the given frame ∂f on the boundary. Without loss of generality, assume $[f]$ is oriented counterclockwise and D^1 has radius 3. Combining Lemma 5 with Lemma 6, there is a monotopy F_3: $[2, 3] \times S^1 \to R^2$ with $\partial F_3(3, x) = \partial f(x)$, and $h(x) = F_3(2, x)$ is a simple analytic loop fringed with $\partial F_3(2, x)/\partial t$. The Riemann Mapping Theorem provides for a univalent analytic map H from the disc of radius 2 to the Jordan domain enclosed by $[h]$. Since h is analytic, this C^1-diffeomorphism extends to one from the closed disc, albeit with a differing boundary loop k, where nevertheless $[k] = [h]$. But $k(x) = H(2, x)$ differs from h by a C^1-regular reparametrization $r: S^1 \to S^1$ with $H(2, r(x)) = h(x)$, and $r'(x) > 0$. Let $z(t) = (2-t)x + (t-1)r(x)$ for $1 \leqq t \leqq 2$. Then the map $\bar{H}(t, x) = H(t, z(t, x))$ has Jacobian

$$J_{\bar{H}} = \det\left(\frac{\partial H}{\partial t} + \frac{\partial H}{\partial x}\frac{\partial z}{\partial t}, \frac{\partial H}{\partial x}\frac{\partial z}{\partial x}\right) = \frac{\partial z}{\partial t} J_H.$$

But $\partial z/\partial x = (2-t) + (t-1)r'(x)$, which is, for each x, a linear function between two positive values. The domain of x is compact, and therefore $\bar{H}$ has positive Jacobian.

Thus $\bar{H}$, as a map of the annulus, can also be considered as a C^1-immersion of the vertical strip $1 \leqq t \leqq 2$ in the (t, x)-plane, periodic in x. By the argument preceding Lemma 2, it is possible to modify $\bar{H}$ to $F_2: [1, 2] \times S^1 \to R^2$, so that $\partial F_2(1, x) = \partial H(1, x)$, and $\partial F_2(2, x) = \partial h(x)$. Set $F_1 = H|$unit disc, and define $F(t, x) = F_i(t, x)$ for $i-1 \leqq t \leqq i$, $i = 1, 2, 3$. It remains to see that F is globally univalent on the disc. Because $J_F > 0$, the winding number of $[f]$ about any point Q in the interior of $[F]$ satisfies the following chain of equalities:

$$\begin{aligned} \text{Cardinality } F^{-1}(Q) &= \sum_{F(P)=Q} \operatorname{sgn} J_F(P) \\ &= \text{Degree } (F, Q) \\ &= \text{degree}\left(S^1 \to S^1 : x \to \frac{f(x)-Q}{|f(x)-Q|}\right) \\ &= \text{TWN}(f) \\ &= +1. \end{aligned}$$

The first and fifth equality hold by hypothesis. The second is one of several equivalent definitions of topological degree, the third is a standard theorem in degree theory, and the fourth is Hopf's classical result [2]. ∎

Proof of Proposition 2. In the service of simplicity, assume in the hypothesis of the proposition that the support of the detour is $[a, b]$ with $0 < a < b < 2\pi$. Let G be considered an immersion of the vertical strip in the Cartesian source plane. See Figure 1. Let the unit disc D^1, with bounding circle S^1, be parametrized with polar coordinates (r, s). Since g is presumed univalent on $[a, b]$, G is also univalent on the rectangular piece $[-e, 0] \times [a-e, b+e]$ for e chosen sufficiently small and positive. Inside this region construct a simple smooth arc C, connecting points $(0, a)$ to $(0, b)$ such that C together with supp (d) enclose the Jordan domain Q smoothly. Let $\gamma: S^1 \to$ Bdy Q be a regular parametrization with frame

$$\partial\gamma(s) = \begin{bmatrix} 0 & 1 & 0 \\ s & 0 & 1 \end{bmatrix} \quad \text{for } a \leqq s \leqq b,$$

and let $\gamma(s)$ occupy C for the remaining values of s. Extend the fringe already specified on $[a, b]$ to $\hat{\gamma}$ on all of $[\gamma]$. By Lemma 7, γ extends to a diffeomorphism $\Gamma: D^1 \to Q$.

On the other hand, the simple loop in the target, consisting in $[d]$ together with $G(C)$, can be parametrized by a map $\eta: S^1 \to R^2$, with $\partial\eta(s) = \partial d(s)$ for $s \in$ supp (d),

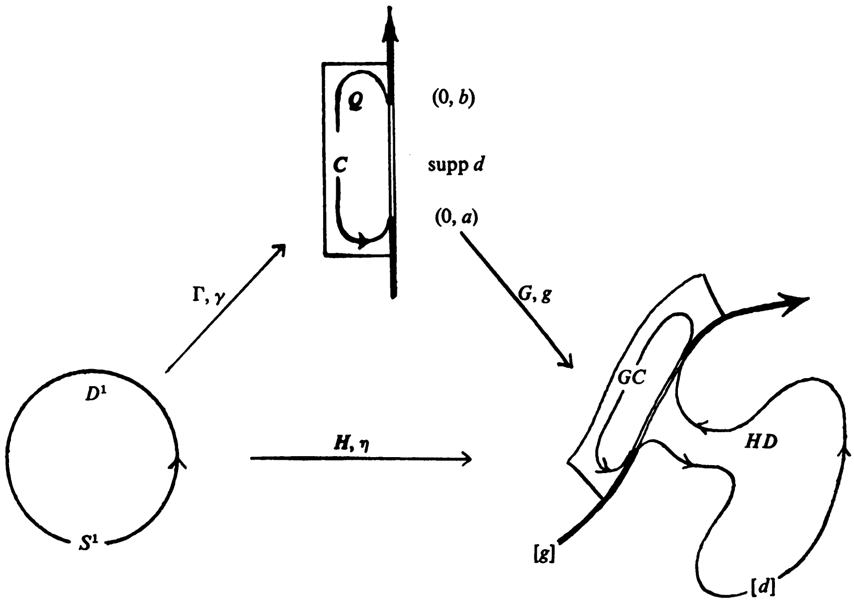


FIGURE 1

and $\partial\eta(s)=\partial(G\circ\gamma)(s)$ otherwise. Again by Lemma 7, η extends to the diffeomorphism H from the disc to the Jordan domain enclosed by $[\eta]$. Finally, setting $dG(t,x)=H\circ\Gamma^{-1}(t,x)$ on Q, and $dG(t,x)=G(t,x)$ otherwise, completes the demonstration. ▮

In practice, only *piecewise normal detours* will be explicitly constructed. Careful smoothing of the corners (so as not to alter the character of the detoured loop) will henceforth be self-understood. In the definition of a piecewise normal detour, only (D2) must be altered so as to allow for tees to happen at $g(a)$ and $g(b)$:

(D′2) $$d(a) = g(a), \qquad d(b) = g(b).$$

Of course (D3) now becomes more manageable if it is restated at the limits as

(D′3) $$0 \neq \operatorname{sgn}\det(d'(a), g'(a)) = \operatorname{sgn}\det(g'(b), d'(b)) = \operatorname{sgn}(d, g).$$

Compound detours. Suppose d_1 is a simple detour of g, and d_2 is a simple detour of d_1g. From all the possibilities, only three will be of interest in this paper.

(C+) The supports of the two detours are disjoint.

In this case the *sum* d_1+d_2 is said to detour g, with $(d_1+d_2)g=d_2(d_1g)$ and $\operatorname{supp}(d_1+d_2)\doteq\operatorname{supp}(d_1)\cup\operatorname{supp}(d_2)$. Additive notation for this case, serves to emphasize the commutativity of detours with disjoint supports.

(C×) Either $\operatorname{supp}(d_2)\subset\operatorname{supp}(d_1)$ or $\operatorname{supp}(d_2)\supset\operatorname{supp}(d_1)$, and $\operatorname{sgn}(d_2, d_1g)=\operatorname{sgn}(d_1, g)$.

In these two cases, the *product* d_2d_1 acts on g, with support the larger of the two supports. Of course, if the second alternative holds, $(d_2d_1)g=d_2g$.

One last formalism is designed to abbreviate "g restricted to supp (d)". Denote by d^* the detour of dg that $g|\text{supp}\,(d)$ constitutes. Evidently supp $(d^*)=$ supp (d), $[g^*]=g(\text{supp}\, d)$, while sgn $(d^*)=-$sgn (d). The detour d^* acts as an *inverse* of d, in the sense that $(d^*d)g=g$ itself.

A formal expression $\mathfrak{D}$, made up of a finite number of permissible sums and products of simple detours is called a *monotone compound detour* if all component detours have the same sign. If these are of mixed sign, the compound is *mixed.*

PROPOSITION 3. *Suppose* $G_{-1}\colon [-1, 0]\times S^1 \to R^2$ *is a monotopy from* g_{-1} *to* h, $\mathfrak{D}$ *is a compound detour of* h *with* $\mathfrak{D}h=k$, *and that* $G_{+1}\colon [0, +1]\times S^1 \to R^2$ *is a monotopy from* k *to* g_{+1}.

(A) *If* $\mathfrak{D}$ *is monotone and all three signs are the same, then* g_{-1} *is monotopic to* g_{+1}.

(B) *If* $\mathfrak{D}$ *is mixed and the monotopies are of unlike sign, then* g_{-1} *is fold-monotopic to* g_{+1}.

Proof. (A) Repeated application of Proposition 2 (in "parallel" for sums, in "series" for products) gives a modified monotopy $\mathfrak{D}G_{-1}$ from g_{-1} to k. But $\partial\mathfrak{D}G_{-1}(0, x)=\partial k(x)=\partial G_{+1}(0, x)$. Therefore

$$G\colon [-1, +1]\times S^1 \to R^2\colon G(t, x) = \mathfrak{D}G_{-1}(t, x) \quad \text{for } -1 \leqq t \leqq 0,$$
$$= G_{+1}(t, x) \quad \text{for } 0 \leqq t \leqq +1,$$

serves as monotopy from g_{-1} to g_{+1}.

(B) Without loss of generality, assume G_{-1} is a positive monotopy. Then G_{+1} is a negative monotopy. By definition, then, the positive monotopy $G^*(t, x) = G_{+1}(1-t, x)$ takes g_{+1} to k. The legitimacy of the compound allows the reassociation of $\mathfrak{D}$ into the sum of two monotone compounds, $\mathfrak{D}^+ + \mathfrak{D}^-$, one positive, the other negative. That $\mathfrak{D}^+G_{-1}$ is a positive monotopy from g_{-1} to $g_0=\mathfrak{D}^+h$ is clear. Now $\mathfrak{D}^-g_0=k$, because $\mathfrak{D}=\mathfrak{D}^+ + \mathfrak{D}^-$. If $\mathfrak{D}^*$ represents the compound obtained by inverting each component of $\mathfrak{D}^-$, then it follows that $\mathfrak{D}^*k=g_0$. Therefore $\mathfrak{D}^*G^*$ is a positive monotopy from g_{+1} to g_0. Consequently,

$$(\mathfrak{D}^*G^*)^*(t, x) = \mathfrak{D}^*G^*(1-t, x)$$

is a negative monotopy from g_0 to g_{+1}. The required fold-monotopy is given by $G(t, x)=\mathfrak{D}^+G_{-1}(t, x)$ for $-1 \leqq t \leqq 0$, and $G(t, x)=(\mathfrak{D}^*G^*)^*(t, x)$ for $0 \leqq t \leqq +1$. ▌

5. **The intersection sequence.** In this section the combinatorial description of a normal loop is developed. Recall that a normal loop g is a smooth immersion of the based, oriented circle into the plane, having a finite number of simple, transverse self-intersections, called *nodes*. It is often more convenient to regard g as an immersion of the interval $[0, 2\pi]$ with $\partial g(0)=\partial g(2\pi)$. The starting point $g(0)=g(2\pi)$, is most conveniently chosen on the boundary of the unbounded complementary component of $[g]$. The start is counted among the nodes as an

honorary initial member of the intersection sequence. With the obvious modifications, the combinatorial descriptions given below adapts equally well to a piecewise normal loop having a finite number of corners. Such a piecewise normal loop will happen naturally, in this paper, as a subloop (the starting point is a corner), or as a detoured loop (the detour has a finite number of corners).

The *intersection sequence* $W(g)$, of a (piecewise) normal loop g is a finite, totally ordered set of *indices*, say $\{0, 1, 2, \ldots, n\}$, assigned to the nodes of $[g]$ as follows:

(W1) The starting point of g is denoted by N_0.

(W2) Following the orientation of $[g]$, N_1 is the first node encountered.

(W3) The remaining nodes are numbered consecutively, skipping those already assigned a number.

To each node N_k, there are two parameter values $0 \leqq x_k' < x_k'' \leqq 2\pi$. The kth *subloop* is the map $g_k = g|[x_k', x_k'']$. This subloop inherits its parametrization from g. Evidently $g_0 = g$ itself. The order on the indices of $W(g)$ is just the numerical order of the left end points of the domain intervals of the subloops.

Suppose $0 \leqq i < j \leqq n$ in $W(g)$. Already, $x_i' < x_j' < x_j''$ and $x_i' < x_i''$. So there are precisely three possible locations for x_i'', which determine one of three exclusive relations between i and j:

($i \supset j$) read *i contains j*, if $x_i' < x_j' < x_j'' < x_i''$,

($i \mid j$) read *i disjoins j*, if $x_i' < x_i'' < x_j' < x_j''$,

(iLj) read *i left links j*, if $x_i' < x_j' < x_i'' < x_j''$.

It follows that $0 \supset j$, for all $j > 0$. Moreover, if $W_j = \{i \mid j \supseteq i\}$, then W_j is the intersection sequence $W(g_j)$. $W_0 = W(g)$ itself. By abuse of language, relations predicated of indices will also be predicated of the corresponding subloops and nodes.

An intersection sequence $W(g)$ is said to be *properly nested* if the relation (iLj) never obtains in W. A subloop g_j is *internally linked* if W_j is not properly nested. It is *externally linked*, if j links in W, but for no proper subloop k of j, $j \supset k$, is k linked in W. A properly nested intersection sequence W, can be represented by a simply connected plane graph (*tree graph*) under the transitive relation $\supset$. Conventionally, 0 is set at the top and the indices are read from left to right. The *depth* of an index i in a properly nested W is the number of segments in the unique graph-path from i to the initial index 0. If both relations L and | are null in W, W is a *chain*, and the relation $\supset$ is here redundant with $<$. See the example in §7.

Directing attention to the topology of the image $[g]$, a subloop g_k is said to be *exterior* if

(X1) for $k=0$, the starting point $g(0) = N_0$ lies on the closure of the unbounded complementary component of $[g]$, Clos $C_\infty[g]$,

(X2) for $k \neq 0$, the points $g(x_k' - \xi)$ and $g(x_k'' + \xi)$ lie in $C_\infty[g_k]$ for all ξ sufficiently small and positive.

A loop satisfying (X1) is traditionally said to *start outside*.

There is a canonical properly nested subsequence of *essential indices*, EW of W.

Its members are discovered as follows:

(E1) The initial index 0 is essential.

(E2) If i_0 is essential, then $i_1 = \min\{j \mid i_0 \supset j\}$ is essential;

(E3) and $i_{k+1} = \min\{j \mid i_0 \supset j \text{ and } i_k \mid j\}$ is essential, for $k = 1, 2, 3, \ldots$, however many such so called *principal* subindices of i_0 there are *under* i_0.

Observe that for q essential, $EW_q = W_q \cap EW$. The *reduced subloop* corresponding to i_0 is the closed Jordan loop(3)

$$[g/i_0] = g[x'_{i(0)}, x'_{i(1)}] \cup \bigcup_{k=1}^{r-1} g[x''_{i(k)}, x'_{i(k+1)}] \cup g[x''_{i(r)}, x''_{i(0)}].$$

The chief properties of EW are collected in the following

PROPOSITION 4. *If W is the intersection sequence of a normal loop g then*

(A) *EW is properly nested.*

(B) *If W is already properly nested, the $EW = W$.*

(C) *If g starts outside, then every essential index is also exterior.*

(D′) *If g starts outside and W is not properly nested, there is at least one essential subloop i_0 that is externally linked,*

(D″) *any such linking of i_0 occurs on the reduced subloop $[g/i_0]$,*

(D‴) *and if i_0 is linked once, then it is linked at least twice.*

Proofs. (A) Let $i<j$ both be essential and $q_0 = \max\{k \mid k \text{ essential}, k \supseteq i, k \supseteq j\}$. If $q_0 = i$, then $i \supset j$. Otherwise there are two principal indices $q_r | q_s$ under q_0 so that $q_r \supseteq i$ and $q_s \supseteq j$. But in that case, $x''_i < x''_{q(r)} < x'_{q(s)} < x'_j$, whence $i | j$.

(B) Let W be properly nested and j in W. If $j=0$ the j is essential. Otherwise, let $i_0 = \max\{k \mid k \text{ essential and } k \supset j\}$. Let $i_1, \ldots, i_r$ be the principals under i_0. From $i_0 \supset j$ follows $i_1 \leqq j$. If $i_1 = j$, j is essential. Otherwise, there is a k with $i_k < j \leqq i_{k+1}$. From $i_k \not\supset j$ follows $i_k | j$. But then $j = i_{k+1}$ and j is essential.

(C) Assume q_0 is essential, exterior by inductive hypothesis and q_k is principal under q_0. Consider the initial subarc of $[g/q_0]$

$$L_k = g[x'_{q(0)}, x'_{q(1)}] \cup g[x''_{q(1)}, x'_{q(2)}] \cup \cdots \cup g[x''_{q(k-1)}, x'_{q(k)}].$$

From $q_0 \supset q_k$, and q_0 exterior, it follows that $N_{q(0)} \in C_\infty[g_{q(k)}]$. Unless q_k is also exterior, L_k must leave the unbounded, and enter a bounded complementary component of $[g_{q(k)}]$, crossing at, say, N_j. Then $x'_{q(k)} < x''_j < x''_{q(k)}$ and either $x'_{q(0)} < x'_j \leqq x'_{q(1)}$ or $x''_{q(s)} < x'_j \leqq x'_{q(s+1)}$. From this follows that $q_0 \supset j$, $j\mathrm{L}q_k$ and either $j = q_1$ or $j = q_{s+1}$, which cannot be by the definition of the essential indices.

(D′) Suppose $EW \neq W$ and j is not essential. Set

$$i_0 = \max\{k \mid k \text{ essential and } k \supset j\}.$$

There is an s with $i_s < j < i_{s+1}$. The choice of i_0 precludes $i_s \supset j$, the choice of i_{s+1} precludes $i_s | j$, hence $i_s \mathrm{L} j$, and some essential index links.

(3) Here and subsequently, the typographical convention for double subscripts is used: $x'_{i(r)}$ instead of x'_{i_r}.

(D″) Next, suppose q is essential, linked and of maximal depth in EW. It follows that no proper essential subloop of g_q is linked in W. A fortiori, no loop in EW_q is linked in W_q. Applying (D′), $EW_q = W_q$, and so no proper subloop of g_q is linked in W. So any linking of q in W occurs at points of $[g/q]$.

(D‴) Suppose N_j is such a linking node on $[g/q]$. Because g_q is exterior, it is possible to fix an $e>0$ so that the points $A=g(x'_q-e)$ and $B=g(x''_q+e)$ both lie in $C_\infty[g_q]$. Let x_j be the parameter value of N_j with $x'_q<x_j<x''_q$, and x^*_j its partner. From $N_j \in [g/q]$ follows that exactly one of the two points $C=g(x^*_j+e')$ and $D=g(x^*_j-e')$, e' suitably small, lies in $C_\infty[g_q]$. Removing $g\langle x'_q-e, x''_q+e\rangle \cup g\langle x^*_j-e', x^*_j+e'\rangle$ from $[g]$ leaves two immersed arcs, one connecting A with either C or D, the second connecting B with the remaining point. One of these arcs must cross $[g_q]$ on its way from the unbounded to some bounded complementary component of the subloop. Say this happens at N_k. Since one of the parameters of N_k is off $[g_q]$ and the other is on $[g_q]$, either $k\mathrm{L}q$ or $q\mathrm{L}k$. ▌

To complete the discussion of the intersection sequence W of a normal loop g that starts outside, *signs* are attached to each index as follows:

(S0) sgn (0) = ± provided the point $g(0) \pm tg^{\perp}(0)$ lies in $C_\infty[g]$ for all sufficiently small t positive.

(Sk) sgn (k) = sgn det $(g'(x''_k), g'(x'_k))$.

A properly nested intersection sequence is *precanonical* if either $W=\{0, 1\}$ and sgn (0) = −sgn (1), or all indices of W have the same sign. A precanonical sequence is *canonical* if it is also chained. The topological equivalence class of a canonical normal loop corresponds in a one-to-one way with the integer representing the tangent winding number of the class. See Figure 2.

6. **The normal tubular neighborhood.** The device of the normal tubular neighborhood facilitates the construction of those detours designed to reduce a normal loop g to the canonical loop for TWN(g). Suppose the parameter values $0=z_0<z_1<z_2<\cdots<z_r=2\pi$ are significant for g, in particular the parameter values of all nodes (and corners) of g are among them. A small positive number e *insulates* g, provided that $e<(1/100)\min|z_{k+1}-z_k|$. Then, for example, g is univalent on any interval of length less than $2e$.

Until the remark at the end of this section, g shall be assumed to be smoothly normal. A *tubular neighborhood* (tube) of g is a C^1-immersion $T: [-e, +e]\times S^1 \to R^2$ with $T(0, x)=g(x)$, where e insulates g. In a sense, it is preferable to consider

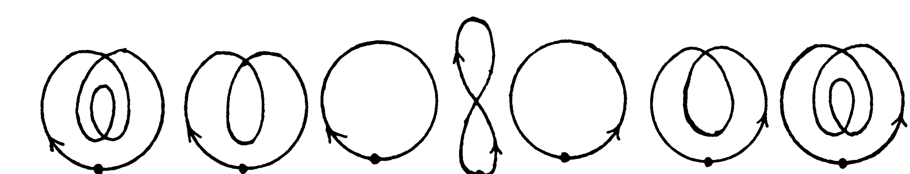


FIGURE 2

a tube (T, e) as a one-parameter family of immersions of $[-e', +e'] \times S^1$, over all $0 < e' \leqq e$. The effect of letting $e' \to 0$ is to shrink the tube so close to $[g]$ as to minimize extraneous self-intersections of the image $[T]$. The additional properties of a *normal* tube provide for the nicest possible self-intersection of $[T]$ near the nodes of $[g]$. Let $\square_i(e)$ denote the square source region $[-e, +e] \times [x_i - e, x_i + e]$, where the ordinate of the (t, x)-plane carries the parameter of g; the values $0 < x_1 < x_2 < \cdots < x_{2n} < 2\pi$ are the parameters of the n nodes of $[g]$. (The reader is warned not to confuse x_i' or x_i'' with x_i, the last is the ith point corresponding to a node, and not necessarily the ith node.) The complement $J(e) = [-e, +e] \times S^1 \setminus \bigcup_i \square_i$ should be thought of as the union of thick strips, one of which contains the base point of the circle in its interior.

DEFINITION. A *normal tubular neighborhood* (T, e) is a C^1-immersion

$$T: [-e, +e] \times S^1 \to R^2$$

having the following three properties:

(T1) For all x, $T(0, x) = g(x)$. Moreover $\partial T(0, x)/\partial t = \hat{g}(x)$ for all, except possibly those x e-close to an x_i.

(T2) If $g(x_i) = g(x_j)$ is a node and r_{ij} is the rigid motion

$$r_{ij}: \square_j \to \square_i: (t, x_j + s) \to (\pm_{ji} s, x_j \pm_{ij} t),$$
$$\pm_{ij} = \operatorname{sgn} \det (g'(x_i), g'(x_j)),$$

then $T_j = T|\square_j = T_i \circ r_{ij}$.

(T3) Furthermore, T shall be *unipotent*(4) on $J(e)$ and *bipotent* on $\square_i$ as specified by (T2). See Figure 3.

Thus T mimics the normal loop g in most of its essential features.

(1) T is an embedding of $J(e)$.

(2) $T\square_i$ and $T\square_j$ are identical or disjoint, according to whether $g(x_i) = g(x_j)$ or not.

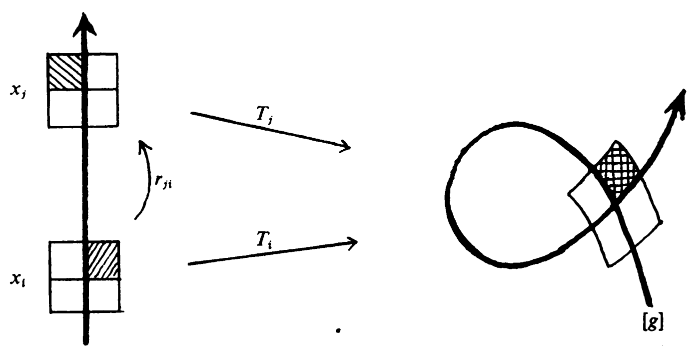


FIGURE 3

(4) Recall that a map $f: X \to Y$ is *univalent* on a subset $Z \subset X$ if Cardinality $f^{-1}(z) \cap Z = 1$ for every $z \in Z$. It is *unipotent* on Z if Cardinality $f^{-1}(z) = 1$ for all $z \in Z$.

(3) T is univalent on any strip $[-e, +e]\times[x-e, x+e]$.

(4) The tube restricted to $[-e', +e']\times S^1$ is also normal for all $0<e'\leqq e$.

(5) $TJ(e)$ is disjoint from every $T\square_i(e)$.

(6) On $T\square_i$, the arc $T[-e, +e]\times\{x_i\}=g[x_i^*-e, x_i^*+e]$.

The construction of a normal tubular neighborhood proceeds by a process of welding similar to the usual C^∞-welding [3], but somewhat simpler for the C^1 case.

LEMMA 8. *Let (T_i, e_i), $i=0, 1$, be two tubes of g, and $[a_0, a_1]$ a parameter interval for g. There is a transition tube (T, e), $e\leqq\min\{e_i\}$, such that for $x\leqq a_0$, $T=T_0$; for $a_1\leqq x$, $T=T_1$.*

Such a welding of tubes is said to occur on $[a_0, a_1]$.

Proof. (A) The proof is simpler under the additional hypothesis that

$$\partial T_0(0, x) = \partial g(x) = \partial T_1(0, x).$$

Let $z\colon [a_0, a_1]\to[0, 1]$ be a C^1-splice function, with $\partial z(a_i)=[i; 0]$, $i=0, 1$, and for $a_0<x<a_1$, $0<z<1$ and $0<z'$. On the interval in question, set $T(t, x)=(1-z(x))\cdot T_0(t, x)+z(x)T_1(t, x)$. Because $\partial T=(1-z)\partial T_0+z\partial T_1+[0; 0, (T_1-T_0)z']$, it appears that $\partial T(0, x)=\partial g(x)$, hence $J_T(0, x)>0$. By continuity, T is an immersion for $|t|\leqq e$, e sufficiently small and positive. At the endpoints T matches up with T_0 and T_1, i.e. $\partial T(t, a_i)=\partial T_i(t, a_i)$, $i=0, 1$.

(B) Suppose g has two different fringes f_i, $i=0, 1$, on $[a_0, a_1]$. Make $f=f_0$ for $x\leqq a_0$ and $f=f_1$ for $a_1\leqq x$, while on $[a_0, a_1]$, $f(x)=(1-z(x))f_0(x)+z(x)f_1(x)$. If $m=\min\det(f_i, g')$, $i=0, 1$, $x\in[a_0, a_1]$, then m is positive. Computation reveals that $\det(f, g')\geqq(1-z)m+zm=m>0$ for $a_0<x<a_1$. But at the endpoints,

$$\det(f(a_i), g'(a_i)) = \det(f_i(a_i), g'(a_i)),$$

both being positive. Hence, f is also a fringe of g, *fusing* f_0 to f_1 on the interval $[a_0, a_1]$.

(C) To complete the proof of the lemma, trisect the interval, $a_0<b<c<a_1$. Set fringe $f_i(x)=\partial T_i(0, x)/\partial t$. Fuse f_0 to f_1 on $[b, c]$ as in (B). Construct the tube

$$T'(t, x) = g(x)+\frac{1}{2}\int_{x-t}^{x+t} f(s)\, ds,$$

valid for $|t|\leqq e'$, some e'. Weld T_0 to T' on $[a_0, b]$ and T' to T_1 on $[c, a_1]$ by means of (A). The entire construction holds for some $e\leqq\min\{e_0, e_1, e'\}$. ∎

The canonical thickening of g given by

$$S(t, x) = g(x)+\frac{1}{2}\int_{x-t}^{x+t} \hat{g}(s)\, ds,$$

gives a tubular neighborhood satisfying (T1) at least (Lemma 4). At the nodes, S must be considerably modified to assure (T2). This modification is welded into S at each node.

LEMMA 9. *Let $x_i \neq x_j$, $g(x_i)=g(x_j)$ be the node N of* $[g]$.

(A) *The map* $T_j\colon \square_j(e) \to R^2$; $(t, x_j+s) \to g(x_j+s)+g(x_i \pm_{ij} t)-N$ *is a* C^1-*embedding for e sufficiently little.*

(B) *Let such a* T_j *be constructed for every* x_j, *choosing a common e. Then for e sufficiently small,* $[T_i]$ *and* $[T_j]$ *are identical or disjoint depending on whether* $g(x_i) = g(x_j)$ *or not.*

(C) *In the former case* T_j *is related to* T_i *by the rigid motion* r_{ij}.

Proof. For the first assertion, observe that the Jacobian of T_j reduces to $\pm_{ij}$ det $(g'(x_i), g'(x_j))$ at N. This quantity is positive. The inverse function theorem applies. The second assertion is clear. Assertion (C) is a matter of computation. Suffice it to point out that $-(\pm_{ij}) = \pm_{ji}$. ∎

If T_i is welded to S near and on either side of x_i, it is clear that (T2) will also be true. Specifically, choose $e_0>0$, small enough to insulate $[g]$. Next, choose $e_1 \leqq e_0$ to work for S; next choose $e_2 \leqq e_1$ to serve each T_i, welding S to T_i on the intervals $[x_i-e_2, x_i-\frac{1}{2}e_2]$ and $[x_i+\frac{1}{2}e_2, x_i+e_2]$. These welds will hold for some $e_3 \leqq e_2$. Now (T2) will still hold for e_4 chosen smaller than min $\{\frac{1}{2}e_2, e_3\}$. For (T3), it remains to trim the width of the tube (T, e_4) so constructed. As $e \to 0$, (T, e) does not cease to satisfy (T1) and (T2). Being an immersion, there is a point $e_5 < e_4$ beyond which T is univalent on any strip of length and width less than $2e_5$.

LEMMA 10. *If* (T, e) *satisfies* (T1) *and* (T2) *and is univalent on any* $[-e, +e] \times [x-e, x+e]$, *then there is* $0<e' \leqq e$ *so that* (T, e') *also satisfies* (T3).

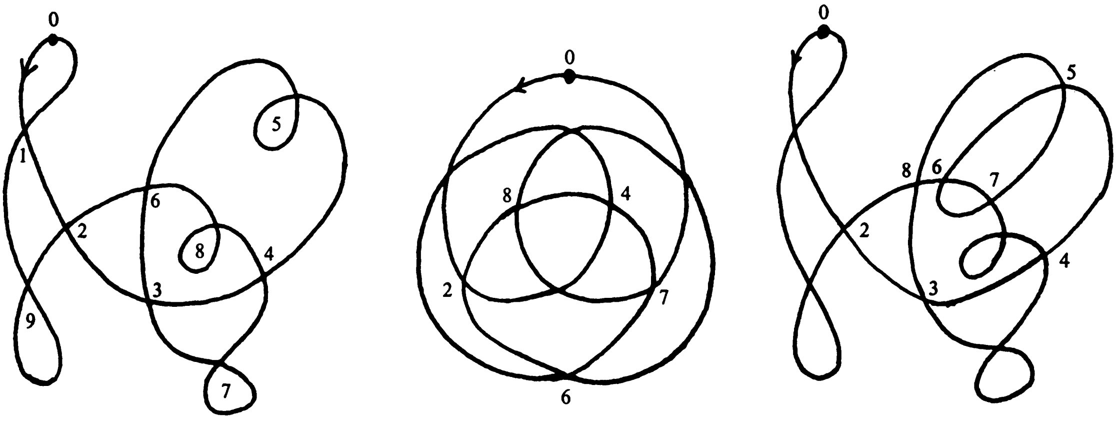
Proof. Suppose $e \geqq e_n \to 0$, so that there is a pair of source points $Y_n' \neq Y_n''$ in $[-e_n, +e_n] \times S^1$ with common image $T(Y_n')=T(Y_n'')=Z_n$. In the compact target $T[-e, +e] \times S^1$ a subsequence of $\{Z_n\}$ converges to Z_0, whence a subsequence of $\{Y_n'\}$ converges to Y_0', and one of $\{Y_n''\}$ to Y_0'', so that $T(Y_0')=T(Y_0'')=Z_0$. Because $e_n \to 0$, both Y_0' and Y_0'' are on the ordinate, so Z_0 lies on $[g]$. Then either $Y_0'=Y_0'' =(0, y_0)$ or Z_0 is a node of $[g]$. In the former case, there is a pair $Y_n' \neq Y_n''$ in $[-e, +e] \times [y_0-e, y_0+e]$. On this strip, however, T is univalent. Hence the second case obtains. Here, $Y_0'=(0, x_i)$, $Y_0''=(0, x_j)$, $g(x_i)=g(x_j)=Z_0$. So there is a pair with Y_n' in $\square_i(e)$ and Y_n'' in $\square_j(e)$. Unless $r_{ji}(Y_n')=Y_n''$, there would be a second distinct point, namely the point $r_{ji}(Y_n')$ in $\square_j(e)$, having the same image as Y_n''. Again, by univalence, this cannot be. ∎

REMARK. Close inspection of the method in Lemma 9 leads to an analogous construction of a normal tubular neighborhood for a piecewise normal immersion. Such a tube makes a unipotent corner at each of the corners of the loop.

7. **Reduction to canonical form.** This section is devoted to the demonstration of the following

PROPOSITION 5. *Let g be a normal loop starting outside, with* $TWN(g) \geqq 0$.

(A) *There is a mixed sum of simple detours* $\mathfrak{U}$ *so that* $W(\mathfrak{U}g)=EW(g)$.



α α* α**

FIGURE 4

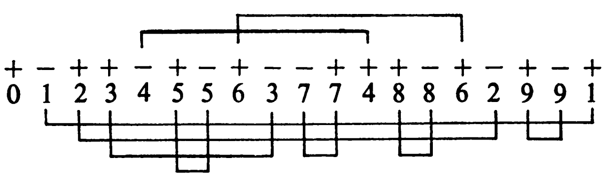
(B) *There exists further, a mixed sum of simple detours* $\mathfrak{A}$, *whose supports are disjoint from the support of* $\mathfrak{U}$, *so that* $(\mathfrak{U}+\mathfrak{A})g$ *is precanonical.*

(C) *Unless* $TWN=0$ *or* 1, *in which case* $(\mathfrak{U}+\mathfrak{A})g$ *is already canonical, there is a monotone sum* $\mathfrak{B}$, *of simple, positive detours, with support disjoint from those of* $\mathfrak{U}$ *and* $\mathfrak{A}$, *so that* $(\mathfrak{U}+\mathfrak{A}+\mathfrak{B})g$ *is canonical.*

EXAMPLE. A simple, but nontrivial example of the methods in this section is given below. The reader is referred to the subsequent text for the details.

Consider the loop α, drawn in Figure 4.

The *word* is obtained by writing the parameter values $\{x_0'=0, x_i', x_i'', x_9''\}$ in their natural order, placing sgn (i) above the symbol x_i', and $-\text{sgn}\,(i)$ above x_i''. Now delete all but the indices and the signs, drawing the display



Thus the essential subsequence of α is given by the first tree graph in Figure 5.

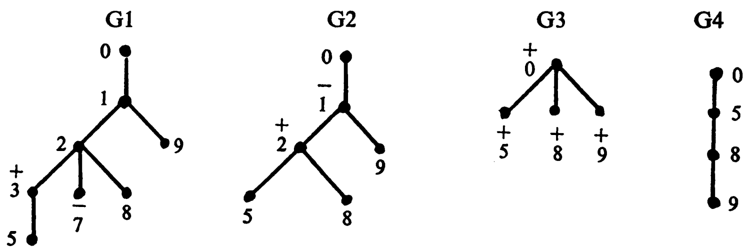
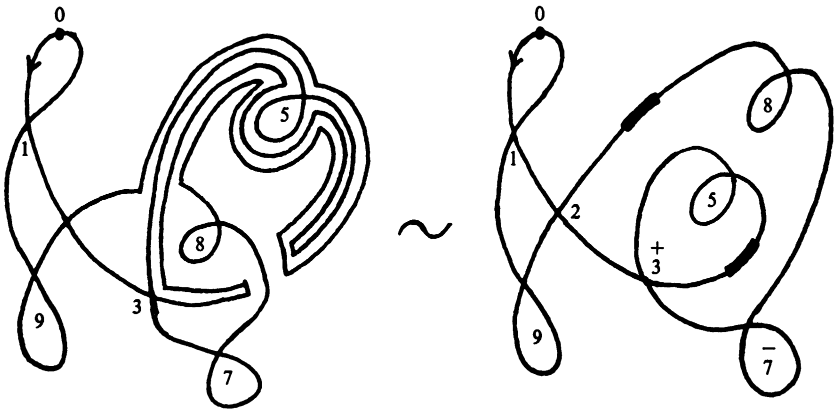


FIGURE 5



β γ

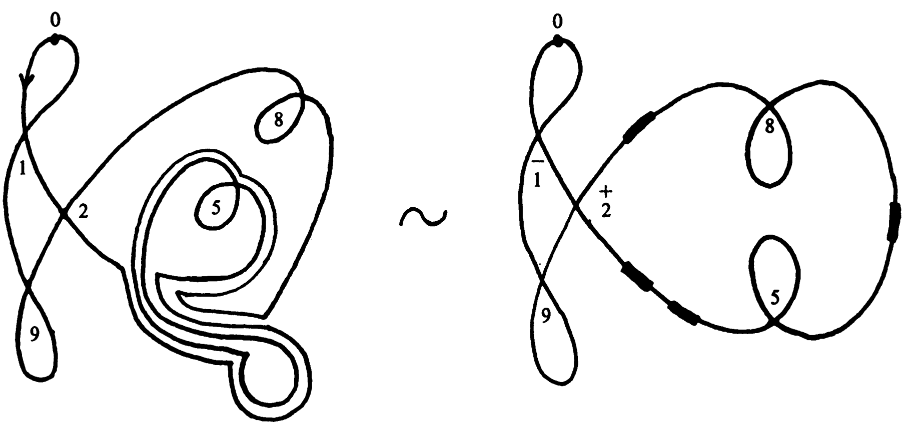
FIGURE 6

In $W(\alpha)$, while 2 has interior linking at 4 and 6, 3 is exteriorily linked. The unlinking detours of $\mathfrak{U}$ are indicated in loop β of Figure 6 as they will actually be constructed. Loop γ to its right, is topologically equivalent to β; heavy traces mark the detours. This topological transformation is presented for visual purposes only, in fact; subsequent detours are added to β. The unlinked loop has graph G1 of Figure 5.

The $\mathscr{L}$ shaped piece determined by $\overset{+}{3}$ and $\overset{-}{7}$ in γ, is detoured, as shown in loop δ of Figure 7. The graph is now G2, its topological transform is shown to its right.

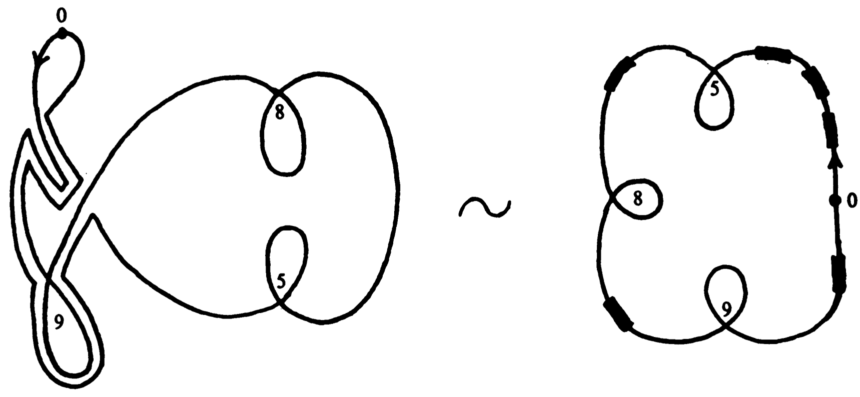
The $\mathscr{S}$ shaped piece determined by $\overset{-}{3}$ and $\overset{+}{2}$ in ε, is detoured, as shown in loop ζ of Figure 8. The graph is now G3, its topological transform η, is precanonical.

The chaining detours are indicated in ϑ of Figure 9. The graph is G4. The loop ι is topologically equivalent to it, and is canonical for TWN=4.



δ

FIGURE 7



ζ η

FIGURE 8

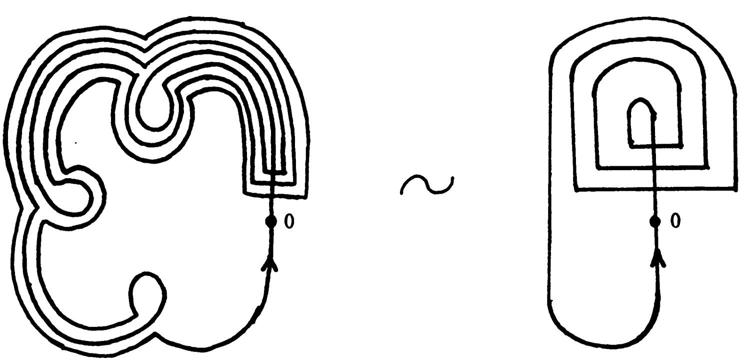
Hosiery. The most frequently encountered detour will be one that relocates a given node, introducing in the process temporary nodes easily located along a normal tubular neighborhood. Let $[a, b]$ be a parameter interval of a normal loop g and (T, e) a normal tube of g. Since the image

$$T[-e, +e]\times[a, b] = [T, e; a, b]$$

is closed, connected and bounded, $D=\text{Bdy } C_\infty[T, e; a, b]$ is a Jordan loop. Suppose next that $g(a)$ is the node N of $[g]$, with a^* its other parameter value. If $b+e<a^*$, then

$$D' = T[-e, +e]\times\{a\} = g[a^*-e, a^*+e]$$

is an arc of D and its complementary arc $D''=\text{Clos }(D\backslash D')$, suitably parametrized, carries a simple detour d of g with support $[a^*-e, a^*+e]$ and with sign$=$sgn det $(g'(a), g'(a^*))$. The immersed arc $g[a, b]$ lies entirely in the bounded complementary component C_0D, except of course, for N on D', and possibly the point $N''=g(b)$ on D''. This last eventuality suggests singling out two admissible cases



ϑ

FIGURE 9

Case 1. N'' is a node of $[g]$ and $a<b^*<b$. Here $N''\in C_0D$.

Case 2. For every node value y, $a<y<b$, $|b-y^*|>e$.

In this latter case,

$$D''' = T[-e, +e]\times\{b\} = g[b^*-e, b^*+e]$$

is a subarc of D''. Set $D'''=\varnothing$ if Case 1 obtains. The open arc(s) $D^\natural = D\backslash(D'\cup D''')$ intersects $[g]$ at points of the form $T(e', y)=g(y^*+e'')$ exclusively, where y is a node value, with $a<y<b$, $|e'|=|e''|=e$, $y^*\notin[a, b]$. In Case 2, $D'''\cap[g]=N''$ alone.

Evidently, the above construction can be performed with the roles of a and b interchanged, in which case D' shall be $g[b^*-e, b^*+e]$. To abbreviate this construction of the *hose detour* on $[a, b]$, detouring the node $g(c)$, where c is either a or b, write

$$d = \text{Hose } [T, e; a, b; c].$$

The unlinking compound $\mathfrak{U}$. Let g be a normal loop, starting outside, which is not properly nested. By Proposition 4(D), g has an exteriorily linked essential subloop at, say, the index q. Let $\{j_1, j_2, \ldots, j_m\}$, $m\geqq 2$, be the indices that link q, in the order of their occurrence on $[g_q]$.

In the Example, in α, $q=3$, $j_1=4$, and $j_2=6$. (Of course, more complicated exterior linking is possible. In loop α^* of Figure 4, the subloop at 6 is a candidate for q, with $j_1=7$, $j_2=4$, $j_3=8$, $j_4=2$.) That the two conditions for exterior linking are necessary for the successful unlinking of an essential subloop can be inferred from loop α^{**} of Figure 4. Here subloop 2 is itself not properly nested. Subloop 3, on the other hand, has its own subloop at 5 linked at 6 and 7. To unlink this loop, two unlinking procedures have to be performed, the first on subloop 5, the second on subloop 3.

Returning to the general case, the nodes $N_{j(k)}$ all lie on $[g/q]$. Let $N_{j(k)}=g(y_k)$, where $x'_q<y_k<x''_q$. The partner $y^*_k\notin[x'_q, x''_q]$. Let (T, e) be a normal tube on g, choose the decreasing sequence $e_k=e/k$, $k=1, 2, \ldots, m+1$; set $y'_k=y_1+e-e_k$. Call $e_{m+1}=e'$. For $2\leqq k\leqq m$, let $d_k=\text{Hose }[T, e_k; y'_k, y_k; y_k]$. The $(m-1)$ supports are disjoint. The effect of the sum $(d_2+\cdots+d_m)g=\bar{g}$ is to pull the linking nodes around $[g_q]$, stacking them up near $N_{j(1)}=N'_{j(1)}$, in the same order. Set $N'_{j(k)}=g(y'_k)$; the nodes $N_{j(k)}$ are replaced by the nodes $N'_{j(k)}$.

Observe, that for $2\leqq p<k\leqq m$, $[d_k]$ crosses $[g]$ at the points $g(x^*_p\pm e_k)$, which lie on $[d^*_p]$ because $e_p>e_k$. Thus the only new nodes are the $N'_{j(k)}$. To remove these, the piece $g[y_1-e, y_1+e]$ must be detoured. Consider the Jordan loop

$$D = C_0[g_q]\cap \text{Bdy } [T, e'; x'_q, x''_q]$$

running "parallel" and inside the boundary of the bounded complementary component of $[g_q]$ that has the node N_q on its closure, $C_0[g_q]$. Because $e'<e_k$, $k=2, 3, \ldots, m$, D meets $[\bar{g}]$ exactly once on each $[d_k]$, and then at the point $T(e'', y'_k)$, where $e''=-\text{sgn}(q)e'$. These points all lie on the arc

$$D = T\{e''\}\times[y_1-e, y_1+e] \quad \text{of } D.$$

Let D'' be its complementary arc, then $D'' \cap [\tilde{g}] = \varnothing$. To complete the construction of the detour $[d_1]$ with support $[y_1 - e, y_1 + e]$, join the loose ends of D'' by short transverse pieces to the corresponding ends of supp (d_1). Thus $d_1\tilde{g}$ makes a U-turn at $y_1 - e$, running off shore Bdy $C_0[g_q]$ in the opposite direction, until it makes another U-turn near $y_1 + e$ and rejoins $[g]$.

The (piecewise) normal loop $\tilde{g} = d_1\tilde{g} = (d_1 + d_2 + \cdots + d_m)g$ has precisely m fewer nodes than g: $W(\tilde{g}) = W(g) \setminus \{j_1, j_2, \ldots, j_m\}$. Because $EW(g)$ is properly nested and q is essential, none of the removed j_k were essential. Hence $EW(\tilde{g}) = EW(g)$. The supports of the detours are each close to the nodes removed. For subsequent constructions it is necessary to choose the insulation small enough to keep the subsequent supports disjoint from those already supporting detours. This precaution will be self understood henceforth. An inductive argument now permits the construction of $\mathfrak{U}$, required for part (A) of Proposition 5.

Preparation for the next step.

LEMMA 11. *Let g be a properly nested normal loop. If $W(g)$ has two indices of unlike sign, neither of which is the initial one, then there must be two indices $i < j$ of unlike sign such that*

(L) *either $i \mid j$ and they are consecutive principal indices under*

$$q = \max\{k \mid k \supset i \text{ and } k \supset j\},$$

(S) *or $0 \neq i \supset j$ and j is principal under i.*

Proof. Reflection on the tree graph representation of $W(g)$ makes this obvious. ▌

In fact, the strict alternative to (S) is stronger than (L). If (S) does not occur, then (L) does, and with $q = 0$. The end of the next two constructions is to design a pair of simple detours d_i and d_j, with disjoint supports close to the respective nodes N_i and N_j, so that $W((d_i + d_j)g) = W(g) \setminus \{i, j\}$.

The L-construction. Without loss of generality, assume that sgn $(i) = +1$, hence sgn $(j) = -1$. The curve

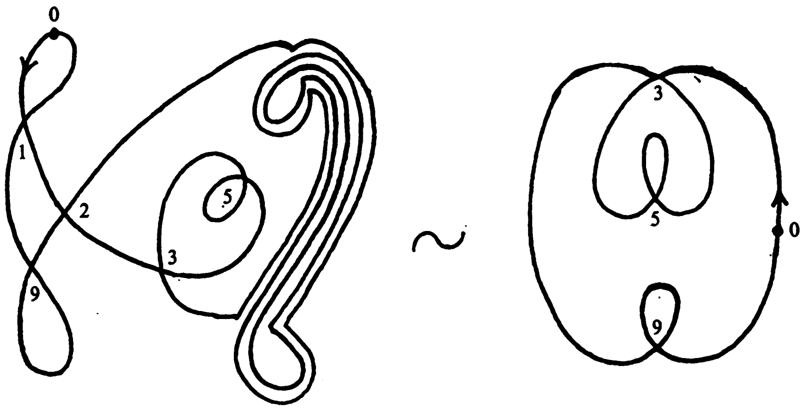
$$g[x_i' - e, x_i'] \cup [g/i] \cup g[x_i'', x_j'] \cup [g/j] \cup g[x_j'', x_j'' + e]$$

which is essentially the stretch $g[x_i' - e, x_j'' + e]$ with the subloops at i and j reduced, has the topological shape of $\mathscr{L}$.

In the example, the subloops $i = 3, j = 7$ of loop γ in Figure 6 displays this shape. The detours are indicated in loop δ, Figure 7. It may as well be pointed out here that there is no particular reason for choosing this pair. The pair 7 and 8 are also candidates for this reduction. The form of the detours is indicated in loop δ' of Figure 10. Subsequent to this choice, the precanonical form eventually obtained is shown by the curve η' in the same figure.

Returning to the general case, Figure 11, set $e_k = e/k$, $k = 1, 2, 3$, and construct

$$\bar{d}_j = \text{Hose}\,[T, e_3; x_i', x_j'; x_j'], \qquad d_i = \text{Hose}\,[T, e_2; x_i'', x_j''; x_i''].$$



δ' η'

FIGURE 10

It is clear that $\bar{d}_j$ detours the piece $g[x_j''-e_3, x_j''+e_3]$, pulling it around $[g_i]$, cutting $[g]$ at $g(x_i'-e_3)$. But this point lies on $g[x_i'-e_2, x_i'+e_2]$, which is detoured in a similar manner by d_i. Examination of $\bar{g}=(d_i+\bar{d}_j)g$, shows that, while N_i and N_j have disappeared, new nodes appear at $X_i=[d_i]\cap[\bar{d}_j]=T(e_3, x_i'+e_2)$ and at $X_j=[d_i]\cap[g]=g(x_j''+e_2)$. By replacing the Γ-shaped piece

$$T\{e_3\}\times[x_i'+e_3, x_i'+e_1]\cup T\{-e_3\}\times[x_i''+e_3, x_j'],$$

of $[\bar{d}_j]$ by the short cut, $T\{-e_1\}\times[x_i''+e_3, x_j']$, a detour d_j is obtained, with support $[x_j''-e_3, x_j''+e_1]$. Now $[d_i]\cap[d_j]=\varnothing$ and $X_j\in[d_j^*]$. The construction is complete.

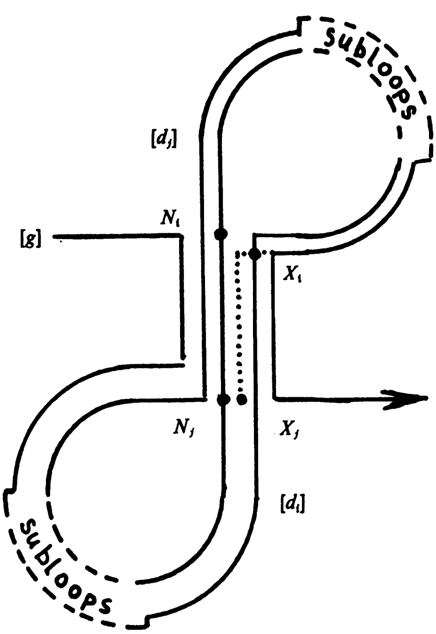


FIGURE 11

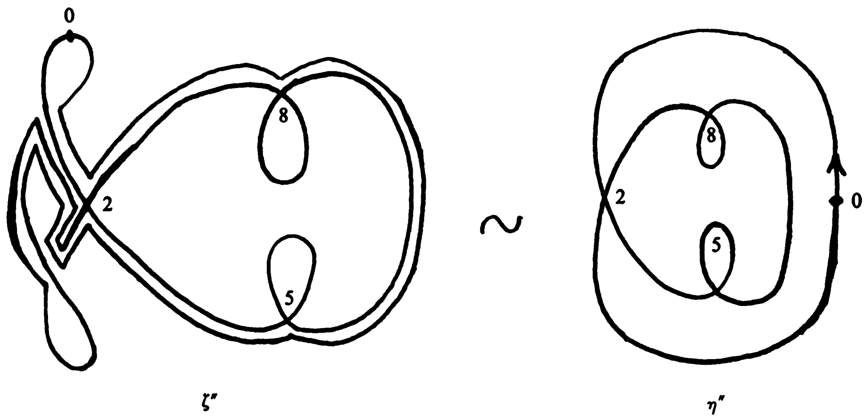


FIGURE 12

The S-construction. Without loss of generality assume sgn $(i)=-1$. So, sgn $(j)=+1$. The stretch $g[x'-e, x''+e]$ with all subloops reduced, except the reduced subloop of j,

$$g[x_i'-e, x_i'] \cup [g/i] \cup [g/j] \cup g[x_i'', x_i''+e],$$

has the topological shape of $\mathscr{S}$.

In the Example, on loop ε of Figure 7, there is $i=1$ and $j=2$. The intended detours are shown in loop ζ of Figure 8. Again, this choice is not unique. Another choice, $i=1$ and $j=9$, is shown by loop ζ'' of Figure 12, with the resulting precanonical shape shown to its right, η''.

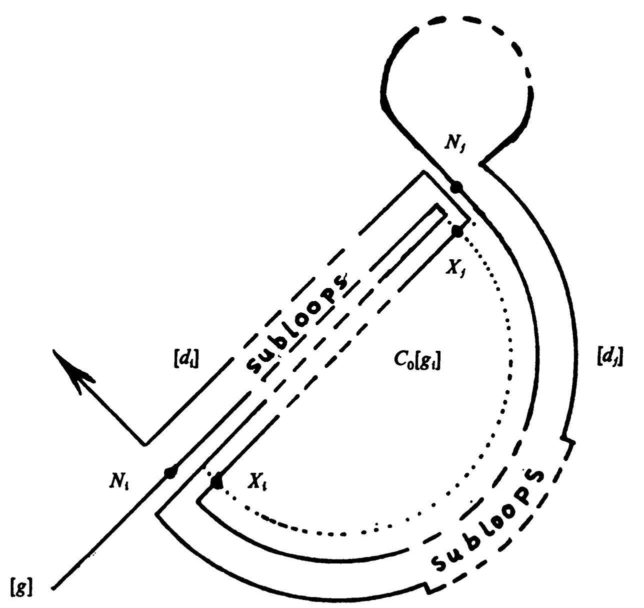


FIGURE 13

Returning to the general case, see Figure 13, set $e_k = e/k$, $k = 1, 2$, and construct

$$d_i = \text{Hose}\,[T, e_1; x_i', x_j' - e_2; x_i'], \qquad \bar{d}_j = \text{Hose}\,[T, e_1; x_j'', x_i'' - e_2; x_j''].$$

The detour d_i removes N_i, but reintersects $[g]$ at $g(x_j' - e_2)$, which however, lies on $[\bar{d}_j^*]$. Nevertheless, $\bar{g} = (d_i + \bar{d}_j)g$ has two new nodes, both on $[d_i] \cap [d_j]$. These two points are $X_i = T(e_1, x_i'' - e_1)$ and $X_j = T(e_1, x_j'' + e_1)$. Consider the Jordan loop

$$D = C_0[g_i] \cap \text{Bdy}\,[T, e_2; x_i', x_i''],$$

which is essentially Bdy $C_0[g_i]$, shrunk inside a little ways. Let D' be that arc of D from $T(e_1, x_j'' + e_2)$ to $T(e_1, x_i'' - e_2)$ which does not cross $[d_i]$. If D'' is its complementary arc, and d_j is $\bar{d}_j$ with D' replacing D'' in $[\bar{d}_j]$, the $[d_i] \cap [d_j] = \varnothing$. This completes the required construction.

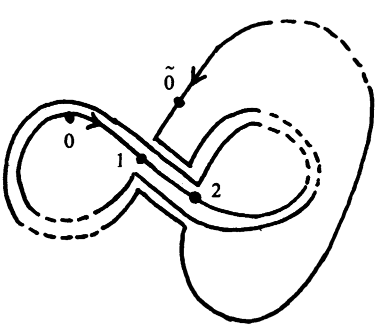
Remark. It is worth noting that both detours of an S-compound have the same sign, namely sgn (j). In an L-compound, the signs are necessarily opposite with sgn $(d_i) =$ sgn (i) and sgn $(d_j) =$ sgn (j). Keeping track of the signs of the detours, while not important for the purposes of this paper, has important consequences in applications that will be given elsewhere.

Reaching precanonical form. By means of these two constructions it is possible to reduce a properly nested loop to one, all of whose indices, save possibly the initial one, have the same sign. If g is such that all indices of $W(g)\backslash\{0\}$ have like sign, opposite that of 0, and if the number of nodes $n \geqq 2$, the starting point can be resigned by removing two further nodes as follows. (Recall that $n = 0$ or 1 means that g is already canonical.)

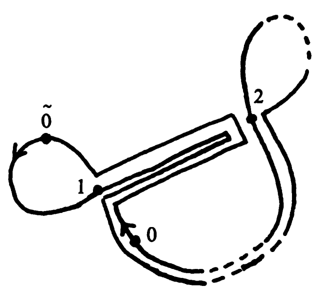
Without loss of generality, assume sgn $(0) = -1$. The two possible situations are illustrated by the loops in Figure 14.

Case $(1 \supset 2)$. Reparametrize g temporarily by $\tilde{g}(x) = g(x + x_1'' - e)$. This defines a permutation $p: W \to \tilde{W}$ so that $N_i = \tilde{N}_{p(i)}$ with $\tilde{N}_{p(0)} = g(x_1'' - e)$. Observe that $p(1) = 1$. Both $p(1)$ and $p(2)$ are now principal under $p(0)$, but of opposite sign,

$$\text{sgn}\,(p(1)) = -\text{sgn}\,(p(2)) = -1.$$



Case $1 \supset 2$



Case $1|2$

Figure 14

Performing an L-compound detour removes these two nodes, and makes the old starting point $g(0)$ a positive starting point. The resulting loop is canonical.

Case $(1|2)$. This time reparametrize $[g]$ to start at $g(x_1'+e)$. Then $p(2)$ is principal under $p(1)$, and they now have unlike sign, $\operatorname{sgn}(p(1))=-\operatorname{sgn}(p(2))=-1$. An S-compound effects the reduction of $W(g)$ to the precanonical form.

Chaining a precanonical loop. Let g be precanonical, with the number of nodes $n\geqq 1$, all having the same sign. Without loss of generality, assume these signs are all positive. The effect of the following n positive simple detours is, this time, not to remove nodes, but to relocate them near the starting point in a different order, so that the resulting loop is chained.

In the Examples, precanonical loops η, η', and η'', once chained, all have the form of ι. However, the order differs: for η, it is as indicated 9, 8, 5. For η', it is 9, 3, 5; and for η'', it is 2, 8, 5.

In general now, let $p\colon W\to W$ be the permutation induced by the reverse of the numerical order of the latter endpoints of the subloop intervals

$$\{x_1'', x_2'', \ldots, x_n'', x_0'' = 2\pi\}.$$

Then $p(0)=0$ and

$$i\supset j \Rightarrow p(i) < p(j), \qquad i|j \Rightarrow p(i) > p(j).$$

Let (T, e) be the usual normal tube. Assign

$$y_j = p(j)e/(n+1), \qquad e_j = e-y_j, \qquad d_j = \text{Hose}\,[T, e_j; y_j, x_j'; x_j'].$$

Thus $[d_j^*]=g[x_j''-e_j, x_j''+e_j]$, all detours are positive, and the new node at $\tilde{N}_{p(j)}=g(y_j)$ is also positive. In fact, this is the only intersection of $[d_j]$ with $[g]$, except possibly near a node N_i where $i<j$. Suppose first that $i\supset j$. Because here x_i'' does not belong to $[0, x_j']$, $[d_j]$ crosses the segment $g[x_i'-e, x_i'+e]$ at the two points $g(x_i''\pm e_j)$. But $e_i>e_j$, and so these crossing points lie in $[d_i^*]$. Next, suppose $i|j$. This time N_i is a node of the curve $g[0, x_j']$ and $[d_j]$ passes by N_i picking up a corner, but no intersection with $[g]$. In this case $e_i<e_j$ and $[d_i]$ lies closer to $g[0, x_i']$ than $[d_j]$ does, running "parallel" to $[d_i]$ back to near N_0. In both cases, then, $[d_i]\cap[d_j]=\varnothing$ for all i, j. The $\tilde{N}_{p(j)}$ are the only nodes of $\tilde{g}=(d_1+\cdots+d_n)g$, which is canonical.

Thus, unlinking, deleting oppositely signed subloops, possibly resigning the starting point, and finally, chaining, establishes Proposition 5.

REMARK. These constructions confirm a number of combinatorial results of Titus [6] and Whitney [10]. In particular two are included here as examples. Let W be the intersection sequence of a normal loop g (parametrized so as to start outside).

(1) (Whitney) $$TWN(g) = \sum_{j\in W(g)} \operatorname{sgn}(j),$$

(2) (Titus) $$\sum_{j\in W(g)} \operatorname{sgn}(j) = \sum_{j\in EW(g)} \operatorname{sgn}(j).$$

If d is a simple detour of g, then $[d]\cup[d^*]$ is a Jordan loop, and so $\mathrm{TWN}(dg)=\mathrm{TWN}(g)$. Consequently, detours do not change the tangent winding number.

It is obvious from visual inspection that (1) is true for g a canonical loop. Chaining a precanonical loop merely permutes the indices, without changing their signs. Formula (1) remains true for precanonical g. Suppose, in the reduction to precanonical form, the starting point had to be resigned. It will be recalled that this amounted to deleting the first two nodes (their common sign was opposite that of the starting point). Thus (1) still holds for loops, all of whose indices are of like sign, opposite that of the starting point. In each of the L- and S-compounds, two oppositely signed indices were delted. So (1) holds for properly nested loops.

Finally, a closer look at the proof of (D) of Proposition 4 will verify the assertion: An externally linked essential subloop is so linked, by pairwise oppositely signed indices. Thus unlinking does not change the algebraic index sum, which confirms (2), finishing the confirmation of (1).

8. **Bridging two canonical loops.** In this section it is shown how two canonical normal loops (of the same tangent winding number) can be detoured to occupy the same intermediate normal loop. Let g_i, $i=\pm 1$, be two normal loops that exhibit the following primitive properties:

(1) Both loops are canonical.

(2) They are *separated*, in the sense that each lies in the unbounded complementary component of the other.

(3) $\mathrm{TWN}(g_i)\geqq 0$, $i=\pm 1$.

(4) In both cases, the starting point is outside.

This last condition permits the construction of some simple curve h, connecting the starting points, called a *bridge*. The bridge is chosen so that

(B1) $h\colon [-1, +1]\to R^2$ is a C^1-embedding,

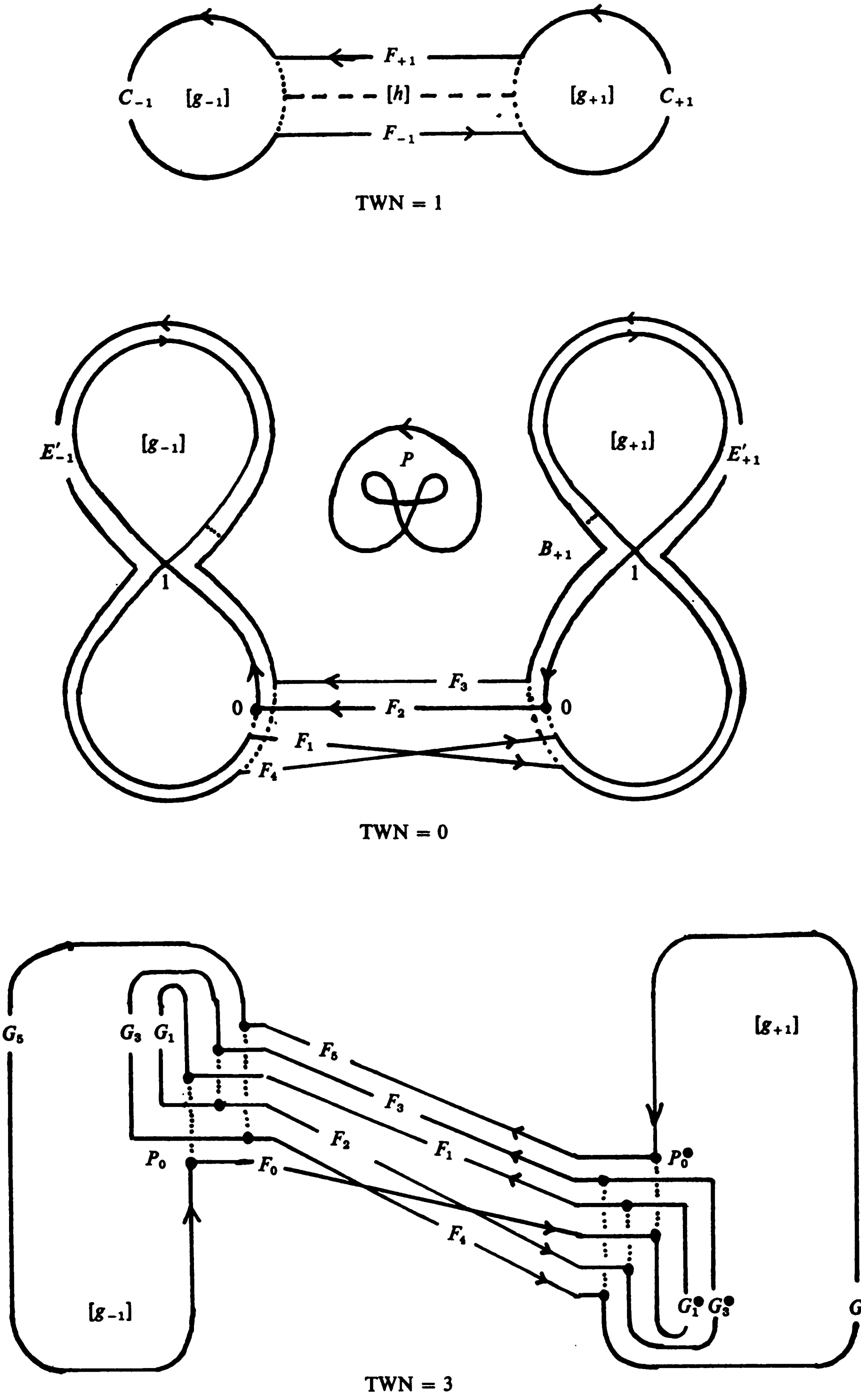
(B2) $h(i)=g_i(0)$, $i=\pm 1$,

(B3) $\operatorname{sgn}\det(h'(-1), g'_{-1}(0))=\operatorname{sgn}\det(g'_{+1}(0), h'(+1))$,

(B4) $[h]$ meets $[g_i]$ only at the tee at $g_i(0)$.

It is not difficult to see how the prescriptions in §6 permit the construction of a normal tubular neighborhood (H, e), of the manifold collection $[g_{-1}]\cup[h]\cup[g_{+1}]$, with $e>0$ small enough for whatever purpose. The source plane is prepared so, that the parameter of g_{-1} runs north on the vertical at $t=-1$, the parameter of h runs east on the abscissa between -1 and $+1$, and the parameter of g_{+1} runs south on the vertical at $t=+1$. Three cases are illustrated in Figure 15. They are $\mathrm{TWN}=1, 0$ and >1.

Case $\mathrm{TWN}=+1$. Each loop is an embedded circle with positive orientation. The tube $[H]$ has the general shape of a dumbbell. Choose simple curves $F_i=H[-1,+1]\times\{ie\}$, $i=\pm 1$, parallel to $[h]$, with F_{+1} directed westward, and F_{-1} directed eastward. Let $[d_i^*]=g_i[-e, +e]$ with $C_i=\operatorname{Clos}[g_i]\backslash[d_i^*]$. Parametrize the curve $F_i\cup C_{-i}\cup F_{-i}$ as the simple detour $[d_i]$, $i=\pm 1$. Then the Jordan loop $C_{-1}\cup F_{-1}\cup C_{+1}\cup F_{+1}$ carries both curves $[d_ig_i]$.



F_{+1}
C_{-1}
$[g_{-1}]$
$[h]$
$[g_{+1}]$
C_{+1}
F_{-1}
TWN = 1
E'_{-1}
$[g_{-1}]$
P
$[g_{+1}]$
E'_{+1}
B_{+1}
1
1
F_3
0
F_2
0
F_1
F_4
TWN = 0
G_5
G_3
G_1
F_5
F_3
F_2
F_1
P_0
F_0
F_4
$P_0^\bullet$
$[g_{+1}]$
$[g_{-1}]$
$G_1^\bullet$
$G_3^\bullet$
$G_5^\bullet$
TWN = 3

FIGURE 15

Case TWN$=0$. Each loop is an embedded figure eight with $\operatorname{sgn}(0)=-\operatorname{sgn}(1)=+1$ for convenience. Consider the following construction: Set $e=e_0/3$, where e_0 is the width of an acceptable tube H. Build $E_j=\text{Bdy } C_\infty H([i-e, i+e]\times S^1)$, $i=\pm 1$, which is a Jordan loop enclosing the corresponding figure eight. Let

$$E_i'' = H\{i-ei\}\times[-2e, e] \quad \text{and} \quad E_i' = \text{Clos } E_i \backslash E_i''.$$

Both are to be oriented counterclockwise. Next, build

$$F_3 = H[-1+e, +1-e]\times\{e\}, \quad \text{oriented westerly,}$$

$$F_2 = H[-1, +1]\times\{0\}, \quad \text{oriented westerly,}$$

$$F_1 = H[\text{graph of } x = e(3-e+t)/(e-2),\ -1 \leqq t \leqq +1-e], \quad \text{easterly,}$$

$$F_4 = H[\text{graph of } x=e(3-e-t)/(e-2),\ -1+e \leqq t \leqq +1], \quad \text{easterly.}$$

The curve $F_1 \cup E'_{+1} \cup F_3 \cup E'_{-1} \cup F_4 \cup g_{+1}[e, 2\pi] \cup F_2$ supports a parametrization d_{-1}, under which it becomes a positive detour of $[d^*_{-1}]=g_{-1}[2\pi-e, 2\pi]$. This is, obviously, not a simple detour. Rather, it has three self intersections. However, temporarily introducing the short connecting piece $B_{+1}=H[1, 1+e]\times\{x_1'-2e\}$, (remember that the parameter of the right loop runs southward on $t=+1$!), oriented to pass from E'_{+1} to $[g_{+1}]$, it is clear that d_{-1} is the product of two simple positive detours, the second detouring B_{+1}. Likewise, the curve $F_2 \cup g_{-1}[0, 2\pi-e] \cup F_1 \cup E'_{+1} \cup F_3 \cup E'_{-1} \cup F_4=[d_{+1}]$ is a positive compound, detouring $[d^*_{+1}]=g_{+1}[0, e]$. The pretzel shaped curve

$$P = F_4 \cup g_{+1}[e, 2\pi] \cup F_2 \cup g_{-1}[0, 2\pi-e] \cup F_1 \cup E'_{+1} \cup F_3 \cup E'_{-1}$$

has TWN$=0$, with three nodes, and carries both $[d_i g_i]$, $i=\pm 1$.

Case TWN>1. Each loop has the shape somewhat like that of an embedded snail. In the application, each g_i will have been prepared as in the part on chaining a precanonical loop of §7. To ease the description, assume for once, that the corners of the chaining detours nearest the starting point have not been smoothed. The construction here is a direct continuation of the chaining construction. Therefore it shall be assumed that chaining has been performed with width parameter $e/2$, where e is suitable for H.

Examining the source squares $\square_i=[-e, +e]\times[i-e, i+e]$, $i=\pm 1$, it happens that the minorah shaped, oriented one-complexes in the source:

$$\Psi_i = H^{-1}(H\square_i \cap [g_i]) \cap \square_i, \qquad i = \pm 1,$$

are symmetrical under the central reflection $(t, x)\to(-t, -x)$. Recall that the node of $[g_i]$ occur at the points $H(i, kie')$, $i=\pm 1$, $k=1, 2, \ldots, n$, $e'=e/2(n+1)$. On Ψ_{-1} mark the points $P_0=(-1, 0)$, $P_1=(-1, (n+1)e')$, $P_{2k}=(-1+ke', (n+1-k)e')$, and $P_{2k+1}=(-1+ke', (n+1+k)e')$, $k=1, 2, \ldots, n$. This peculiar indexing is so chosen that g_{-1} runs through the sequence $HP_0, HP_1, HP_2, \ldots,$ in that order. Mark $2(n+1)$ further points Q_j in $\square_{-1}$ according to the prescription

$$\text{Ordinate } Q_j = -1+(n+1)e', \qquad \text{Abscissa } Q_j = \text{Abscissa } P_j.$$

The H-image of the simple polygon $P_jQ_jQ^{\bullet}_{j+1}P_{j+1}$ is denoted by F_j, where $Q^{\bullet}$ is centrally symmetric with Q. These paths are oriented according to the rule:

F_{odd} is directed westward, F_{even} is directed eastward. The oddly indexed polygons are "parallel" and so, mutually disjoint. So are the evenly indexed ones, except for F_0, which intersects each F_{even} exactly once. No F_{odd} meets an F_{even}. The points $\{HP_j\}$ decompose $[g_{-1}]$ into simple, disjoint arcs. Denote the arc from HP_j to HP_{j+1} by the symbol G_j.

Assign the superscript $\bullet$ also to the H-images of objects obtained under the central reflection in the source. Observe that $[g_{-1}]$ is the ordered union of the arcs G_j, and $[g_{+1}] = \bigcup_j G_j^{\bullet}$. Consider the following simple positive detours:

$$[d_k] = F_{2k} \cup G^{\bullet}_{2k+1} \cup F_{2k+1} \quad \text{detouring } [d_k^*] = G_{2k},$$
$$[\bar{d}_k] = F_{2k-1} \cup G_{2k-1} \cup F_{2k} \quad \text{detouring } [\bar{d}_k^*] = G^{\bullet}_{2k}.$$

The piecewise normal loop

$$F_0 \cup G_1^{\bullet} \cup F_1 \cup G_1 \cup \cdots \cup F_{2j} \cup G^{\bullet}_{2j+1} \cup F_{2j+1} \cup G_{2j+1} \cup \cdots \cup G_{2n+1}$$

carries both $[(d_0+d_1+\cdots+d_n)g_{-1}]$ and $[(\bar{d}_0+\bar{d}_1+\cdots+\bar{d}_n)g_{+1}]$. This intermediate loop is also canonical.

9. **Assembly of the proof.** Let g_i, $i = \pm 1$, be two C^1-regular immersions of the circle in the plane.

(1) The work in the foregoing section was done on two separated loops. The first step is to obtain a separation of images. Let the target plane be coordinatized by the complex parameter w, write $R^2(w)$. Choose a point w_0, with imaginary part sufficiently less than $\min\{\operatorname{Im} w \mid w \text{ on } [g_{-1}] \cup [g_{+1}]\}$, to insure that the set $[g_{-1}] \cup [g_{+1}]$ subtends an arc of less than 45° from w_0. Let another plane have complex parameter z. Consider the function F: $R^2(z) \to R^2(w) : F(z) = w_0 + z^2$. Let P^+ be the open half plane in the target given by $\operatorname{Im} w > \operatorname{Im} w_0$. Let the first and third quadrants in the source be given as

$$Q_i = \left\{ z \mid (1+i)\frac{\pi}{2} < \arg z < (2+i)\frac{\pi}{2}, i = \pm 1 \right\}.$$

Then $F_i = F|Q_i$ is an orientation preserving diffeomorphism of Q_i onto P^+. Thus, $f_i = F_i^{-1} \circ g_i$, $i = \pm 1$, are two separated regular loops in $R^2(z)$. Without loss of generality, assume $\operatorname{TWN}(g_i) \geqq 0$.

Notational convention. The following notational convention will hold until step (7). Let $R^2(z)$ be just R^2, f_{-1} be just f, and f_{+1} is written $\bar{f}$. The bar over a letter shall serve to indicate the analogous object for the third quadrant, that has been constructed in the first under the name of that letter. Thus $Q = Q_{-1}$ and $\bar{Q} = Q_{+1}$.

(2) From Proposition 1, there is a positive monotopy G of f, so that $f_1 = Gf$, is normal. (Analogously, there is a $\bar{G}$ with $\bar{G}\bar{f} = \bar{f}_1$ normal. This sort of explicit remark for the third quadrant will generally be dispensed with.) The width e for

both G and $\bar{G}$, and for all subsequent tubular neighborhoods must be chosen sufficiently small for any construction to remain in the appropriate quadrant.

(3) Pick an orientation preserving reparametrization $f_2(x)=f_1(r(x))$, $r'>0$, so that f_2 starts outside. Moreover, make certain that the starting points $f_2(0)$ and $\bar{f}_2(0)$, are not centrally opposite, so that they lie in the same open half plane, call it Q_0.

(4) By Proposition 5, there is a mixed sum of simple detours, so that $f_3=(\mathfrak{U}+\mathfrak{A}+\mathfrak{B})f_2$ is canonical. It was seen in the third case of §8, that for TWN > 1, the chaining detours $\mathfrak{B}$ were correlated to the tube H. So for this case, it is well to construct first the bridge h and the tube H, for the two separated precanonical loops $(\mathfrak{U}+\mathfrak{A})f_2$ and $(\bar{\mathfrak{U}}+\bar{\mathfrak{A}})\bar{f}_2$. The image $[H]$ must stay away from the singularity of F (at the origin), to insure that $F\circ h$ still be an immersion. If $[h]$ is so chosen as to remain in the open half plane Q_0, and the width of H is sufficiently small, all this will be accomplished.

(5) According to §8, there is a positive compound $\mathfrak{D}$, such that $f_4=\mathfrak{D}f_3$ differs from the corresponding $\bar{f}_4=\bar{\mathfrak{D}}\bar{f}_3$ only by a reparametrization of $[f_4]=[\bar{f}_4]$.

Remark. At this point in the process all auxiliary constructions have been made. The care that has been lavished on keeping track of the parametrization while building the detours has served its purpose. Provided one keeps the proper names of the nodes, subarcs, detours etc. for easier reference, no ambiguity is introduced in assuming that in (5) $f_4=\bar{f}_4$, and that there are (appropriately reparametrized) detours such that

$$(\dagger)\qquad \mathfrak{D}(\mathfrak{U}+\mathfrak{A}+\mathfrak{B})f_1 = \bar{\mathfrak{D}}(\bar{\mathfrak{U}}+\bar{\mathfrak{A}}+\bar{\mathfrak{B}})\bar{f}_1.$$

(6) The situation now becomes somewhat different for each of the three cases of §8. The goal is to rewrite (†) as

$$(*)\qquad (\mathfrak{M}+\bar{\mathfrak{N}}^*\mathfrak{P})f_1 = (\bar{\mathfrak{M}}+\mathfrak{N}^*\bar{\mathfrak{P}})\bar{f}_1 = f_0$$

so that all detours shown in (*) are positive.

Case TWN$=1$. Here $(\mathfrak{U}+\mathfrak{A})f_1$ is already canonical. The detour $\mathfrak{D}$ consists in a single positive simple detour, call it p, here, which is supported so close to N_0 that the sum $(\mathfrak{U}+\mathfrak{A}+p)$ is permissible. Moreover $[p]$ occupies all of $[(\bar{\mathfrak{U}}+\bar{\mathfrak{A}})\bar{f}_1]$, except of course, for $[\bar{p}^*]$ near $\bar{N}_0$. Hence $[p]\supset[\bar{d}]$ for every detour $\bar{d}$ in $\bar{\mathfrak{U}}+\bar{\mathfrak{A}}$. In particular, if $\bar{d}$ is negative, $\bar{d}^*$ is a positive detour of p. So the positive compound $\bar{d}^*p$ is a positive detour of f_1. Splitting $\bar{\mathfrak{U}}+\bar{\mathfrak{A}}=\bar{\mathfrak{N}}+\bar{\mathfrak{M}}$, with $\bar{\mathfrak{N}}$ negative and $\bar{\mathfrak{M}}$ positive, the detour $\bar{\mathfrak{N}}^*p$ is a positive detour of f_1. It follows that

$$(\mathfrak{U}+\mathfrak{A}+\bar{\mathfrak{N}}^*p)f_1 = (\bar{\mathfrak{M}}+\bar{p})\bar{f}_1.$$

The analogous argument leads to equation (*), setting $p=\mathfrak{P}$ here.

Case TWN$=0$. Here $(\mathfrak{U}+\mathfrak{A})f_1$ is canonical, and by the remark above, one may as well assume that both N_0 and $\bar{N}_0$ are positive starting points. The compound $\mathfrak{D}$ here, consists in the product $p=p''p'$, of simple positive detours. Recall that p', supported near N_0, proceeds through the tube sheathing h, to the third quadrant covering some of $(\bar{\mathfrak{U}}+\bar{\mathfrak{A}})\bar{f}_1$, and the little piece B_{+1} near the only node. Then p'', detouring B_{+1}, covers the remainder (except for the small piece near $\bar{N}_0$), returns through the tunnel to the first quadrant and runs clear around $(\mathfrak{U}+\mathfrak{A})f_1$ without intersecting it. If $\bar{d}$ belongs to $\bar{\mathfrak{R}}$, its support is not near the last remaining node, nor near the starting point. Thus $[\bar{d}]$ belongs either to $[p']$ or to $[p'']$. In short, $\bar{d}^*p$ is a permissible positive product detour of f_1. As in the previous case, write $\mathfrak{P}=p''p'$, and equation (*) follows.

Case TWN>1. Please refer to the notation of §8. Let $\mathfrak{D}$ be written as $\mathfrak{D}=p_0''+\cdots+p_n''$, where $[p_k''^*]=G_{2k}$ and $[p_k'']=F_{2k}\cup G^\bullet_{2k+1}\cup F_{2k+1}$. If $\mathfrak{B}=p_1'+\cdots+p_n'$ is the rechaining compound, so renumbered that $[p_n']\cap G_0$ is the first node of $(\mathfrak{U}+\mathfrak{A}+\mathfrak{B})f_1$ encountered, and $[p_1']\cap G_{2n}$ is the last, then $[p_k''^*]\subset[p_k']$, and $p_k''p_k'=p_k$, $k=1,2,\ldots,n$, is a legitimate positive product detour of f_1. If $\bar{d}\in\bar{\mathfrak{R}}$, then $[\bar{d}]$ is not near the starting point of $\bar{f}_1$, and so is contained in some $G^\bullet_{\text{odd}}$. Therefore $\bar{d}^*$ is a positive detour of some p_k, $k=0,1,\ldots,n$ (set $p_0=p_0''$). Let $\mathfrak{P}=p_0+\cdots+p_n$ and let $\bar{\mathfrak{R}}^*\mathfrak{P}$ be the sum of all appropriately formed products, such that

$$(\mathfrak{U}+\mathfrak{A}+\bar{\mathfrak{R}}^*\mathfrak{P})f_1 = (\bar{\mathfrak{M}}+\bar{\mathfrak{P}})\bar{f}_1$$

holds.

Explicitly, a typical summand of $\bar{\mathfrak{R}}^*\mathfrak{P}$ has the form

$$\left(\sum \bar{d}^*\right)p_k''p_k'$$

summing over all those $\bar{d}$ in $\bar{\mathfrak{R}}$ with $[\bar{d}]\subset G^\bullet_{2k+1}$. The analogous argument again leads to the equation (*).

(7) Recalling Proposition 3, the modified monotopy $(\mathfrak{M}+\bar{\mathfrak{R}}^*\mathfrak{P})G$ takes f to f_0 and $(\bar{\mathfrak{M}}+\mathfrak{R}^*\bar{\mathfrak{P}})\bar{G}$ takes $\bar{f}$ to f_0. Reverting to the original notation of this section, write these two monotopies as $K_i : f_i\sim f_0$, $i=\pm 1$. By the second part of Proposition 3, the two monotopies combine to the folded monotopy K: $[-1,+1]\times S^1\to R^2$ from f_{-1} to f_{+1} with the single fold f_0. It then follows that the composition $G=F\circ K$ is the required fold-monotopy from g_{-1} to g_{+1}.

(8) Finally, let g_i, $i=\pm 1$, be separated loops. Denote by A the annulus $[-1,+1]\times S^1$. The assumption that there necessarily exists a monotopy (without folds!) from g_{-1} to g_{+1} leads quickly to a contradiction. For, assume on the contrary, that $G: A\to R^2$ is an immersion with $G(i,x)=g_i(x)$. Then $G(\text{Bdy } A)=[g_{-1}]\cup[g_{+1}]$, and this union is not connected. The image $G(A)$, on the other hand, is bounded and connected, and therefore $D=\text{Bdy } C_\infty G(A)$ is connected. Because the g_i are located in each others unbounded complementary components, there is a point P on $D\backslash G(\text{Bdy } A)$. Let P' be in $G^{-1}(P)\cap\text{Int } A$. Since G is an immersion, it is a local diffeomorphism at each point of Int A. Hence P is in the interior of the image of G, and not on its boundary, which contradicts the choice of P.

This completes the proof of the Folded Ribbon Theorem for plane immersions of the circle.

References

1. George K. Francis, *The folded ribbon theorem*, Dissertation, Univ. of Michigan, 1967. (University Microfilms Abstract 67-15,622, Vol. 28,6, 1967.)

2. Heinz Hopf, *Über die Drehung der Tangenten und Sehnen ebener Kurven*, Compositio Math. **2** (1935), 50–62.

3. Morris L. Marx and Roger F. Verhey, *Polynomial extensions of normal curves*, (to appear).

4. John Milnor, *Lectures on the h-cobordism theorem*, Princeton Univ. Press, Princeton, N. J., 1965.

5. Valentin Poenaru, *On regular homotopy in codimension one*, Ann. of Math. **83** (1966), 257–265.

6. Charles J. Titus, *A theory of normal curves and some applications*, Pacific J. Math. **10** (1960), 1083–1096.

7. ———, *Characterizations of the restriction of a holomorphic function to the boundary of a disk*, J. Analyse Math. **18** (1967), 351–358.

8. ———, *The combinatorial topology of analytic functions on the boundary of a disk*, Acta Math. **106** (1961), 45–64.

9. Roger F. Verhey, *Diffeomorphic invariants of immersed circles*, (to appear).

10. Hassler Whitney, *On regular closed curves in the plane*, Compositio Math. **4** (1937), 276–284.

University of Michigan,
Ann Arbor, Michigan