# 1. Introduction

Objekttyp: **Chapter**

Zeitschrift: **L'Enseignement Mathématique**

Band (Jahr): **43 (1997)**

Heft 3-4: **L'ENSEIGNEMENT MATHÉMATIQUE**

PDF erstellt am: **12.07.2024**

*L'Enseignement Mathématique*, t. 43 (1997), p. 207–252



# AN ALGORITHM FOR CELLULAR MAPS OF CLOSED SURFACES

by Warren DICKS and H. H. GLOVER

ABSTRACT. The purpose of this article is to use diagrammatic methods to give proofs, accessible to algebraists, of some important topological results of H. Kneser, A. L. Edmonds, and R. Skora; we then describe some consequences for homomorphisms between surface groups. Cellular maps between two-dimensional CW-complexes can be represented by diagrams which, in turn, can be interpreted algebraically in terms of fundamental groupoids. For diagrams representing cellular maps between closed surfaces, we show how to apply certain homotopy equivalences algorithmically to obtain a normal-form map, which is a branched covering, or a pinching followed by a covering, or a map which collapses a graph of punctured spheres to a graph immersed in the one-skeleton of the target surface. We then indicate how the algorithm can be expressed entirely in terms of formal manipulations with presentations of surface groupoids, yielding algebraic proofs of results about homomorphisms between surface groups.

## 1. INTRODUCTION

We begin by recalling some basic concepts.

1.1. DEFINITIONS. Let $\beta$ be a map between closed surfaces (without boundary).

Then $\beta$ is a *branched covering* if deleting finitely many points from the source and from the target yields a covering.

We say that $\beta$ is a (possibly trivial) *pinching* if it is obtained by collapsing, to a point, a compact subsurface with a single boundary component.

The (geometric) *degree* of $\beta$, denoted $\mathcal{G}(\beta)$, is the least non-negative integer $d$ such that there is a map $\beta'$ homotopic to $\beta$, such that the inverse image under $\beta'$ of some 2-disk consists of $d$ 2-discs, each mapped homeomorphically by $\beta'$ to the chosen disk.

Results of H. Kneser [11, p. 354], A.L. Edmonds [3], and R. Skora [16] show that if $\mathcal{G}(\beta)$ is nonzero, then $\beta$ is homotopic either to a branched covering or to the composite of a pinching followed by a covering. In the case where $\mathcal{G}(\beta)$ is zero, $\beta$ is homotopic to a map which is not surjective. Thus one of the following holds:

(a) $\beta$ is homotopic to a map which is not surjective;

(b) $\beta$ is homotopic to the composite of a pinching followed by a covering;

(c) $\beta$ is homotopic to a branched covering.

In case (a), $\mathcal{G}(\beta) = 0$, and in case (b) (resp. (c)), $\mathcal{G}(\beta)$ is given by the degree of the covering (resp. branched covering).

This allows one to compute the degree via homological means, which is the essence of a classical result of Kneser [10], [11]. Edmonds and Skora further discuss cases where the surfaces are not necessarily closed, but we wish to restrict our attention to the closed case.

The main purpose of this article is to prove these Kneser-Edmonds-Skora results using diagrammatic techniques developed by van Kampen, Lyndon, and Ol'shanskii. We shall give an algorithm which applies homotopy equivalences to a cellular map between closed surfaces, and yields a map in normal form, that, in the non-zero degree case, is a branched covering, or, after pinching, is a covering, while, in the degree zero case, the source surface is expressed as a union of spheres with various punctures based at the poles, and these punctured spheres are collapsed to arcs, to give a graph immersed in the one-skeleton of the target surface. Recall that a graph map *immersion* is a locally injective graph map, so that the induced map of fundamental groups is injective. In particular, the algorithm yields the degree of the map.

In the non-zero degree case, we then have the Kneser-Edmonds-Skora result, and, in the zero degree case, we recover preliminary steps towards results previously obtained by several authors, notably Zieschang [17], Edmonds, Skora, Ol'shanskii [15], and Grigorchuk and Kurchanov [7]. The present article is very much in the spirit of Ol'shanskii's article.

The proofs by Edmonds and Skora are brief, simple, direct, and essentially algorithmic, but are not readily expressible in algebraic terms. Our proof, although substantially longer, has the feature that it deals throughout with closed surfaces, without cutting them up, and uses elementary homotopy operations which readily lend themselves to algebraic intrepretation. So we claim that we have fulfilled our main objective of giving algebraic proofs of the substantial group-theoretic consequences of these topological theorems. There is a natural motivation to have algebraic proofs of algebraic results, especially when they are obtained topologically, and our work fits into this scheme in a useful way; for example, it can be used in the (algebraic) proof of Theorem 4.9 of [2].

To give an idea of the sort of algebraic consequences of the algorithm, it is convenient to introduce some terminology.

1.2. DEFINITIONS. Recall that a group $G$ is called a *surface group* if $G$ is the fundamental group of a closed surface, or, equivalently, $G$ has a *surface group presentation*, by which we mean a one-relator presentation $\langle S \mid r\rangle$ such that $r$ is cyclically reduced, and each element of $S$ occurs exactly twice in $r$, with exponent 1 or $-1$, and the face-adjacency relation on $S \cup S^{-1}$ has only one equivalence class. Here face-adjacency is the equivalence relation determined by identifying $s_1^{\epsilon_1}$ and $s_2^{\epsilon_2}$ whenever $s_1^{\epsilon_1} s_2^{-\epsilon_2}$ occurs in the cyclic expression of $r$.

It follows that $G$ is a surface group if and only if $G$ has a presentation

$$\left\langle x_1, y_1, \ldots, x_n, y_n, z_1, \ldots z_m \;\middle|\; (x_1, y_1) \cdots (x_n, y_n) z_1^2 \cdots z_m^2 \right\rangle,$$

for some non-negative integers $m$, $n$, and here one can arrange that either $m$ or $n$ is zero. Recall that $(x, y)$ denotes $\bar{x}\bar{y}xy$, where overlines denote inverses.

There is an associated *orientation map* $\epsilon\colon G \to \{\pm 1\}$ which, for the latter presentation, sends the $x_i$ and the $y_i$ to 1, and the $z_j$ to $-1$. The kernel of $\epsilon$ is denoted $G^+$. We say $G$ is *orientable* (resp. *unorientable*) if $(G : G^+)$ is 1 (resp. 2). Thus a surface group is orientable if it is the fundamental group of an orientable closed surface. By the *orientation module* we mean the $\mathbf{Z}G$-module $\Omega$ which consists of the abelian group $\mathbf{Z}$, on which each $g \in G$ acts as $\epsilon(g) \in \{\pm 1\} = \mathrm{Aut}(\mathbf{Z})$.

The finite surface groups have order 1 or 2, and are the fundamental groups of the two-sphere and the projective plane, respectively. The infinite surface groups correspond to the surface group presentations in which the relator has length at least 4, and these are the fundamental groups of the *aspherical* closed surfaces, that is, closed surfaces whose universal covers are contractible.

A homotopy class of continuous maps between path-connected topological spaces determines an equivalence class of homomorphisms between their fundamental groups, where equivalence corresponds to composition with an inner automorphism of the target group. For aspherical closed surfaces, this correspondence between equivalence classes of morphisms is bijective, so, in quite a strong sense, the study of homotopy classes of continuous maps between aspherical closed surfaces is much the same as the study of homomorphisms between infinite surface groups.

1.3. DEFINITIONS. Let $\alpha\colon G_1 \to G_2$ be a homomorphism of surface groups.

If $\alpha$ together with the orientation maps of $G_1$ and $G_2$ form a commuting triangle, we say that $\alpha$ is *orientation-true*, and otherwise $\alpha$ is *orientation-false*. Thus $\alpha$ is orientation-true if and only if $\alpha^{-1}(G_2^+) = G_1^+$.

For any surface group presentations $G_1 = \langle S_1 \mid r_1\rangle$, $G_2 = \langle S_2 \mid r_2\rangle$, there exists a homomorphism of free groups $A\colon \langle S_1 \mid \ \rangle \to \langle S_2 \mid \ \rangle$ which induces $\alpha$, and then there exist a non-negative integer $d$, elements $w_1, \dots, w_d$ of $\langle S_2 \mid \ \rangle$, and elements $\epsilon_1, \dots, \epsilon_d$ in $\{1, -1\}$, such that, in $\langle S_2 \mid \ \rangle$,

$$A(r_1) = \prod_{i=1}^{d} w_i r_2^{\epsilon_i} w_i^{-1}. \tag{1.1}$$

The *degree* of $\alpha$, denoted $\mathcal{G}(\alpha)$, is the least value of $d$ which occurs as we range over all the possible choices at our disposal. If $G_1$ or $G_2$ is finite, this concept is rather degenerate and we shall not be discussing this case. If $G_1$ and $G_2$ are infinite, the algorithm given in this article provides a lifting $A'\colon \langle S_1 \mid \ \rangle \to \langle S_2 \mid \ \rangle$ of $\alpha$, and an expression of $A'(r_1)$ as a product of $\mathcal{G}(\alpha)$ conjugates of $r_2^{\pm 1}$.

Notice that if $\mathcal{G}(\alpha) = 0$, then $\alpha$ factors through the natural surjection $\langle S_2 \mid \ \rangle \to \langle S_2 \mid r_2\rangle$; conversely, if $\alpha$ factors through any map from a free group $F$ to $\langle S_2 \mid r_2\rangle$, we can use the freeness of $F$ to factor this map through the natural surjection. Thus $\mathcal{G}(\alpha) = 0$ if and only if $\alpha$ factors through a free group $F$. By replacing $F$ with the image of $\alpha$ in $F$, we see that $\mathcal{G}(\alpha) = 0$ if and only if $\alpha$ factors through a surjective map to a free group.

Kneser's homological calculation of the degree, in the formulation of Skora [16], yields the following.

1.4. THEOREM (Kneser [10], [11]). *Let $\alpha\colon G_1 \to G_2$ be a homomorphism of infinite surface groups, and consider an equation* (1.1) *arising from some lifting of $\alpha$.*

(i) *If $\alpha$ is orientation-true, then $\mathcal{G}(\alpha) = \left|\sum_{i=1}^{d} \epsilon_i\, \epsilon(w_i)\right|$, where the map $\epsilon\colon \langle S_2 \mid \ \rangle \to \{\pm 1\}$ is induced from the orientation map of $G_2$.*

(ii) *If $\alpha$ is orientation-false, and either $d$ is even, or the index $(G_2 : \operatorname{Im}\alpha)$ is infinite, then $\mathcal{G}(\alpha) = 0$.*

(iii) *If $\alpha$ is orientation-false, and $d$ is odd, and $(G_2 : \operatorname{Im}\alpha)$ is finite, then $\mathcal{G}(\alpha) = (G_2 : \operatorname{Im}\alpha)$.*

1.5. REMARK. Under pullback along $\alpha$, the orientation module $\Omega_2$ for $G_2$ becomes a $G_1$-module, again denoted $\Omega_2$, and $\alpha$ induces a change of groups map in cohomology $H^2(\alpha, \Omega_2)\colon H^2(G_2, \Omega_2) \to H^2(G_1, \Omega_2)$. By Poincaré duality, $H^2(G_2, \Omega_2) \simeq \Omega_2 \otimes_{\mathbf{Z}G_2} \Omega_2 \simeq \mathbf{Z}$ with trivial $G_2$-action, and

$$H^2(G_1, \Omega_2) \simeq \Omega_1 \otimes_{\mathbf{Z}G_1} \Omega_2 \simeq \begin{cases} \mathbf{Z} & \text{if } \alpha \text{ is orientation-true,} \\ \mathbf{Z}/2\mathbf{Z} & \text{if } \alpha \text{ is orientation-false,} \end{cases}$$

with trivial $G_1$-action. Using a lifting $A$ and an equation (1.1), and techniques such as those used in the proof of Theorem V.4.9 of [1], one can calculate that, up to sign, $H^2(\alpha, \Omega_2)$ acts as multiplication by $\sum_{i=1}^{d} \epsilon_i\, \epsilon(w_i)$.

Hence, if $\alpha$ is orientation-true, the non-negative integer $\left|\sum_{i=1}^{d} \epsilon_i\, \epsilon(w_i)\right|$ which occurs in Theorem 1.4 (a) is independent of the lifting chosen to get (1.1), and the theorem says that, in this case, there exists a lifting such that all the $\epsilon_i \epsilon(w_i)$ are equal.

Even if $\alpha$ is orientation-false, the parity of $d$ (which is the parity of $\sum_{i=1}^{d} \epsilon_i\, \epsilon(w_i)$) is independent of the lifting chosen to get (1.1), and will be called the *parity* of $\alpha$, which is either *even* or *odd*.

In particular, if $\alpha$ is any homomorphism of infinite orientable surface groups, and $G_1 = \langle S_1 \mid r_1 \rangle$, $G_2 = \langle S_2 \mid r_2 \rangle$, are surface group presentations, then there exists a homomorphism of free groups $A\colon \langle S_1 \mid \ \rangle \to \langle S_2 \mid \ \rangle$ which induces $\alpha$, such that $A(r_1)$ is a product of $\mathcal{G}(\alpha)$ conjugates of $r_2$ (or of $r_2^{-1}$), with no conjugate of $r_2^{-1}$ (resp. $r_2$) needed in this expression.

The Kneser-Edmonds-Skora results give even more information about homomorphisms between infinite surface groups, but we shall postpone making the precise statements until Section 4.

In outline, the paper is structured as follows. In Section 2, we present some of the terminology we will use, describe some preliminary constructions, and recall how to associate, with a homomorphism between surface groups, a cellular map between surfaces which realizes the homomorphism. A cellular map between surfaces can be visualized as a labelled diagram, and, in Section 3, we give the algorithm for homotoping a diagram until a normal form is reached. In Section 4, we describe consequences for group homomorphisms, such as Kneser's Theorem determining degrees, and Nielsen's Theorem [14, Section 26] that every automorphism of a surface group lifts to an automorphism of the covering free group which sends the surface relator to a conjugate of itself or its inverse. In Section 5 we indicate how the algorithm can be described in terms of formal manipulations of presentations of surface groupoids, by describing a trivial example which illustrates the algorithm.

## 2. DIAGRAMS OF CELLULAR SURFACE MAPS

In this section we introduce the setting in which we shall work, and describe the connection with group theory.

2.1. DEFINITIONS. By a *two-dimensional CW-complex* $X$ we shall mean a combinatorial CW-complex of dimension at most two, in which each cell has a preferred orientation. Formally we have the following situation.

As a set, $X$ is the disjoint union of three sets $V$, $E$, $F$, whose elements are called the *vertices*, *edges*, and *faces*, of $X$, respectively.

There are given maps $\iota$, $\tau$, from $E$ to $V$, and, for each edge $e$, the vertices $\iota e$, $\tau e$ are called the *initial* and *terminal* vertices of $e$, respectively. If $\iota e = \tau e$, we say that $e$ is a *loop*. For each vertex $v$ we understand that $\iota v = v = \tau v$.

We write $E^{\pm 1}$ for the Cartesian product $E \times \{1, -1\}$, and for any $(e, \epsilon) \in E^{\pm 1}$ we write $e^{\epsilon}$ for $(e, \epsilon)$. We identify $e = e^1$. We use the same conventions for the faces. For a vertex $v$, we understand that $v^1 = v = v^{-1}$.

For $e \in E$, we define $\iota(e^{-1}) = \tau e$, and $\tau(e^{-1}) = \iota e$.

Each face $f$ of $X$ has an associated *boundary cycle* which is a finite alternating sequence

$$\partial f = v_0, e_1^{\epsilon_1}, v_1, \ldots, v_{n-1}, e_n^{\epsilon_n}, v_n, \tag{2.1}$$

where $n \geq 0$, the $v_i$ are vertices, $v_n = v_0$, the $e_i$ are edges, each $\epsilon_i$ is $\pm 1$, and $\iota(e_i^{\epsilon_i}) = v_{i-1}$, $\tau(e_i^{\epsilon_i}) = v_i$. We define

$$\partial f^{-1} = v_n, e_n^{-\epsilon_n}, v_{n-1}, \ldots, v_1, e_1^{-\epsilon_1}, v_0 .$$

It is thus implicit that we are assigning to each closed two-cell a polygonal structure, and a distinguished vertex where the boundary cycle begins and ends. Notice that we are allowing vertices of valence one, so a boundary cycle need not be reduced.

A one-dimensional CW-complex, that is, a two-dimensional CW-complex with no faces, will be called a *graph*.