Math. Ann. 199, 131—153 (1972)

Winding Numbers on Surfaces. II

Applications

D. R. J. Chillingworth

§ 0. Introduction

This is a sequel to the paper [7] in which a theory of *winding numbers* of regular closed curves on surfaces, based on [16, 18], was extended to a theory of winding numbers of homotopy classes of closed curves on surfaces. We give here two applications of the theory.

The main application is to the problem of determining when a given element of the fundamental group of a compact surface can be represented by simple closed curves: we are able to obtain a workable algorithm which uses the formulae for calculating winding numbers given in [7]. Earlier algorithms for solving this problem using non-euclidean geometry have been given by Reinhart [17] and Călugăreanu [3], and a combinatorial algorithm has been given by Zieschang [19, 21].

The second application is to the investigation of the structure of the *homeotopy group* of a closed surface, i.e. the group of homeomorphisms modulo those isotopic to the identity. Although the results produced at this stage are not essentially new, the fact that some concrete results can be obtained does suggest that a fuller understanding of the algebraic properties of winding numbers may lead to more powerful methods for analyzing homeotopy groups.

The definitions and notations used throughout are those of [7]. A *surface* is a connected separable 2-manifold, called a *closed surface* if it is compact and with empty boundary. Any compact orientable or non-orientable surface may be regarded as a 2-sphere with a finite number $r \geqq 0$ of open discs removed and a finite number $g \geqq 0$ of handles attached in the orientable case or a finite number $n \geqq 0$ of Möbius strips attached in the non-orientable case; the *genus* of the surface is g or n.

The author wishes to thank the referee for his many helpful comments, in the light of which the original version of this paper was substantially rewritten.

§ 1. Geometric Preliminaries

Let M be a compact surface, and x a point of M. In all that follows we shall assume that M is not a sphere, since the problems with which we are concerned are trivial in that case.

We begin with some technical constructions which provide a machinery for relating the algebraic structure of $\pi_1(M, x)$ to the geometry of M. These are taken directly from [19] or [22].

We recall from [7, § 7] (see [19, 22]) that every compact surface M admits a *canonical curve system* Σ described as follows. If M is orientable with boundary $\partial M \neq \emptyset$ then Σ consists of a system of oriented simple closed curves α_i, β_j and σ_k $(1 \leqq i, j \leqq g; 1 \leqq k \leqq r)$ with the following properties:

(i) the curves all meet at x and are otherwise disjoint;

(ii) σ_k together with a component $\varrho_k \subset \partial M$ bounds an annulus R_k whose interior is disjoint from all the other curves of Σ;

(iii) cutting along the curves of Σ dissects M into the disjoint union of the R_k together with a disc whose boundary runs

$$\alpha_1 \beta_1 \alpha_1^{-1} \beta_1^{-1} \alpha_2 \beta_2 \alpha_2^{-1} \beta_2^{-1} \ldots \alpha_g \beta_g \alpha_g^{-1} \beta_g^{-1} \sigma_1 \sigma_2 \ldots \sigma_r$$

where "α_i^{-1}" here means "α_i backwards" etc.

The fundamental group $\pi_1(M, x)$ has the presentation

$$\{a_i, b_j, s_k (1 \leqq i, j \leqq g; 1 \leqq k \leqq r) | d = 1\}$$

where a_i, b_j, s_k denote the homotopy classes of α_i, β_j, σ_k respectively, and

$$d = a_1 b_1 a_1^{-1} b_1^{-1} a_2 b_2 a_2^{-1} b_2^{-1} \ldots a_g b_g a_g^{-1} b_g^{-1} s_1 s_2 \ldots s_r .$$

When $\partial M = \emptyset$ we ignore (ii) and dispense with the ϱ_k, σ_k, R_k and s_k. If M is non-orientable with $\partial M \neq \emptyset$ then Σ consists of a system of oriented simple closed curves η_i, σ_k $(1 \leqq i \leqq n; 1 \leqq k \leqq r)$ satisfying (i), (ii) above together with:

(iii)′ cutting along the curves of Σ dissects M into the disjoint union of the R_k together with a disc whose boundary runs

$$\eta_1 \eta_1 \eta_2 \eta_2 \ldots \eta_n \eta_n \sigma_1 \sigma_2 \ldots \sigma_r .$$

In this case $\pi_1(M, x)$ has the presentation

$$\{u_i, s_k (1 \leqq i \leqq n; 1 \leqq k \leqq r) | d' = 1\}$$

where u_i, s_k denote the homotopy classes of η_i, σ_k respectively and

$$d' = u_1^2 u_2^2 \ldots u_n^2 s_1 s_2 \ldots s_r .$$

When $\partial M = \emptyset$ we again ignore (ii) and dispense with ϱ_k, σ_k, R_k and s_k.

The following definition also comes from [19] or [22]:

(1.1) Definition. When M is orientable a *dual dissection* Σ^*, corresponding to the canonical curve-system Σ, consists of a collection of oriented simple closed curves α_i^*, β_j^* $(1 \leqq i, j \leqq g)$ and (if $\partial M \neq \emptyset$) arcs σ_k^* $(1 \leqq k \leqq r)$ with the following properties:

(1) the curves and arcs all have a common base-point q but are otherwise disjoint;

(2) σ_k^* connects q to a point of ϱ_k;

(3) each α_i^* (or β_j^*, σ_k^*) meets α_i (or β_j, σ_k respectively) exactly once, and meets none of the other curves of Σ;

(4) cutting along the curves and arcs of Σ^* dissects M into a disc D whose boundary runs

$$\alpha_1^* \beta_1^{*-1} \alpha_1^{*-1} \beta_1^* \alpha_2^* \beta_2^{*-1} \dots \alpha_g^{*-1} \beta_g^* \sigma_1^* \varrho_1 \sigma_1^{*-1} \sigma_2^* \varrho_2 \sigma_2^{*-1} \dots \sigma_r^* \varrho_r \sigma_r^{*-1}.$$

(Note the difference between this and the expression for d.)

If $\partial M = \emptyset$ there are no arcs σ_k^*.

When M is non-orientable a *dual dissection* Σ^* consists of oriented simple closed curves η_i^* $(1 \leqq i \leqq n)$ and arcs σ_k^* $(1 \leqq k \leqq r)$ satisfying (1), (2) above and

(3′) each η_i^* or σ_k^* meets η_i or σ_k respectively exactly once, and meets none of the other curves of Σ

(4′) cutting along the curves and arcs of Σ^* dissects M into a disc D whose boundary runs

$$\eta_1^* \eta_1^* \eta_2^* \eta_2^* \dots \eta_n^* \eta_n^* \sigma_1^* \varrho_1 \sigma_1^{*-1} \sigma_2^* \varrho_2 \sigma_2^{*-1} \dots \sigma_r^* \varrho_r \sigma_r^{*-1}.$$

Again if $\partial M = \emptyset$ there are no arcs σ_k^*.

(1.2) Definition. Define an equivalence relation "$\sim$" on the set of all words in a given collection of symbols ("letters") by $w_1 \sim w_2$ if and only if w_1 is a cyclic permutation of w_2. Each equivalence class is called a *cyclic word*, and the equivalence class of w is denoted by (w).

Any word w in the generators of $\pi_1(M, x)$ and their inverses represents an element of $\pi_1(M, x)$ which we denote by $[w]$. If $w_1 \sim w_2$ then $[w_1]$ and $[w_2]$ are conjugate elements.

Let γ be an oriented closed curve not passing through q, and in *general position* with respect to Σ^* (i.e. γ meets Σ^* in at most a finite number of points none of which is a self-intersection point of γ, and at each of which γ crosses over the curve of Σ^*). Let χ^* be any curve or arc of Σ^*, and let p be a point of $\gamma \cap \chi^*$. By assigning to a neighbourhood of p the orientation induced by the path qp along χ^* (in the positive direction) from some chosen fixed orientation at q, we may give meaning to the statements "γ cuts χ^* from right to left (positively) at p" or "γ cuts χ^* from left to right (negatively) at p".

(1.3) Definition. The *total word* ("Ablesung" in [19, 22]) of γ with respect to Σ^* is a cyclic word in the letters $a_i^{\pm 1}$, $b_j^{\pm 1}$, $s_k^{\pm 1}$ $(1 \leqq i, j \leqq g;$ $1 \leqq k \leqq r)$ if M is orientable or $u_i^{\pm 1}$, $s_k^{\pm 1}$ $(1 \leqq i \leqq n;$ $1 \leqq k \leqq r)$ if M is non-orientable, obtained from γ as follows. Choose any point y on γ and not on any curve or arc of Σ^*, and proceed once round γ in the positive direction. At each point where γ crosses a curve or arc of Σ^* write

$$\begin{aligned} &a_i \text{ if } \gamma \text{ crosses } \alpha_i^* \text{ positively},\\ &a_i^{-1} \text{ if } \gamma \text{ crosses } \alpha_i^* \text{ negatively},\\ &b_j \text{ if } \gamma \text{ crosses } \beta_j^* \text{ positively},\\ &\text{etc.} \end{aligned}$$

The sequence of letters so obtained gives a word w, and the *total word* of γ is then defined to be the cyclic word (w), which we denote by $\mathrm{tw}(\gamma)$. Clearly, if we had chosen a different starting point on γ we would have obtained a word w' with $(w')=(w)$, and so the total word does not depend on the choice of starting point.

The curves α_i, η_i etc. have total words (a_i), (u_i) etc., and it is easy to see that if two closed curves have the same total word they are (freely) homotopic.

Given a cyclic word (w) it is always possible to construct a closed curve γ on M with $\mathrm{tw}(\gamma)=(w)$. However, if w represents an element of $\pi_1(M,x)$ which is known to contain simple closed curves, such a γ may not necessarily itself be simple. For example, on a torus a closed curve whose total word is $(a_1^3 b_1^2)$ cannot be simple although it is homotopic to a simple curve (e.g. with total word $(a_1^2 b_1 a_1 b_1)$). In [19, 22] Zieschang uses a method of "reduction" of a word w to a word W with the property that if the element $[w]\in\pi_1(M,x)$ contains simple closed curves then some element conjugate to $[w]$ contains a simple closed curve whose total word is (W). We follow this method here.

(1.4) Definition. A word w in the $a_i^{\pm 1}$ etc. (or $u_i^{\pm 1}$ etc.) is *completely reduced* if it satisfies the following conditions:

(i) w is cyclically reduced (i.e. no two cyclically adjacent letters in w are mutual inverses);

(ii) neither w nor any cyclic permutation of w contains any subword of the element $d^{\pm 1}$ (or $d'^{\pm 1}$), or of any cyclic permutation of $d^{\pm 1}$ (or $d'^{\pm 1}$), whose length is more than half the length of d (or d');

(iii) if w or any cyclic permutation of w contains a subword which consists of exactly half of a cyclic permutation of $d^{\varepsilon}(\varepsilon=\pm 1)$ then this half contains the first element of d (i.e. $a_1, u_1(g,n>0)$ or $s_1(g,n=0)$) when $\varepsilon=+1$, or the last element of d^{-1} (i.e. a_1^{-1}, u_1^{-1} or s_1^{-1}) when $\varepsilon=-1$.

If w is any given word it is easy to see that in general a completely reduced word W may be obtained from w by reducing w cyclically, replacing "long" subwords of $d^{\pm 1}$ (or $d'^{\pm 1}$) by shorter words, replacing "half-length" subwords if necessary, reducing cyclically again if possible, and so on. We call this process *complete reduction*. Exceptions to this rule occur when M is a torus ($g=1, r=0$) or a Klein bottle ($n=2, r=0$), since for example the word $a_1 b_1 a_1$ in the first case or the word $u_1 u_2 u_1 u_2$ in the second case cannot be reduced to words satisfying (iii). However, these cases are treated separately in § 3.

Note that $[W]$ is conjugate to $[w]$ in $\pi_1(M, x)$.

If W is completely reduced and $(W')=(W)$ then W' is completely reduced. Therefore we also refer to the *cyclic word* (W) as being *completely reduced*. The element $1 \in \pi_1(M, x)$ is represented by the empty word which is conventionally regarded as being completely reduced.

The following two lemmas are taken straight from [19] with appropriate changes in notation. The proof of the first is elementary, whereas the second is more difficult and is based on the methods of [8].

(1.5) Lemma. *Any simple closed curve γ based at x on M may be deformed by an isotopy (which in general moves x) to obtain a simple closed curve whose total word is completely reduced.*

(1.6) Lemma. *Suppose the number of generators of $\pi_1(M, x)$ is at least 5. If W and W' are completely reduced words such that $[W]$ and $[W']$ are conjugate in $\pi_1(M, x)$ and $[W]$ contains simple closed curves then $(W)=(W')$.*

(1.7) Lemma. *If $c \in \pi_1(M, x)$ contains simple closed curves then any element conjugate to c contains simple closed curves.*

Proof. This follows immediately from the well-known elementary fact that any inner automorphism of $\pi_1(M, x)$ is induced by a homeomorphism of M (see e.g. [20]). The inner automorphism defined by an element z is induced by a homeomorphism of M which may be described as the result of sliding x once round a closed curve representing z.

We shall now assume until § 3 that the number of generators of $\pi_1(M, x)$ is at least 5: thus every word w in the generators of $\pi_1(M, x)$ and their inverses can be reduced completely, and if $[w]$ contains simple closed curves the resulting completely reduced cyclic word depends only on the conjugacy class of $[w]$ in $\pi_1(M, x)$.

Lemmas (1.5)–(1.7) can be combined to yield the following:

(1.8) Lemma. *Let $c \in \pi_1(M, x)$ be a non-trivial element and let w be a word representing c. Let W be a word obtained by reducing w completely. Then there exists an oriented closed curve γ in general position with respect to Σ^* such that the total word of γ is (W), and γ can be chosen to be simple if and only if c contains simple closed curves.*

We now turn attention to the geometric situation. Recall that Σ^* dissects M into a disc D. Let γ be any closed curve in general position with respect to Σ^* on M, and not nullhomotopic. On D the curve becomes a finite collection $\mathscr{C}$ of oriented paths whose end-points lie on ∂D and are identified in pairs to form γ from $\mathscr{C}$ when segments of ∂D are identified in pairs to form M from D. Let D^0 denote the interior of D (i.e. $D^0 = D - \partial D$). It is easy to see that γ may be deformed by a homotopy if necessary to ensure that $\mathscr{C}$ satisfies the following conditions:

(1) each path of $\mathscr{C}$ is a simple arc;

(2) any pair of distinct arcs of $\mathscr{C}$ are either disjoint or meet at just one point of D^0 where they actually cross;

(3) no three arcs of $\mathscr{C}$ have a common point of intersection.

(1.9) Definition. If γ is an oriented closed curve as described in Lemma (1.8) and such that $\mathscr{C}$ satisfies conditions (1), (2) and (3) above, then γ is called an *elementary representative curve* for the conjugacy class of c in $\pi_1(M, x)$.

We recall from [7] that an orientation-preserving closed curve γ in M is said to be *direct* if there is a covering of M such that γ lifts to a simple closed curve. (In fact in [7] we choose an infinite cyclic covering, but this restriction is unimportant.) Lemma (2.4) of [7] characterizes these direct curves by the absence of null-homotopic loops. An orientation-reversing closed curve γ is *direct* if 2γ (the curve obtained by going twice round γ) is direct.

(1.10) Lemma. *An elementary representative curve is direct.*

Proof. If γ is an orientation-preserving elementary representative curve which is not direct then by Lemma (2.4) of [7] it contains a null-homotopic loop. By (1) the loop cannot be disjoint from Σ^*, and so $\mathrm{tw}(\gamma) = (w)$ for some w containing a null subword, i.e. a subword representing $1 \in \pi_1(M, x)$. However, a word containing a null subword must fail to satisfy (i) or (ii) of Definition (1.4): as $\pi_1(M, x)$ has at least five generators this fact follows from Reidermeister's solution of the word problem [15, pp. 200–204] or from [22, Satz IV.15]. (See [22, p. 130] or [19, p. 26] for details of the argument when $\partial M \neq \emptyset$; see also [19, p. 20] for other references.) Similarly, if γ is orientation-reversing and not direct then it is easy to see that ww contains a null subword and that this is impossible when (w) is completely reduced.

§ 2. Construction of the Algorithm

From now until Lemma (2.8) (excepting Lemma (2.4)) in order to save space we shall only give proofs of lemmas in the case when M is *orientable*. However, the results remain valid in the non-orientable case, winding numbers (see below) being taken mod 2 where appropriate. The detailed proofs for the general case can be found in [4, pp. 127–147].

Let M be given some fixed differentiable structure. A closed curve in M will be called *regular* if it has a continuous non-vanishing tangent: see [7].

The Simple Case

Suppose $c \neq 1 \in \pi_1(M, x)$ contains simple closed curves, and let γ be a simple elementary representative curve for the conjugacy class C of c. Without loss of generality γ may be chosen to be a regular curve. Since γ is simple the arcs of $\mathscr{C}$ are mutually disjoint. We assume that c is not one of the elements $a_i^{\pm 1}$, $b_j^{\pm 1}$, $s_k^{\pm 1}$, so that $\mathscr{C}$ contains at least two arcs. Let κ_1, κ_2 be any two arcs of $\mathscr{C}$, and let v be a smooth arc in D^0 joining a point p_1 on κ_1 to a point p_2 on κ_2, otherwise disjoint from $\kappa_1 \cup \kappa_2$, and chosen so that v meets no other arc of $\mathscr{C}$ more than once. Let ξ_1' denote the arc in D obtained by going along κ_2 in the negative direction to p_2, then along v^{-1} to p_1, and finally along κ_1 in the positive direction; let ξ_2' be the arc going along κ_2 in the positive direction to p_2, then along v^{-1} to p_1, and then along κ_1 in the negative direction. Approximate ξ_1', ξ_2' by smooth arcs ξ_1, ξ_2 which are the same as ξ_1', ξ_2' outside a closed disc B with $v \subset B^0 \subset B \subset D^0$ (where B^0 denotes the interior of B), and which meet just once, touching or crossing according to whether the orientations of κ_1, κ_2 at p_1, p_2 are the same or opposite with respect to an orientation of v. See Fig. 1, which illustrates the two possibilities. Furthermore, assume that the point $p = \xi_1 \cap \xi_2$ lies on no arc of $\mathscr{C}$, and that ξ_i meets each arc κ of $\mathscr{C} - (\kappa_1 \cup \kappa_2)$ exactly once or not all depending on whether the end-points of ξ_i separate those of κ on ∂D or not $(i = 1, 2)$.

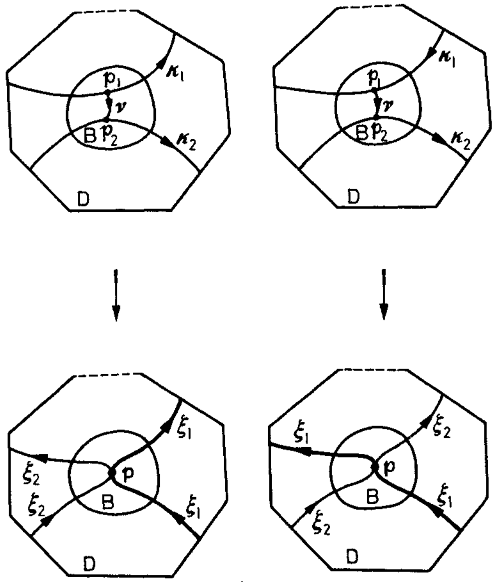

Fig. 1

The arcs ξ_1 and ξ_2 and those of $\mathscr{C}-(\kappa_1 \cup \kappa_2)$ fit together to give (with a suitable parametrization) a regular curve γ' on M. Let δ denote the curve which starts at p and proceeds along ξ_1 in the positive direction and then around the rest of γ' until first returning to p (along ξ_2), and let ε be the curve which starts at p and proceeds along ξ_2 in the positive direction and then around γ' until first returning to p (along ξ_1). Choose parametrizations such that each of δ, ε is regular, and γ' is the composition of δ with ε. If $\operatorname{tw}(\gamma)=(W)$ then some cyclic permutation of W is of the form uv where $(u)=\operatorname{tw}(\delta)$ and $(v)=\operatorname{tw}(\varepsilon^{-1})$, and $\operatorname{tw}(\gamma')=(uv^{-1})$.

Let Q be an open disc neighbourhood of q (the common intersection point of the curves of Σ^*) not meeting γ' and not containing the basepoint x of M, and let $N=M-Q$. The curves γ, δ, ε and γ' can be thought of as curves in N. Any word w in the letters $a_i^{\pm 1}$, $b_j^{\pm 1}$, $s_k^{\pm 1}$ represents an element of $\pi_1(N, x)$ as well as an element of $\pi_1(M, x)$: we continue to denote the element of $\pi_1(M, x)$ by $[w]$, and we denote the element of $\pi_1(N, x)$ by $[w]'$.

In Lemmas (2.1)–(2.7) which follow we shall understand *nullhomotopic* to mean nullhomotopic in N, and by a *null word* or *null subword* we shall mean a word (subword) which represents $1 \in \pi_1(N, x)$. Clearly a word which represents $1 \in \pi_1(N, x)$ also represents $1 \in \pi_1(M, x)$.

(2.1) Lemma. *If $u \neq v$ then δ, ε and γ' are not nullhomotopic and are direct (as curves on N).*

Proof. If δ is nullhomotopic then $\operatorname{tw}(\gamma)$ contains a null subword, which is impossible since γ is an elementary representative curve. If δ is not direct on N then by [7, Lemma (2.4)] it contains a nullhomotopic loop λ. Since γ is simple the arcs of $\mathscr{C}$ are mutually disjoint so the vertex of λ lies in K where

$$K=(\xi_1 \cup \xi_2) \cap (\mathscr{C}-(\kappa_1 \cup \kappa_2)).$$

But ξ_1, ξ_2 are arcs meeting only once so $\lambda \not\subset D^0$ and hence $\operatorname{tw}(\gamma)$ again contains a null subword, which is impossible. Therefore δ is direct. Similar arguments apply to ε.

If γ' is nullhomotopic then $[uv^{-1}]'=1 \in \pi_1(N, x)$ and since uv is cyclically reduced this implies $u=v$. If γ' is not direct it contains a nullhomotopic loop μ which cannot lie in D^0 and by the above cannot be contained in δ or in ε. The vertex k of μ must lie in K, and this then implies that μ contains the whole of δ or the whole of ε. In the first case uv^{-1} is of the form uzy where z, y are subwords and either uz is null and y is non-empty, or yu is null and z is non-empty. Therefore either $vu=y^{-1}uu$ or $uv=uuz^{-1}$ where the arc kp_1p_2 or p_1p_2k of γ gives rise to the subword uu. It is easy to verify that this is impossible since γ is simple and k lies between the arcs κ_1 and κ_2 in D. Thus μ cannot contain the whole of δ.

Similarly, it cannot contain the whole of ε. This is a contradiction, so γ' is direct.

We now invoke the theory of *winding numbers* developed in [7]. The following is a brief summary of part of the theory that we shall need here, with the non-orientable case included for completeness. We refer to [7, §§ 1, 2 and 5] for precise definitions and further explanations.

The *winding number* $\omega_X(\gamma)$ of an oriented regular closed curve γ with respect to a (continuous) non-vanishing vector field X on a (smooth) surface N ($\partial N \neq \emptyset$) is defined to be the number of times the tangent vector to γ rotates relative to the X-vector on going once round γ, reduced mod2 if N is non-orientable[1]. If N is orientable the sign depends on a choice of orientation of N. If two curves are direct and homotopic (but not nullhomotopic) they have the same winding number ([7, Theorem (5.3)]). Homotopy classes (base-point free) correspond to conjugacy classes of elements of $\pi_1(N, x)$, and so we can define the *winding number* $\omega_X(C)$ *of the conjugacy class* $C \subset \pi_1(N, x)$ as an element of $\mathbb{Z}$ (N orientable) or $\mathbb{Z}_2$ (N non-orientable) by

$$\omega_X(C) = \omega_X(\gamma)$$

where γ represents C and is direct. Finally, for $c \neq 1 \in \pi_1(N, x)$ let

$$\omega_X(c) = \omega_X(C)$$

where C = conjugacy class of c, and let $\omega_X(1) = 1$.

Now let X be any continuous non-vanishing vector field on the surface $N = M - Q$ which we are considering. Such a vector field exists since $\partial N \neq \emptyset$.

If w is a word representing an element $[w]'$ of $\pi_1(N, x)$ we shall write $\omega_X(w)$ instead of $\omega_X([w]')$.

(2.2) Lemma. $\omega_X(\gamma') = \omega_X(\delta) + \omega_X(\varepsilon)$.

Proof. Immediate, since γ' is the composition $\delta \circ \varepsilon$.

(2.3) Lemma. *If* $u \neq v$ *then* $\omega_X(uv^{-1}) = \omega_X(u) + \omega_X(v^{-1})$ *where* u, v *are as in Lemma* (2.1).

Proof. Immediate from Lemmas (2.1) and (2.2) and the definition of $\omega_X(u)$ etc.

This is a key result in the construction of the algorithm. It completes our study of the simple case.

[1] In the non-orientable case with γ orientation-reversing it is necessary to choose a starting-point x_0 on γ at which the tangent to γ is in the same direction as $X(x_0)$, and to transport an orientation at x_0 around γ: see [7, § 5].

The Non-Simple Case

The following lemma applies whether M is orientable or not.

(2.4) Lemma. *For every conjugacy class C containing elements of the form $[W]$ where W is completely reduced and not of the form VV for any subword V there is a regular elementary representative curve γ with* $\mathrm{tw}(\gamma)=W$ *and with the following property. If ξ and η are any two paths in γ with the same beginning points and same end points, and $\xi\cap\eta$ consists of a finite number of points, then the closed curve consisting of the composition $\xi\circ\eta^{-1}$ is not nullhomotopic in N.*

Proof. Let γ_1 be any regular elementary representative curve for C with $\mathrm{tw}(\gamma)=W$, and suppose ξ and η are two paths of γ_1 described as above, but with $\xi\circ\eta^{-1}$ nullhomotopic (in N). Let ξ' be a path covering ξ in the universal cover M' of M and let η' be a lift of η to M' starting at the starting point of ξ'. Each of ξ' and η' is an arc since γ is direct (Lemma (1.10)). Since $\xi\circ\eta^{-1}$ is nullhomotopic it follows by the homotopy lifting property that η' ends at the end point of ξ'. Suppose without loss of generality that the number of points $\xi'\cap p^{-1}(P_1)$ is greater than or equal to the number of points of $\eta'\cap p^{-1}(P_1)$ where P_1 is the set of double points of γ_1 and $p: M'\to M$ is the covering map; if this is not so, rename ξ and η. Let ξ'' and η'' denote arcs in $p^{-1}(\gamma_1)$ obtained by extending ξ' and η' slightly, i.e. to include no more points of $p^{-1}(P_1)$ or points of $p^{-1}(\gamma_1\cap\Sigma^*)$.

Let the number of points of $\eta'\cap p^{-1}(P_1)$ (including the end-points of η') be $n(\geqq 2)$, the number of those belonging to $\xi'\cap\eta'$ being $k(\geqq 2)$. There is a homotopy of ξ'' (keeping its end-points fixed) to a path $h\xi''$ near η'' which meets η'' exactly once or not at all and which meets $p^{-1}(\gamma_1)$ a total of $n-k+1$ or $n-k$ times respectively (not counting the end-points of $h\xi''$). Suppose $h\xi''$ to be chosen sufficiently close to η'' so that $h\xi''\cap p^{-1}(\Sigma^*)$ has the same number of points as $\eta''\cap p^{-1}(\Sigma^*)$. Since the end-points of ξ' project to distinct points of M (otherwise ξ and η would be closed curves and $\gamma=\xi\circ\eta$ would be homotopic in N to 2ξ, contradicting the hypothesis on the conjugacy class C) the homotopy of ξ'' to $h\xi''$ can be chosen to induce a homotopy of γ_1 on M, keeping η fixed, to a curve γ_2 whose number of self-intersections is *less* than that of γ_1. Since $\xi\circ\eta^{-1}$ is nullhomotopic on N and $\pi_1(N,x)$ is the free group on the generators of $\pi_1(M,x)$ it follows that the contribution to $\mathrm{tw}(\gamma_1)$ from ξ is exactly the same as the contribution from η, which is the same as the contribution from $ph\xi''$, and so $\mathrm{tw}(\gamma_1)=\mathrm{tw}(\gamma_2)$. Without altering $\mathrm{tw}(\gamma_2)$ or increasing the number of self-intersections we may now replace γ_2 by a regular curve which is an elementary representative curve. By repeating this whole process a finite number of times we must eventually obtain a regular curve γ_m which is an elementary representative curve for C satisfying the condition of the lemma, since a curve with no self-intersections automatically satisfies the condition. The lemma is thus proved.

An elementary representative curve satisfying the condition of Lemma (2.4) will be called *minimal.*

Suppose now that $c \in \pi_1(M, x)$ does not contain simple curves. Suppose the conjugacy class C of c is as in Lemma (2.4), and let γ be a minimal regular elementary representative curve for C lying in N. The arcs of $\mathscr{C}$ cannot all be disjoint. Assume the construction of $\mathscr{C}$ and the parametrization of γ to be such that whenever two arcs $\kappa_1, \kappa_2 \in \mathscr{C}$ meet, the tangents to κ_1 and κ_2 at $r = \kappa_1 \cap \kappa_2$ coincide.

Let κ_1, κ_2 be two arcs of $\mathscr{C}$ which meet; let $r = \kappa_1 \cap \kappa_2$. Let δ_r be the curve which begins at r and proceeds along κ_1 and then continues around γ until first returning to r, and let ε_r be the curve which begins at r and proceeds along κ_1^{-1} and then continues around γ^{-1} until first returning to r, each parametrized so that γ is the composition $\delta_r \circ \varepsilon_r^{-1}$. Now the composition $\delta_r \circ \varepsilon_r$ has two cusps at r; let γ_r' be a regular curve in N, which is the same as $\delta_r \circ \varepsilon_r$ outside a closed disc B where $r \in B^0 \subset B \subset D^0$ and $(B \cap \mathscr{C}) \subset \kappa_1 \cup \kappa_2$, obtained by rounding off the cusps of $\delta_r \circ \varepsilon_r$ at r in such a way that $B \cap \gamma_r'$ is the disjoint union of two arcs. If $\mathrm{tw}(\gamma) = (W)$ then $(W) = (uv)$ where $(u) = \mathrm{tw}(\delta_r)$, $(v) = \mathrm{tw}(\varepsilon_r^{-1})$, and $\mathrm{tw}(\gamma_r') = (uv^{-1})$.

In order to save space in the proofs which follow we now revert to our assumption that M is *orientable*. The results still hold, however, in the non-orientable case.

Clearly δ_r, ε_r are not nullhomotopic, since γ is direct.

(2.5) Lemma. *The curves δ_r and ε_r are direct on N.*

Proof. Suppose δ_r contains a nullhomotopic loop λ with vertex s. Since γ is direct, λ must contain r. Thus $\lambda = \xi \circ \eta^{-1}$ where ξ and η are both paths in γ from s to r. This contradicts the minimality of γ and hence δ_r contains no nullhomotopic loop and is therefore direct. Similarly, ε_r is direct.

Now let X be any continuous non-vanishing vector field on N as before.

(2.6) Lemma. *If γ_r' is also direct on N then*

$$\omega_X(uv^{-1}) = \omega_X(u) + \omega_X(v^{-1}) \pm 1 .$$

Proof. By choosing B sufficiently small the contributions to $\omega_X(\gamma_r')$ from δ_r and ε_r outside B may be made arbitrarily close to $\omega_X(\delta_r)$, $\omega_X(\varepsilon_r)$ respectively, and the contribution from $\gamma_r' \cap B$ may be made arbitrarily close to ± 1 since the contributions from rounding off the cusps are both arbitrarily close to $+\frac{1}{2}$ or to $-\frac{1}{2}$. See Fig. 2. Hence $\omega_X(\gamma_r')$ may be made arbitrarily close to $\omega_X(\delta_r) + \omega_X(\varepsilon_r) \pm 1$, and since winding numbers are

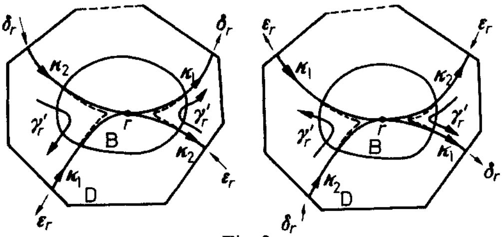

Fig. 2

integer-valued we have

$$\omega_X(\gamma_r') = \omega_X(\delta_r) + \omega_X(\varepsilon_r) \pm 1 .$$

The result then follows from Lemma (2.5) and the definitions of $\omega_X(u)$ etc.

Note. In the non-orientable case it is necessary to pay close attention to starting-points and to transportation of orientations around curves.

(2.7) Lemma. *Every minimal elementary representative curve γ for C either contains at least one double point r_* for which γ_{r_*}' is direct on N, or has total word (W) with $W = V^m$ where V is a subword and $m \geqq 2$.*

Proof. For $\kappa \in \mathscr{C}$ let $e(\kappa)$ denote the end-point (as opposed to the beginning-point) of κ. If $\kappa_i \cap \kappa_j \neq \emptyset$ and $r = \kappa_i \cap \kappa_j$ let $D(r)$ denote the "triangle" (a topological 2-disc) in D bounded by the arcs $re(\kappa_i)$, $e(\kappa_i)\, e(\kappa_j)$, $e(\kappa_j) r$ in κ_i, ∂D and κ_j, respectively, and not containing the rest of $\kappa_i \cup \kappa_j$. For $\kappa_1, \kappa_2 \in \mathscr{C}$ write $\kappa_1 \sim \kappa_2$ if, after identifying segments of ∂D to give Σ^*, the points $e(\kappa_1)$ and $e(\kappa_2)$ lie on the same curve χ of Σ^* and κ_1 and κ_2 meet χ from the same side. Let $\tau(\kappa)$ denote the successor to κ, i.e. the arc of $\mathscr{C}$ which follows κ in the sequence of the arcs which describes γ; define inductively $\tau^0(\kappa) = \kappa$, $\tau^n(\kappa) = \tau(\tau^{n-1}(\kappa))$, $n = 1, 2, \ldots$. Given any double point r of γ we have $r = \kappa_i \cap \kappa_j$ for some κ_i, $\kappa_j \in \mathscr{C}$; let $n(r) \leqq \infty$ be the least upper bound of all integers n such that $\tau^s(\kappa_i) \sim \tau^s(\kappa_j)$ for all s with $0 \leqq s \leqq n$. If $\kappa_i \nsim \kappa_j$ put $n(r) = -1$. Note that if $n(r) = \infty$ then $\tau^s(\kappa_i) \sim \tau^s(\kappa_j)$ for all integers s, and it is easy to see that in that case some non-trivial cyclic permutation of W is the same as W (where $(W) = \mathrm{tw}(\gamma)$) and hence that W is of the form V^m where V is a subword and $m \geqq 2$. We thus assume $n(r) < \infty$ for the remainder of the proof.

Let n_0 be the minimum of $n(r)$ taken over all double points r, and let r_0 be a double point for which $n(r_0) = n_0$. Number the arcs of $\mathscr{C}$ so that $r_0 = \kappa_1 \cap \kappa_2$, and construct the curve γ_{r_0}' as above. If γ_{r_0}' is direct we take $r_* = r_0$. If γ_{r_0}' is not direct it contains a nullhomotopic loop λ with vertex r_1, say, and since $\mathrm{tw}(\gamma)$ is completely reduced λ must contain at least one of the two arcs of $\gamma_{r_0}' \cap B_0$ (where B_0 denotes the small disc containing r_0

inside which the cusps of $\delta_{r_0}\circ\varepsilon_{r_0}$ are rounded off to give γ'_{r_0}) otherwise γ would contain λ. In fact, since γ is minimal, λ must contain both arcs of $\gamma'_{r_0}\cap B_0$, since otherwise two arcs ξ, η going from r_1 to r_0 (or r_0 to r_1) in γ would be such that $\xi\circ\eta^{-1}$ is nullhomotopic. Hence λ contains the whole of δ_{r_0} or the whole of ε_{r_0}, and by renumbering κ_1 and κ_2 if necessary we may assume λ contains δ_{r_0}. This means u must be of the form pq and $v=phq$ where p, q, h are (possibly empty) subwords. The complementary loop to λ in γ contributes the subword h. We have $r_1\in\tau^l(\kappa_2)$ where $l=$ length of p (taken to be zero if p is empty).

Let $\kappa_3=\tau^l(\kappa_2)$, and let κ_4 be the other arc of $\mathscr{C}$ containing r_1 (i.e. $r_1=\kappa_3\cap\kappa_4$).

First suppose $n_0\geqq 0$, so $\kappa_i\sim\kappa_j$ for all κ_i, κ_j with $\kappa_i\cap\kappa_j\neq\emptyset$. Corresponding to the fact that u and v both begin with the subword p we have

$$\tau^s(\kappa_1)\sim\tau^s(\kappa_2)\,,\quad s=0,1,\ldots,l-1$$

and then since $v=phq$ and u also ends with q we have

$$\tau^s(\kappa_4)\sim\tau^{l+s}(\kappa_1)\,,\quad s=0,1,\ldots,l'-1$$

where $l'=$ length of q. However, $\tau^{l'}(\kappa_4)=\kappa_1$ and $\tau^{l+l'}(\kappa_1)=\kappa_2$ and so in fact

$$\tau^s(\kappa_4)\sim\tau^{l+s}(\kappa_1)\,,\quad s=0,1,\ldots,l'+n_0\,.$$

By definition of $n(r_1)$ we have

$$\tau^s(\kappa_3)\sim\tau^s(\kappa_4)\,,\quad s=0,1,\ldots,n(r_1)$$

and hence

$$\tau^{l+s}(\kappa_1)\sim\tau^{l+s}(\kappa_2)\,,\quad s=0,1,\ldots,\min(n(r_1),l'+n_0)$$

because $\kappa_3=\tau^l(\kappa_2)$. Therefore by definition of $n_0=n(r_0)$ we have

$$l+\min(n(r_1),l'+n_0)\leqq n_0\,.$$

The choice of r_0 implies $n_0\leqq n(r_1)$ and so

$$l=0$$

and

$$n(r_1)=n_0=n(r_0)$$

as l' cannot be zero if $l=0$. Thus

$$\kappa_3=\kappa_2$$

and r_1 lies between r_0 and $e(\kappa_2)$ on κ_2. Also the sub-arc $r_1e(\kappa_4)$ of κ_4 is disjoint from the sub-arc $r_0e(\kappa_1)$ of κ_1, since if $r'\in r_1e(\kappa_4)\cap r_0e(\kappa_1)$ there would exist two arcs ξ,η from r' to r_0 in γ with $\xi\circ\eta^{-1}$ nullhomotopic, thus contradicting the minimality of γ. Hence the "triangle" $D(r_1)$ is strictly contained in $D(r_0)$.

If $n_0 = -1$ then we arrive at the same results as these, although the proof has to be worded slightly differently.

Now either γ'_{r_1} is direct (in which case we take $r_* = r_1$) or by replacing r_0 by r_1 in the above (remember $n(r_1) = n_0$) we obtain a double point r_2 with $n(r_2) = n_0$ and $D(r_2)$ strictly contained in $D(r_1)$. The point r_2 lies on $\kappa_2 = \kappa_3$ or on κ_4 according to which of δ_{r_1} or ε_{r_1} is contained in the nullhomotopic loop of γ'_{r_1}. If γ'_r is not direct for any double point r we thus obtain by induction a sequence of double points $r_0, r_1, r_2, \ldots$ with $n(r_i) = n_0$ and $D(r_{i+1})$ strictly contained in $D(r_i)$, $i = 0, 1, 2, \ldots$. The sequence must have repeats since the number of double points is finite. However, it is impossible for the sequence to repeat if $D(r_{r+1}) \subset D(r_i)$ strictly. This contradiction shows that there must exist some double point r_* for which γ'_{r*} is direct, which proves the lemma.

We now have all the ingredients that are needed for the construction of the algorithm, except in the case when $\mathrm{tw}(\gamma)$ is of the form (V^m) $(m \geqq 2)$. However, this case will easily be disposed of using the following well-known result. For a proof see e.g. [17] or [19], or [10, Theorem 4.2].

(2.8) Lemma. *Let $c = b^m \in \pi_1(M, x)$ where b is non-trivial and $m \geqq 2$. Then c contains simple closed curves if and only if $m = 2$ and b contains simple closed orientation-reversing curves.*

§ 3. The Algorithm

We have been assuming above that the number of generators of $\pi_1(M, x)$ is at least 5. In order to give a comprehensive solution to the problem of finding whether an element of $\pi_1(M, x)$ contains simple closed curves we shall first consider the exceptional cases (i.e. where $\pi_1(M, x)$ has at most four generators) and then give the algorithm for the general case (including M non-orientable). In cases (i)–(vi) we shall by abuse of notation not distinguish between an element of $\pi_1(M, x)$ and a word which represents it.

Case (i): $\pi_1(M, x) = \{s_1, \ldots, s_r | s_1 \ldots s_r = 1\}$ where $1 \leqq r \leqq 4$. *(M = disc with $r-1$ holes.)*

When $r = 1$ there is no problem. Otherwise, we observe that $\pi_1(M, x)$ can be presented as the free group generated by $s_1, \ldots, s_{r-1}$. The whole of §§ 1 and 2 continue to be valid with *cyclic reduction* (see Definition (1.4)) in the free group playing the role of *complete reduction* in the original presentation. For example, the analogues of Lemmas (1.8) and (1.10) continue to hold in this context. We can therefore apply to this case the algorithm as it is formulated for the general case (vii) below, with s_r replaced by $s_{r-1}^{-1} \ldots s_1^{-1}$ and "completely reduced" then replaced by "cyclically reduced". Of course, these remarks apply equally well when $r > 4$.

Case (ii): $\pi_1(M, x) = \{u_1 | u_1^2 = 1\}$. *(M = projective plane.)*

Clearly both homotopy classes contain simple closed curves.

Case (iii): $\pi_1(M, x) = \{u_1, s_1 | u_1^2 s_1 = 1\}$. *(M = Möbius strip.)*

Here $\pi_1(M, x)$ can be considered as the infinite cyclic group generated by u_1, and the only non-trivial elements represented by simple closed curves are $u_1^{\pm 1}$ and $u_1^{\pm 2}$.

Case (iv): $\pi_1(M, x) = \{u_1, s_1, s_2 | u_1^2 s_1 s_2 = 1\}$. *(M = Möbius strip with (open) disc removed.)*

Bearing in mind Case (iii) it is easy to show that the only non-trivial elements represented by simple closed curves are those conjugate to u_1, u_1^2, $u_1 s_1$, $(u_1 s_1)^2$, s_1, $u_1^2 s_1$, or their inverses. In order to check whether or not a given word represents an element conjugate to one of these one replaces s_2 by $s_1^{-1} u_1^{-2}$ and then checks for the conjugacy in the free group generated by s_1 and u_1.

Case (v): $\pi_1(M, x) = \{a_1, b_1 | a_1 b_1 a_1^{-1} b_1^{-1} = 1\}$. *(M = torus.)*

It is well known that the element $a_1^p b_1^q$ contains simple curves if and only if $(p, q) = (0, 0)$, $(0, \pm 1)$, $(\pm 1, 0)$ or p and q are coprime. This may easily be proved by observing that any non-trivial element c contains simple closed curves if and only if there is an automorphism of $\pi_1(M, x)$ taking c to a_1. A different proof can be found in [17].

Case (vi): $\pi_1(M, x) = \{u_1, u_2 | u_1^2 u_2^2 = 1\}$. *(M = Klein bottle.)*

It is easy to deduce from the results about isotopy classes of simple curves on M in [12] that the only simple curves are those whose homotopy classes are conjugate to 1, $u_1^{\pm 1}$, $u_2^{\pm 1}$ or $u_1^{\pm 2}$. A full list of these conjugate elements may be computed easily by writing $u_1 = a$, $u_1 u_2 = b$ so that $\pi_1(M, x) = \{a, b | aba^{-1} b = 1\}$, and it consists of the elements 1, $(u_1 u_2)^{\pm 1}$, $u_1^{\pm 2}$, and $(u_1 (u_1 u_2)^n)^{\pm 1}$ for any $n = 0, \pm 1, \pm 2, \ldots$.

Case (vii): *The general case.*

Given an element $c \in \pi_1(M, x)$ first express c as a word w in the generators a_i, b_j, s_k (or u_i, s_k) of $\pi_1(M, x)$ and their inverses, and then reduce w completely (Definition (1.4)). Denote the resulting completely reduced word by W. If W is empty (i.e. $c = 1$) or consists of only one letter, then $[W]$, and hence c, contains simple closed curves. Therefore suppose $W = x_1 x_2 \ldots x_p$ where $p \geqq 2$. There are $\frac{1}{2} p(p-1)$ ways of expressing the cyclic word (W) as (uv) where u, v are non-empty subwords: call each such expression a *division* of W.

Let X be any continuous non-vanishing vector field on $N = M - Q$, where Q is an open neighbourhood of the base-point of a dual dissection of M as in § 2.

If c contains simple curves it follows from the results in § 2, culminating in Lemma (2.3), that

$$\omega_X(uv^{-1}) = \omega_X(u) + \omega_X(v^{-1})$$

for every division $(W) = (uv)$ with $u \neq v$.

If c does not contain simple curves, then by Lemmas (2.6), (2.7), and (2.8) either $W = V^m (m \geqq 2)$ or there exists a division $(W) = (uv)$ with

$$\omega_X(uv^{-1}) \neq \omega_X(u) + \omega_X(v^{-1}).$$

Therefore the final result may be stated as follows:

The element c contains simple curves if and only if a completely reduced word W obtained from c as in § 1 *satisfies one of the conditions:*

(i) *W is empty* (i.e. $c = 1$) *or contains only one letter,*

(ii) *$W = V^2$ where $[V]$ contains an orientation-reversing simple closed curve,*

(iii) *W contains at least* 2 *letters, $W \neq V^m$ for any V and any $m \geqq 2$, and for every division $(W) = (uv)$ the equation*

$$\omega_X(uv^{-1}) = \omega_X(u) + \omega_X(v^{-1})$$

is satisfied, where X is any continuous non-vanishing vector field on $N = M - Q$.

Note. In case (ii) the algorithm must be applied in turn to $[V]$ (which is completely reduced since (V^2) is completely reduced), but case (ii) will not arise a second time since if $V = Z^2$ then $[V]$ can contain only orientation-preserving curves.

The algorithm is clearly not complete, however, until we are able effectively to compute whether or not the equation

$$\omega_X(uv^{-1}) = \omega_X(u) + \omega_X(v^{-1})$$

is satisfied for each division $(W) = (uv)$. To do this we will invoke the Formulae 1 or 2 of [7, § 7]: see below.

It is to be expected that if the equation is satisfied for one choice of vector field X it will be satisfied for all vector fields: whether c contains simple curves or not does not depend on X. Now $\omega_X(u)$ means $\omega_X([u]')$, but we cannot necessarily apply Formulae 1 (1′) or 2 from [7] immediately, since we are using the generators of $\pi_1(M, x)$ to generate the free group $\pi_1(N, x)$. However, to apply the machinery of [7] to N we need only introduce another generator s_{r+1} corresponding to the boundary of Q. If $\bar{u}$ denotes a *cyclically reduced* word obtained from u (here $\bar{u}$ is not necessarily completely reduced) then $\bar{u}$ contains no subword which is null in $\pi_1(N, x)$. Hence $\omega_X([u]') = \omega_X([\bar{u}]')$ since $[u]'$ is conjugate to $[\bar{u}]'$, and then calculating $\omega_X([\bar{u}]')$ by Formulae 1′ or 2 of [7] we have

$$\omega_X(u) - \varphi_X(\bar{u}) = T(\bar{u}) \qquad (M \text{ orientable})$$

or

$$\omega_X(u) - \varphi'_X(\bar{u}) = T(\bar{u}) \quad \mathit{mod}\, 2 \quad (M \text{ non-orientable}),$$

where $\varphi_X(\varphi'_X)$ is the unique homomorphism $\pi_1(N, x) \to \mathbb{Z}(\mathbb{Z}_2)$ coinciding with the function ω_X on the chosen generators of $\pi_1(M, x)$ (here we identify a word in the generators with the element of $\pi_1(N, x)$ which it represents), and where $T(\bar{u})$ is defined as follows:

$$T(\bar{u}) = P(\bar{u}) - (\bar{A}(\bar{u}) + B(\bar{u}) + \bar{S}(\bar{u}))$$

where $\bar{u} = y_1 y_2 \ldots y_q (q > 1)$ and (with y_{q+1} denoting y_1)

$P(\bar{u})$ = number of y_m's $(1 \leqq m \leqq q)$ for which y_m^{-1} comes *before* y_{m+1} in the ordering

$$a_1, b_1^{-1}, a_1^{-1}, b_1, a_2, b_2^{-1}, a_2^{-1}, b_2, \ldots, a_g^{-1}, b_g, s_1, s_1^{-1}, s_2, \ldots, s_r^{-1}$$

for M orientable, or the ordering

$$u_1, u_1^{-1}, u_2, u_2^{-1}, \ldots, u_1, u_n^{-1}, s_1, s_1^{-1}, s_2, \ldots, s_r^{-1}$$

for M non-orientable,

$\bar{A}(\bar{u})$ = number of y_m's which are a_i^{-1} for some i,

$B(\bar{u})$ = number of y_m's which are b_j for some j, and

$\bar{S}(\bar{u})$ = number of y_m's which are s_k^{-1} for some k.

If $q = 1$ we define $T(\bar{u}) = 0$ if $\bar{u} \neq u_i^{-1}$ $(1 \leqq i \leqq n)$, or $T(\bar{u}) = 1 \in \mathbb{Z}_2$ if $\bar{u} = u_i^{-1}$, since $\omega_X(c) + \omega_X(c^{-1}) = 0$ or 1 according to whether c contains orientation-preserving or -reversing curves (see Proposition (2.2) of [7]).

There are similar expressions involving $\omega_X(v^{-1})$ and $\omega_X(uv^{-1})$.

Now $\varphi_X(\bar{u}) = \varphi_X(u)$ etc., and

$$\varphi_X(uv^{-1}) = \varphi_X(u) + \varphi_X(v^{-1})$$

since φ_X is a homomorphism. The same applies to φ'_X. Hence the formula

$$\omega_X(uv^{-1}) = \omega_X(u) + \omega_X(v^{-1})$$

reduces to

$$T(\overline{uv^{-1}}) = T(\bar{u}) + T(\overline{v^{-1}})$$

in the orientable case, or

$$T(\overline{uv^{-1}}) \equiv T(\bar{u}) + T(\overline{v^{-1}}) \, mod\, 2$$

in the non-orientable case, and is patently independent of the vector field X. Therefore we see that condition (iii) in the algorithm may be replaced by:

(iii)′ *W contains at least 2 letters, $W \neq V^m$ for any V and for any $m \geqq 2$, and*

$$T(\overline{uv^{-1}}) \equiv T(u) + T(\overline{v^{-1}}) \, mod\, 2$$

for every division $(W) = (uv)$.

A few examples illustrating the working of the algorithm are given in [6].

§ 4. Application of Winding Numbers to Homeomorphisms of Surfaces

By [9], [11, 13], and [12, 14], [5] it is known that any homeomorphism of a closed surface M may be expressed as a product of

(i) homeomorphisms isotopic to the identity;

(ii) "twists" (defined in [11]) about members of a finite family of simple closed curves as shown in Fig. 3 (M orientable) or in [5, Fig. 2] (M non-orientable);

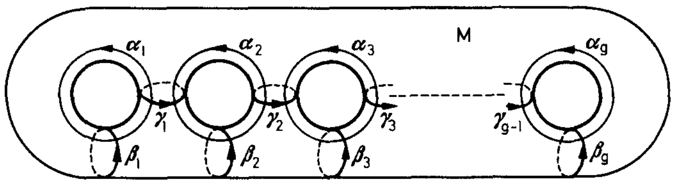

Fig. 3

and if M is orientable

(iii) a fixed orientation-reversing homeomorphism;

or if M is non-orientable

(iii)′ a Y-homeomorphism, as defined in [12].

A homeomorphism h keeping x fixed induces an automorphism $h_*: \pi_1(M, x) \to \pi_1(M, x)$, and if x is not fixed then h_* can be defined only up to inner automorphism. However, if C is a conjugacy class of elements in $\pi_1(M, x)$ then $h_* C$ is a well-defined conjugacy class. If $\omega_X(k_* C)$ is known[2] (for some vector field X) whenever k is a homeomorphism isotopic to the identity, to a twist or to a Y-homeomorphism, then a knowledge of $\omega_X(h_* C)$ for a general h leads to information about the composition of h in terms of twists, etc. In particular, taking $h =$ identity may throw some light on the problem of finding a set of relations for the generators of the *homeotopy group* or *mapping class group* Λ_M which is the group of homeomorphisms of M factored by those isotopic to the identity. (See e.g. [1, 2] for some recent results relating to this problem.)

The following three lemmas concerning the effects of homeomorphisms on winding numbers are given without proof. The first is trivial (since $k_* C = C$), the second is elementary, and the third is a little more complicated technically but essentially a matter of straighforward verification. All non-orientable surfaces will be supposed to have even genus so that winding numbers of conjugacy classes are defined, and vector fields will be assumed to be *continuous and non-vanishing* on M minus a point.

[2] We recall from [7, § 6] that for a closed surface M winding numbers of conjugacy classes are defined as elements of $\mathbf{Z}_{2g-2}$ (M orientable) or $\mathbf{Z}_2$ (M non-orientable, and of *even* genus), and vector fields are taken to be defined on M minus one point.

(4.1) Lemma. *If k is isotopic to the identity then $\omega_X(k_* C) = \omega_X(C)$ for all C.*

(4.2) Lemma. *Let B, C be conjugacy classes in $\pi_1(M, x)$, where B is represented by a simple closed curve β, and let h_β be a twist about β. Denote by B', C' the homology classes corresponding to B, C with coefficients in $\mathbb{Z}$ if M is orientable (and oriented) or in $\mathbb{Z}_2$ if M is non-orientable, and let $n(\in \mathbb{Z}$ or $\mathbb{Z}_2)$ be the intersection-number $(B' \cdot C')$. Then*

$$\omega_X(h_{\beta *} C) = \omega_X(C) + n\omega_X(B)$$

as an element of $\mathbb{Z}_{2g-2}$ or $\mathbb{Z}_2$ as appropriate.

(4.3) Lemma. *Let y be a Y-homeomorphism defined using a space $Y \subset M$ (see [12]) and let μ_1, μ_2 be two disjoint simple closed orientation-reversing curves on Y whose $\mathbb{Z}_2$-homology classes generate $H_1(Y; \mathbb{Z}_2)$. Then*

$$\omega_X(y_* C) = \omega_X(C) + n_1 + n_2$$

where $n_i \in \mathbb{Z}_2$ $(i = 1, 2)$ is the intersection-number of the homology class C' with the homology class of μ_i.

The following result has a proof similar to that of Lemma (4.2) but is more elementary. These Lemmas (4.2) and (4.4) are mutually independent.

(4.4) Lemma. *With the notation of Lemma* (4.2), $(h_{\beta *} C)' = C' + nB'$.

We now apply the foregoing Lemmas to the word problem in Λ_M.

Let M be a closed orientable surface of genus g. By [13], the group G_M of homeomorphisms of M is generated by an orientation-reversing homeomorphism, homeomorphisms isotopic to the identity, and twists about curves α_i $(1 \leqq i \leqq g)$, β_i $(1 \leqq i \leqq g)$ and γ_i $(1 \leqq i \leqq g-1)$ as in Fig. 3. Let K denote the union of all these curves, and let N be a regular neighbourhood of K in M, i.e. a smooth closed neighbourhood of K which contracts to K. All twists may be assumed to take place in N. If $h \in G_M$ let $\{h\} \subset \Lambda_M$ denote the isotopy class of h.

Observe that $H_1(M; \mathbb{Z})$ is the free abelian group generated by the homology classes of the α_i and β_i, and $H_1(N; \mathbb{Z})$ is similarly generated by the homology classes of the α_i, β_i and γ_i.

(4.5) Lemma. *Let $h \in G_M$ and let k be a homeomorphism of N given by $k = h_1 \dots h_r$ where $\{h\} = \{h_1\} \dots \{h_r\}$, and each h_i is a "generating" twist about a curve in K. Choose one of the α_i, say α_j, and suppose*

$$(k\alpha_j)'' = \sum_{i=1}^{g} \lambda_i \alpha_i'' + \sum_{i=1}^{g} \mu_i \beta_i'' + \sum_{i=1}^{g-1} \nu_i \gamma_i''$$

where here " denotes homology class in N, and the curves of K are oriented as shown in Fig. 3. *Then*

(i) *if the homeomorphism h of M induces the identity on* $H_1(M;\mathbb{Z})$ *we have*

$$\sum_{i=1}^{g} \mu_i = 0, \quad \lambda_j = 1, \quad \lambda_i = 0\,(i \neq j)$$

and (ii) *if h is isotopic to the identity the above equations hold and also*

$$\sum_{i=1}^{g-1} (g-i)\,\mu_i \equiv 0 \mod 2g-2.$$

Proof. On M we have $\beta_i - \beta_{i+1} \sim -\gamma_i$ $(1 \leqq i \leqq g-1)$ and so in case (i) since $h\alpha_j \sim \alpha_j$ on M it follows by equating coefficients of homology classes in M that $\lambda_j = 1$, $\lambda_i = 0$ $(i \neq j)$ and

$$\begin{aligned} 0 &= \mu_1 - \nu_1 \\ 0 &= \mu_2 + \nu_1 - \nu_2 \\ &\vdots \\ 0 &= \mu_{g-1} + \nu_{g-2} - \nu_{g-1} \\ 0 &= \mu_g + \nu_{g-1}. \end{aligned} \tag{$*$}$$

Hence by addition $\mu_1 + \cdots + \mu_g = 0$.

By Lemmas (4.2) and (4.4) it follows that for any vector field X on M minus one point we have

$$\omega_X(k\alpha_j) = \sum_{i=1}^{g} \lambda_i \omega_X(\alpha_i) + \sum_{i=1}^{g} \mu_i \omega_X(\beta_i) + \sum_{i=1}^{g-1} \nu_i \omega_X(\gamma_i)$$

where we may assume the α_i etc. to be regular and k to be smooth. Since β_i, γ_i and β_{i+1}^{-1} are the boundary components of a surface of genus zero in M it follows by [7, Lemma (5.7) and § 6] that

$$\omega_X(\beta_i) + \omega_X(\gamma_i) - \omega_X(\beta_{i+1}) \equiv 1 \mod 2g-2.$$

Now the $\omega_X(\alpha_i)$ and $\omega_X(\beta_i)$ can be given arbitrary values $mod\, 2g-2$ by a suitable choice of X (see [7]; in particular Theorem (5.6) and the end of § 6). In case (ii) since $h\alpha_j \simeq \alpha_j$ we have

$$\omega_X(k\alpha_j) \equiv \omega_X(\alpha_j) \mod 2g-2$$

(see Theorem (5.3) of [7]). Therefore by equating coefficients of the $\omega_X(\alpha_i)$ and $\omega_X(\beta_i)$ the Eqs. (*) of case (i) arise again mod$2g-2$ and we obtain also

$$\sum_{i=1}^{g-1} \nu_i \equiv 0 \quad mod\, 2g-2.$$

Combining this with (*) gives the result.

Note. In view of the note at the end of § 5 in [7], the hypothesis (i) of Lemma (4.5) suffices to prove the entire result modulo 2.

These equations may be used to give information about Λ_M in some particular cases as follows.

(4.6) Corollary. *If $g=2$ it is impossible to express the isotopy class of the twist h_{β_j} ($j=1$ or 2) in terms of the other generators of Λ_M.*

(In fact generators and relations for Λ_M are known when $g=2$; see [2].)

Proof. Suppose $\{h_{\beta_j}\}=\{h\}$ where j is either 1 or 2 and $\{h\}=\{h_1\}\dots\{h_r\}$ where none of the h_i are $h_{\beta_j}^{\pm 1}$. Let $k=h_1\dots h_r$ and apply the lemma to $k^{-1}h_{\beta_j}$ and α_j. The Eqs. (i) and (ii) are

$$\lambda_j=1\,,\quad \lambda_i=0\,(i\neq j)$$
$$\mu_1+\mu_2=0$$
$$\mu_1\equiv 0 \; mod\, 2$$

where μ_i is the coefficient of β_i'' in $(k^{-1}h_{\beta_j}\alpha_j)''$. But $\mu_j=\pm 1$ (depending on the orientation of M) since α_j crosses β_j once, and this is a contradiction.

(4.7) Corollary. *If $g=3$ it is impossible to express $\{h_{\beta_2}\}$ in terms of the other generators, and if $\{h_{\beta_j}\}=\{h\}$ where $h_{\beta_j}^{\pm 1}$ does not appear in $h(j=1,3)$ then $h_{\beta_{4-j}}$ appears in h.*

(The first (and probably both) of these facts can also be easily deduced from known results about generators and relations in Λ_M; see [2].)

Proof. Applying the lemma to $k^{-1}h_{\beta_j}$ and α_j as above gives the equations $\mu_1+\mu_2+\mu_3=0$ and $2\mu_1+\mu_2\equiv 0 \; mod 4$. These cannot hold when $\mu_2=\pm 1$ or when $\mu_1+\mu_3=\pm 1$.

As g increases, the information about Λ_M obtained in this way becomes more complicated and correspondingly less useful. Nevertheless, these examples illustrate at least the possibility of using winding numbers to obtain algebraic results. The possibility would become more real if the structure of $\omega_X:\pi_1(M,x)\to\mathbb{Z}_{2g-2}$ were better understood.

Some similar results may be derived when M is non-orientable; if n is even, using Lemma (4.3). When n is odd an isotopy will not in general preserve winding numbers, and the above methods cannot be used.

If γ is a simple closed curve homologous to zero ($\mathbb{Z}$) on M, which we take to be orientable, then (cf. Lemma (4.2)) a twist about γ preserves the homology class and the winding number with respect to any X of any other curve on M. Since homeomorphisms exist which preserve homology but not winding numbers (see example below) it follows that homology-preserving homeomorphisms are not in general expressible as products of twists about curves which are homologous to zero. However, the following conjecture may be true. If so it would provide a useful link between the geometric and algebraic analyses of homeomorphisms.

Conjecture. *If M is orientable ($\partial M \neq \emptyset$) with a given continuous non-vanishing vector field X on M, and h is a diffeomorphism of M such that*

(i) $h_*: H_1(M;\mathbb{Z}) \to H_1(M;\mathbb{Z})$ *is the identity and*

(ii) $\omega_X(h\gamma) = \omega_X(\gamma)$ *for all regular curves γ on M,*

then h is a product of twists about curves which are homologous to zero relative to ∂M.

Condition (ii) could be replaced by:

(ii)′ $\omega_X(h_* C) = \omega_X(C)$ *for all conjugacy classes C in $\pi_1(M, x)$.*

If $\partial M = \emptyset$ (where X is defined on M minus one point) then the conjecture with (ii)′ is the more natural but is easily seen to be false, since $\omega_X(C)$ is an element only of $\mathbb{Z}_{2g-2}$. For example, consider two simple closed regular curves γ_1, γ_2, forming the oriented boundary of a surface N of genus 2 contained in a closed orientable surface M of genus 3. Then $\omega_X(\gamma_1) + \omega_X(\gamma_2) \equiv 0 \mod 4$ by [7, Lemma (5.7)]. Suppose $\gamma_1 \not\sim 0$, and choose X such that $\omega_X(\gamma_1) = 3$, $\omega_X(\gamma_2) = 1$ (the singularity of X may be chosen not to lie on N). Let $h = h_{\gamma_1}^2 h_{\gamma_2}^{-2}$. Then h preserves homology and preserves $\omega_X(C)$ ($\in \mathbb{Z}_4$) for all C, but if γ meets γ_1 and γ_2 once each then $\omega_X(h\gamma) = \omega_X(\gamma) \pm 4$, so h is not a product of twists about curves homologous to zero.

References

1. Birman, Joan: Automorphisms of the fundamental group of a closed, orientable 2-manifold. Proc. Amer. Math. Soc. **21**, 351—354 (1969).
2. Birman, J., Hilden, M.: Mapping class groups of closed surfaces as covering spaces. Annals of Math. Studies **66**, Princeton 1971.
3. Călugăreanu, G.: Sur les courbes fermées simples tracées sur une surface fermée orientable. Mathematika (Cluj) **8(31)**, 29—38 (1966).
4. Chillingworth, D. R. J.: Thesis, University of Cambridge 1968.
5. Chillingworth, D. R. J.: A finite set of generators for the homeotopy group of a non-orientable surface. Proc. Cambridge Philos. Soc. **65**, 409—430 (1969).
6. Chillingworth, D. R. J.: Simple closed curves on surfaces. Bull. London Math. Soc. **1**, 310—314 (1969).
7. Chillingworth, D. R. J.: Winding numbers on surfaces, I. Math. Ann. **196**, 218—249 (1972).
8. Dehn, M.: Transformation der Kurven auf zweiseitigen Flächen. Math. Ann. **72**, 413—421 (1912).
9. Dehn, M.: Die Gruppen der Abbildungsklassen. Acta Math. **69**, 135—206 (1936).
10. Epstein, D. B. A.: Curves on 2-manifolds and isotopies. Acta Math. **115**, 83—107 (1966).
11. Lickorish, W. B. R.: A representation of orientable combinational 3-manifolds. Ann. Math. **76**, 531—540 (1962).

12. Lickorish, W. B. R.: Homeomorphisms of non-orientable two-manifolds. Proc. Cambridge Philos. Soc. **59**, 307—317 (1963).
13. Lickorish, W. B. R.: A finite set of generators for the homeotopy group of a 2-manifold. Proc. Cambridge Philos. Soc. **60**, 769—778 (1964). Corrigendum, Proc. Cambridge Philos. Soc. **62**, 679—681 (1966).
14. Lickorish, W. B. R.: On the homeomorphisms of a nonorientable surface. Proc. Cambridge Philos. Soc. **61**, 61—64 (1965).
15. Reidermeister, K.: Einführung in die kombinatorische Topologie. New York: Chelsea Publishing Company 1950.
16. Reinhart, B. L.: The winding number on two manifolds. Ann. Inst. Fourier (Grenoble) **10**, 271—283 (1960).
17. Reinhart, B. L.: Algorithms for Jordan curves on compact surfaces. Ann. Math. **75**, 209—222 (1962).
18. Reinhart, B. L.: Further remarks on the winding number. Ann. Inst. Fourier (Grenoble) **13**, 155—160 (1963).
19. Zieschang, H.: Algorithmen für einfache Kurven auf Flächen. Math. Scand. **17**, 17—40 (1965).
20. Zieschang, H.: Über Automorphismen ebener diskontinuierlicher Gruppen. Math. Ann. **166**, 148—167 (1966).
21. Zieschang, H.: Algorithmen für einfache Kurven auf Flächen, II. Math. Scand. **25**, 49—58 (1969).
22. Zieschang, H., Vogt, E., Coldewey, H.-D.: Flächen und ebene diskontinuierliche Gruppen. Lecture Notes in Mathematics, No. **122**. Berlin-Heidelberg-New York: Springer 1970.

Dr. D. R. J. Chillingworth
Department of Mathematics
University of Southampton
Southampton, S 509 5 NH
England

(Received April 15, 1970/April 12, 1972)