Applicable Algebra in Engineering, Communication and Computing manuscript No.
(will be inserted by the editor)

Computing fundamental groups from point clouds

Piotr Brendel · Paweł Dłotko · Graham Ellis · Mateusz Juda · Marian Mrozek

the date of receipt and acceptance should be inserted later

Abstract We describe an algorithm for computing a finite, and typically small, presentation of the fundamental group of a finite regular CW-space. The algorithm is based on the construction of a discrete vector field on the 3-skeleton of the space. A variant yields the homomorphism of fundamental groups induced by a cellular map of spaces. We illustrate how the algorithm can be used to infer information about the fundamental group $\pi_1(K)$ of a metric space K using only a finite point cloud X sampled from the space. In the special case where K is a d-dimensional compact manifold $K \subset \mathbb{R}^d$, we consider the closure of the complement of K in the d-sphere $M_K = \overline{\mathbb{S}^d \setminus K}$. For a base-point x in the boundary ∂M_K of the manifold M_K one can attempt to determine, from the point cloud X, the induced homomorphism of fundamental groups $\phi\colon \pi_1(\partial M_K, x) \to \pi_1(M_K, x)$ in the category of finitely presented groups. We illustrate a computer implementation for K a small closed tubular neighbourhood of a tame knot in $\mathbb{R}^3$. In this case the homomorphism ϕ is known to be a complete ambient isotopy invariant of the knot. We observe that low-index subgroups of finitely presented groups provide useful invariants of ϕ. In particular, the first integral homology of subgroups $G < \pi_1(M_K)$ of index at most 6 suffices to distinguish between all prime knots with eleven or fewer crossings (ignoring chirality). We plan to provide formal time estimates for our algorithm and characteristics of a high performance C++ implementation in a subsequent paper. The prototype computer implementation of the present paper has been written in the interpreted GAP programming language for computational algebra.

1 Introduction

Let X be a finite point cloud sampled from a metric space K. Suppose that the distance between points in X is given but that limited further information about K is given. We are interested

P.D. is supported by the grant DARPA: FA9550-12-1-0416 and AFOSR: FA9550-14-1-0012. GE was partially supported by the European Science Foundation network on *Applied and Computational Algebraic Topology* and by Polish MNSzW, Grant N N201 419639. GE thanks the IST, Austria for its hospitality during the writing of this paper. MM was partially supported by Polish MNSzW, Grant N N201 419639.

Piotr Brendel
Division of Computational Mathematics, Jagiellonian University, ul. St. Łojasiewicza 6, 30-348 Kraków, Poland

Paweł Dłotko
Department of Mathematics, University of Pennsylvania, 209 s. 33rd st., Philadelphia, PA 19104-6395, USA

Graham Ellis
School of Mathematics, National University of Ireland, Galway, Ireland E-mail: graham.ellis@nuigalway.ie

Mateusz Juda
Division of Computational Mathematics, Jagiellonian University, ul. St. Łojasiewicza 6, 30-348 Kraków, Poland

Marian Mrozek, Division of Computational Mathematics, Jagiellonian University, ul. St. Łojasiewicza 6, 30-348 Kraków, Poland E-mail: Marian.Mrozek@ii.uj.edu.pl

in how to infer information on the fundamental group $\pi_1(K)$ using only the data X. Following the standard approach to topological data analysis we associate to X a filtered regular CW-space $K_0^X \subseteq K_1^X \subseteq K_2^X \subseteq \cdots \subseteq K_N^X$ with K_0^X a discrete space and K_N^X a contractible space. The spaces K_i^X are constructed (in one of several possible ways) so that we could expect 'reasonable' metric spaces K to admit a self homotopy equivalence of the form $K \to K_s^X \subseteq K_t^X \to K$ for some large range $s < t$. We do not expect the spaces K_s^X and K_t^X to be homotopy equivalent in general. Tools such as persistent homology can help in identifying a suitable range $s < t$. The image of the homomorphism $\psi_{st}\colon \pi_1(K_s^X) \to \pi_1(K_t^X)$ is an heuristic approximation to $\pi_1(K)$. We describe a procedure that inputs the data X, determines a suitable range $s < t$, and outputs ψ_{st} as a homomorphism of finitely presented groups. The crux of the procedure is an algorithm for computing a finite, and typically small, presentation of the fundamental group of a finite regular CW-space from a discrete vector field on the 3-skeleton of the space. We illustrate an implementation on a point cloud $X \subset \mathbb{R}^3$ and explain how low-index subgroups of finitely presented groups can be used to determine useful information about ψ_{st}.

When K is a d-dimensional compact manifold $K \subset \mathbb{R}^d$ we consider the closure $M = \overline{\mathbb{S}^d \setminus K}$ of the complement of K in the d-sphere $\mathbb{S}^d$ and let ∂M denote the boundary of M with base-point $x \in \partial M$. We describe how our fundamental group procedure, together with standard procedures for low-index subgroups of finitely presented groups, can be used to infer information about the induced homomorphism of fundamental groups $\phi\colon \pi_1(\partial M, x) \to \pi_1(M, x)$ from a knowledge of the finite point cloud X sampled from K. Of particular interest is the case when K is a closed tubular neighbourhood of a tame knot in $\mathbb{R}^3$, the neighbourhood being small enough so as to contain the knot as a deformation retract. In this case the boundary ∂M is a torus and the homomorphism ϕ is known to determine M up to homeomorphism by a result of Waldhausen [32]. (More specifically, Waldhausen's result implies that two Haken 3-manifolds M, M' with connected boundary components are homeomorphic if and only if there exists an isomorphism $\pi_1 M \to \pi_1 M'$ inducing an isomorphism $\pi_1 \partial M \to \pi_1 \partial M'$.) It is known by a result of Gordon and Luecke [16] that two knots are ambient isotopic in $\mathbb{R}^3$ if and only if their complements are homeomorphic. The homomorphism ϕ is thus of some interest as it is a complete knot invariant. We illustrate the computation of ϕ for knots K arising from experimental data on protein backbones. We also use our computer implementation to observe that the first homology of subgroups $G < \pi_1(M_K)$ of index at most 6 suffices to distinguish between all prime knots K with eleven or fewer crossings.

1.1 Prior work

The combinatorially defined *edge-path group* of a connected simplicial complex K, due to Reidemeister, is well-known to be isomorphic to the fundamental group $\pi_1(K)$ (see [30]). It is also well-known that this combinatorial definition and isomorphism extends to connected regular CW-spaces. In this paper we use the terms *edge-path group* and *fundamental group* interchangeably as synonyms. Several authors have described algorithms for implementing Reidemeister's edge-path group. Rees and Soicher [29] use spanning trees and redundant relator searches in their description of an algorithm for finding a small finite presentation of the edge-path group of a 2-dimensional *combinatorial cell complex*; they implement their algorithm in GAP [13] for 2-dimensional simplicial clique complexes of graphs. Letscher [24] uses spanning trees and Tietze elimination/reduction of relators to compute edge-path groups from the 2-skeleta of simplicial complexes arising from knot complements, the knots being produced from experimental data on protein backbones. Palmieri *et al.* [28] have implemented the edge-path group of simplicial complexes in Sage [31]; the implementation uses the 2-skeleton of the complex and calls GAP's Tietze reduction/elimination procedures [13]. Kim *et al.* [22] describe an algorithm for the fundamental groups of 3-dimensional simplicial complexes; their algorithm, which makes use of 3-dimensional cells and the language of general CW-spaces, is applied to 3-dimensional tetrahedral meshes arising in computer vision.

We also mention that there is a large literature on the related problem of algorithmically determining a collection of shortest generating loops for the fundamental group of a space. The

case of oriented combinatorial 2-manifolds is treated by Ericson and Whittlesey in [10]; their algorithm invoves the computation of spanning trees.

1.2 Our contribution

We present an algorithm for computing a presentation of the edge-path group of a finite regular CW-space K based on the observation that a maximal acyclic discrete vector field on the 3-skeleton of K uniquely defines a finite presentation for $\pi_1(K)$ in the case when there is just a single critical 0-cell. The generators of the presentation correspond to critical 1-cells of K, and the relators of the presentation correspond to critical 2-cells. The 3-cells of K help to reduce the number of critical 2-cells. The algorithm thus boils down to a choice of procedure for finding a maximal acyclic discrete vector field on K; we consider one such choice in the present paper. In a subsequent paper [2] we plan to compare several procedures for obtaining discrete vector fields and provide formal time estimates as well as details on the performance of an optimized C++ implementation.

A discrete vector field can be viewed as a means of specifying a sequence of simple homotopy collapses in the sense of Whitehead's simple homotopy theory [35] (*cf.* Example 2). That theory seems to have been inspired by Tietze's theorem in the theory of group presentations asserting that any presentation of a group can be deformed into any other by a sequence of *Tietze moves*. The Tietze moves involve generators and relators and can be regared as being of a 1- and 2-dimensional nature. A simple homotopy collapse is the n-dimensional analogue of a Tietze move. In the context of the present paper, two advantages to simple homotopy collapses over Tietze moves are: (i) they yield a geometrically guided sequence of simplifications of a fundamental group presentation rather than just an algebraically guided sequence; (ii) they allow simplifications of the 3-skeleton and not just of the 2-skeleton.

Our algorithm for finding the homomorphism of finitely presented groups $\psi\colon \pi_1 L \to \pi_1 K$ induced by an inclusion of finite regular CW-space $L \subseteq K$ is based on the observation that an acyclic discrete vector field on K defines a sequence $\psi(e)$ of oriented critical 1-cells of K for each oriented 1-cell e in K. If e happens to be critical in K then $\psi(e)$ is just the sequence e of length 1. For non-critical 1-cells $\psi(e)$ is defined by a recursive formula. The value of the homomorphism ψ on a generator $x \in \pi_1(L)$ is read off directly from the discrete vector field on K by concatenating $\psi(f)$ for f ranging over the sequence of 1-cells in a representative path for $x \in \pi_1 L$.

In order to efficiently represent point clouds as regular CW-space we use a preliminary representation as either a pure simplicial complex, or a pure cubical complex or a pure permutahedral complex. The details of how point clouds are converted to these preliminary representations is explained in Section 2. Homotopy collapse procedures (such as the deformation retract procedures described in [20] and zig-zag deformation retract procedure described in [9]) are then used to reduce the cell structure of the pure complex while retaining its homotopy type. A reduced pure complex is converted to the data type of a regular CW-space.

In order to compute with boundaries of manifolds we use pure permutahedral complexes as the preliminary representation. These complexes are topological manifolds and in this setting the deformation retract procedures preserve homeomorphism type. On converting a pure permutahedral complex to a regular CW-space K we record that K is a manifold. A homeomorphism-type preserving cellular simplification procedure is applied to K before extracting its boundary ∂K as a pure regular CW-subcomplex of K of codimension 1. We give examples involving 3-dimensional manifolds arising from data on protein backbones that illustrate the potential of this approach.

As a more theoretical application we use our computer implementation to observe that the first integral homology of subgroups $G < \pi_1(M_K)$ of index at most 6 suffices to distinguish between all prime knots K with eleven or fewer crossings.

The paper is organized as follows. In Section 2 we recall details on regular CW-space, simplicial complexes, pure cubical complexes, pure permutahedral complexes and their computer representations. We also recall details on discrete vector fields and describe one easily implemented algorithm

for computing a maximal acyclic discrete vector field on finite regular CW-space. In Section 3 we describe an algorithm for computing presentations of fundamental groups, and an algorithm for computing homomorphisms between finitely presented fundamental groups. In Section 4 we describe a procedure for simplifying the cell structure of a regular CW-manifold without changing its homeomorphism type. We then demonstrate how to compute Waldhausen's complete knot invariant for knots arising from experimental data on protein backbones. In Section 5 we explain how low-index subgroups of finitely presented groups can be used to compute invariants of the isomorphism type of a finitely presented group. As an illustration we distinguish between all prime knots K with eleven or fewer crossings.

We have chosen to present theoretical aspects in an informal style. We plan to devote a subsequent more formal paper [2] to a rigorous analysis of our algorithm for finding presentations of fundamental groups and to a discussion of time-estimates and computational efficiencies.

2 Cellular spaces

2.1 Regular CW-spaces

A good introduction to the theory of CW-spaces can be found in [25]. A CW-space K is *regular* if every cell is attached by a map that restricts to a homeomorphism on the boundary of the cell. The complex K is *finite* if it has only finitely many cells. It is n-dimensional if it contains a cell of dimension n and no cell of higher dimension. An n-dimensional CW-space is *pure* if any cell of dimension $k < n$ lies in the boundary of some $(k+1)$-dimensional cell. The cellular structure of a finite regular CW-space K can be encoded as a finite collection of binary valued incidence numbers. From this encoding one can construct a homeomorphic regular CW-space K' for which there exists a homeomorphism $K' \cong K$ which maps cells homeomorphically to cells. We let K^k denote the k-skeleton of K, and e^k_j the jth cell of dimension k. Thus e^k_j is a subspace of K homeomorphic to an open Euclidean ball.

The space K can be represented on a computer as a sequence of lists $B^0, B^1, ..., B^n$, where the jth term of the list $B^k = \{b^k_1, b^k_2, ...\}$ records those $(k-1)$-dimensional cells of K that lie in the boundary of the jth k-dimensional cell of K. For algorithmic efficiency it is best to encode some additional redundant information. Namely, we record the lists $C^0, C^1, ..., C^n$, where the jth term of the list $C^k = \{c^k_1, c^k_2, ...\}$ records those $(k+1)$-dimensional cells of K that contain the jth k-dimensional cell of K in their boundary. Thus c^k_j records the *coboundary* of e^k_j. The lists B^k (or the lists C^k) can be used to construct the face poset of the space K.

2.2 Pure simplicial, cubical and permutahedral complexes

There are several ways to represent a pure simplicial complex on a computer. For present purposes an appropriate representation is simply as a regular CW-space. From this CW representation it is useful to derive and store some extra information for efficient computation of the deformation retracts described in Section 2.3 below. The extra information is a list $R = \{r_1, r_2, \ldots\}$ whose j-th term is a list $r_j = \{(d_1, k_1), (d_2, k_2), ...(d_t, k_t)\}$ of those integer pairs for which the k_i-th simplex of dimension d_i intersects non-trivially with the j-th simplex of dimension n. Each list r_j contains precisely $t = 2^{n+1} - 2$ pairs.

Let $L \subseteq \mathbb{R}^n$ be an additive subgroup generated by some choice of n linearly independent vectors $V_L = \{v_1, \cdots, v_n\}$. Then L acts freely on $\mathbb{R}^n$ as a discrete group of translations with fundamental domain

$$D_L = \{x \in \mathbb{R}^n \ : \ ||x|| \leq ||x - v|| \ \forall v \in L\} \ .$$

When V_L is an orthogonal set the convex polytope D_L is combinatorially equivalent to the n-dimensional hypercube. If we identify $\mathbb{R}^n$ with the hyperplane

$$\mathbb{R}^n = \{(x_1, \ldots, x_{n+1}) \in \mathbb{R}^{n+1} \ : \ x_1 + \cdots + x_{n+1} = 0\}$$

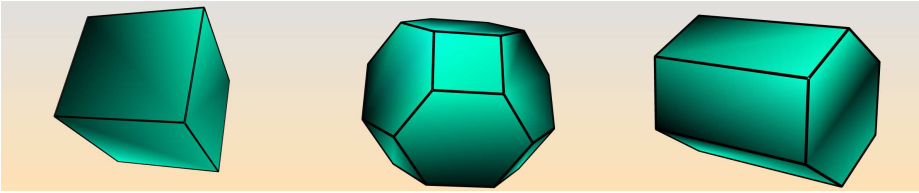

Fig. 1 3-dimensional cube, permutahedron and hexagonal prism.

and take V_L to be the set $v_1 = (-n, 1, 1, \ldots, 1, 1)$, $v_2 = (1, -n, 1, \ldots, 1, 1)$, $\ldots$, $v_n = (1, 1, 1, \ldots, -n, 1)$ of n vectors in this hyperplane, then the convex polytope D_L is the n-dimensional permutahedron. Other choices of V_L can lead to other polytopes such as prisms. A 3-dimensional cube, permutahedron and prism are illustrated in Figure 1.

Any finite subset $\Lambda \subset L$ determines a finite union of polytopes

$$P = \bigcup_{\lambda \in \Lambda} D_L + \lambda \ .$$

Following [9] we call such a union P a *lattice complex* or *L-complex*. When the polytope D_L is combinatorially equivalent to a hypercube we say that P is a *pure cubical complex*. When D_L is combinatorially equivalent to a permutahedron we say that P is a *pure permutahedral complex*. Figure 2 shows a 3-dimensional pure cubical complex and a 3-dimensional pure permutahedral complex, both of which model the trefoil knot.

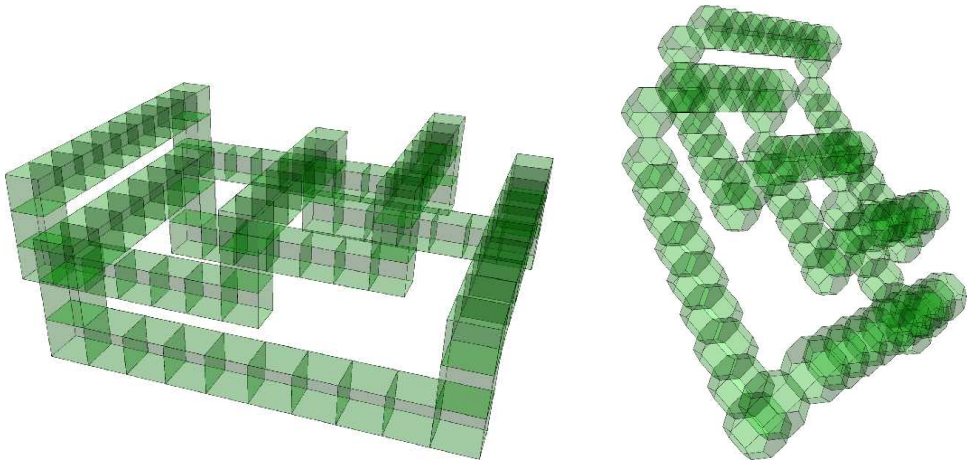

Fig. 2 Pure cubical and permutahedral models of the trefoil knot.

On a computer we represent the L-complex P as an n-dimensional binary array $A = (a_{\lambda_1,\lambda_2\cdots,\lambda_n})$ with

$$a_{\lambda_1,\lambda_2\cdots,\lambda_n} = \begin{cases} 1 \text{ if } D_L + \lambda_1 v_1 + \cdots + \lambda_n v_n \quad \text{lies in } P, \\ 0 \text{ otherwise.} \end{cases}$$

together with the finite set of integer vectors

$$B_L := \{\lambda = (\lambda_1, \ldots, \lambda_n) \in \mathbb{Z}^n \ : \ D_L \cap (D_L + \lambda_1 v_1 + \cdots + \lambda_n v_n) \neq \emptyset\}.$$

The binary array $a_{\lambda_1,\lambda_2\cdots,\lambda_n}$ is often called a *bitmap representation*. We call B_L the *L-ball*. For D_L an n-dimensional hypercube the L-ball contains $3^n - 1$ vectors. For D_L an n-dimensional permutahedron the L-ball contains $2^{n+1} - 2$ vectors.

2.3 Deformation retracts

The representations of n-dimensional pure regular CW-space K given in Section 2.2 have the property that, for low values of n, one can quickly compute the cellular structure of the intersection $\delta(e^n) = \overline{e^n} \cap \overline{K \setminus \overline{e^n}}$ of the closure of an n-dimensional cell $\overline{e^n}$ with the closure of the complement of $\overline{e^n}$ in K. If $\delta(e^n)$ is contractible then the pure complex $\overline{K \setminus \overline{e^n}}$ is a deformation retract of K. (To see this, note that the boundary $\partial e^n = \overline{e^n} \setminus e^n$ is homeomorphic to an $(n-1)$-sphere and that $\delta(e^n)$ is a closed contractible subspace of ∂e^n. Thus $\partial e^n \setminus \delta(e^n)$ is homotopic to a subspace $e^{n-1} \subset K$ that is homeomorphic to an open $(n-1)$-ball. Hence we can view $\overline{K \setminus \overline{e^n}} = K \setminus \{e^n \cup e^{n-1}\}$ as arising from K by a so-called simple homotopy collapse [35]; it is thus a deformation retract.) There are at most 2^{3^n-1} possible reduced CW-structures on $\delta(e^n)$ when K is a pure cubical complex, and at most $2^{2^{n+1}-2}$ possible structures when K is a pure simplicial or permutahedral complex. In these three cases, and for low values of n, it is possible to record the possible contractible $\delta(e^n)$ in a table and thus implement a quick test for when $\overline{K \setminus \overline{e^n}}$ is a deformation retract of K. Such a test is the basis of computational procedures described in [20] for finding small deformation retracts of pure cubical complexes; implementations for the cubical, simplicial and permutahedral settings are available in the software packages [4,8]. More recently an heuristic procedure for computing a sequence of deformation retracts

$$K \overset{\simeq}{\hookrightarrow} K_1 \overset{\simeq}{\hookleftarrow} K_2 \overset{\simeq}{\hookrightarrow} K_3 \cdots \overset{\simeq}{\hookleftarrow} K'$$

was described in [9]. The aim of the procedure is that K' is a pure complex homotopy equivalent to K but with fewer n-cells than K. We say that K' is a *zig-zag deformation retract* of K. An implementation of zig-zag deformation retracts for pure cubical and pure permutahedral complexes is available in [8].

For an arbitrary regular CW-space K containing a k-cell e^k that lies in the boundary of exactly one cell e^{k+1} of dimension $k+1$, we have a deformation retract $K \setminus (e^{k+1} \cup e^k) \hookrightarrow K$. This observation is the basis of procedures for computing small deformation retracts of arbitrary finite regular CW-space; these are implemented in [4,8]. The retract $K \setminus (e^{k+1} \cup e^k)$ is often said to be obtained from K by a *free face collapse*.

2.4 Čech complex and forgetful functors

Algebraic topology involves a range of functors for converting one type of cellular space into another type. We mention two such functors that are particularly useful for fundamental group computations.

The Čech complex construction can be used to associate a simplicial complex $Cech(P)$ to any n-dimensional lattice complex P. The simplicial complex $Cech(P)$ has one vertex for each n-cell in P, and one simplex for each collection of n-cells of P whose closures have non-trivial common intersection. This construction is easily implemented on a computer using the L-ball B_L to determine non-trivial intersections. For each n-cell e^n in P we can construct a small open neighourhood $e^n \subset \overline{e^n} \subset U(e)$ of the closure of e^n such that the open sets $U(e)$ form an open cover of P. Then $Cech(P)$ can be viewed as the nerve of the cover $\{U(e)\}$ and, as such, is well-known to be homotopy equivalent to P. Standard texts such as [30] contain details on the nerve.

A second useful and easily implemented class of functor are the so-called forgetful functors that convert a pure simplicial, cubical or permutahedral complex K to a regular CW-space simply by forgetting some of the structure of K. These forgetful functors are again easily implemented on a computer.

Example 1 Figure 3 (left) shows the backbone of the Thermus Thermophilus protein. The Euclidean coordinates of all atoms in the protein are available from the Protein Data Bank [27]. These coordinates are experimental data. The backbone is obtained by taking the coordinates of

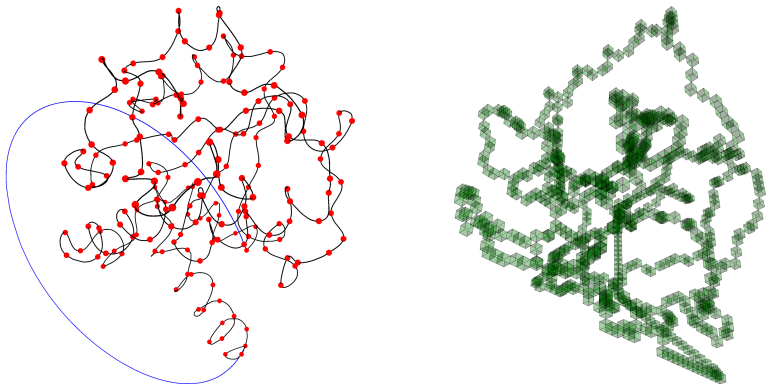

Fig. 3 T.thermophilus 1V2X protein backbone represented as a curve (left) and pure cubical complex (right)

the alpha carbon atom in each of the 191 amino acids in the protein and fitting a curve to this sequence of alpha carbon atoms. The initial and final alpha carbon atoms, which lie close to the 'surface' of the protein structure, have been artificially joined by a curve so as to form a continuous embedding $S^1 \to \mathbb{R}^3$. Any such embedding is referred to as a *knot*. Figure 3 (right) shows a representation of the knot as a 3-dimensional pure cubical complex K. The complex K was formed as a deformation retract of a pure cubical complex K', the complex K' having been constructed so that it contains the image of the knot $S^1 \to \mathbb{R}^3$ as a deformation retract. It can be seen from Figure 3 that the pure cubical complex K is not a topological manifold. However, K can be embedded in a manifold M as follows. Let $A = (a_\lambda)$ be the binary array representing K. Let $A' = (a'_\lambda)$ be a new binary array defined by setting $a'_\lambda = 1$ if $a_{\lambda+\mu} = 1$ for some $\mu \in B_L = \{-1, 0, 1\}^3$, and setting $a'_\lambda = 0$ otherwise. Let L be the pure cubical complex represented by A'. Then L is a topological manifold containing K. In this example it happens that L contains K as a deformation retract. (In other examples it might be necessary to rescale, and use smaller cubes to ensure that K is a deformation retract of L.) The pure cubical complexes K and L contain 1071 and 16408 3-cubes respectively.

As we are interested in the complement $\mathbb{R}^3 \setminus L$ we form a pure cubical complex C with binary array $A'' = (a''_\lambda)$ defined by $a''_\lambda = 1$ if $a'_\lambda = 0$ and $a''_\lambda = 0$ if $a'_\lambda = 1$. To ensure that C is a finite complex we restrict the index λ to a finite range that includes all cases where $a'_\lambda = 1$. The complex C has 691684 3-cubes and is homeomorphic to the complement $I^3 \setminus L$ with I^3 some 3-dimensional solid closed cube whose interior contains the manifold L. By applying the implementation of the zig-zag deformation procedure in [8] we obtain a pure cubical complex C' involving 4892 3-cubes that is homotopy equivalent to C. Considered as a regular CW-space C' has 77077 cells. A smaller, but homotopy equivalent, regular CW-space can be obtained as $C'' = Cech(C')$. The CW-space C'' has a total of 51607 cells and is 5-dimensional. In general the Cech complex of a pure cubical complex will be of higher dimension than the original complex. A deformation retract $C''' \subset C''$ can be computed using the default procedure in [8]. The regular CW-space C''' has a total of 30743 simplicial cells, compared to a total of 5674743 cubical cells in the homotopy equivalent CW-space C. Moreover, the CW-space C''' is of dimension 2.

2.5 Discrete vector fields

Many algorithms on CW-spaces work more efficiently when a space has fewer cells. The requirement that a CW-space be regular often necessitates the inclusion of more cells than would be needed in a non-regular CW-decomposition of the space. The difficulty with non-regular CW-spaces is, however, that it is not so clear how best to represent them combinatorially on a computer without loosing homotopy-theoretic information. We opt for a compromise representation involving a triple (X, Y, h) where X is a regular CW-space, Y is a (possibly) non-regular CW-space and $h\colon X \to Y$

is a homotopy equivalence. This way all information is stored in X, while Y can be used in algorithms. We specify h and Y using the following notion.

Definition 1 A *discrete vector field* on a regular CW-space X is a collection of formal arrows $s \to t$ where

1. s, t are cells of X with $\dim(t) = \dim(s) + 1$ and with s lying in the boundary of t. We say that s and t are *involved* in the arrow, that s is the *source* of the arrow, and that t is the *target* of the arrow.
2. any cell is involved in at most one arrow.

An example of a discrete vector field on a regular CW-decomposition of a torus is illustrated in Figure 4. The term *discrete vector field* is due to Forman [11]. In an earlier work [19] Jones calls

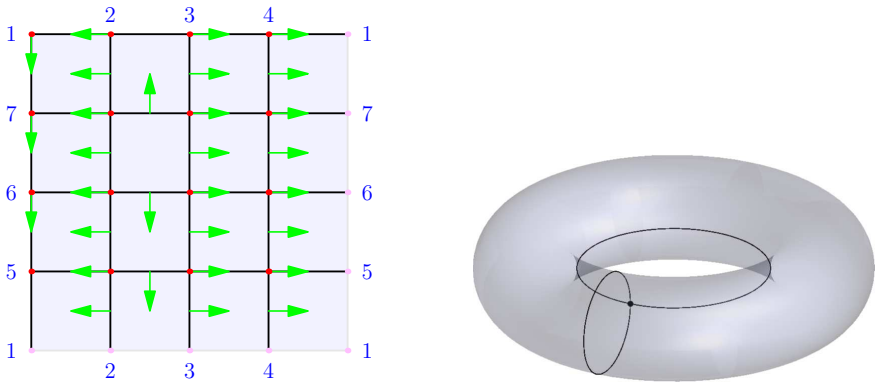

Fig. 4 A regular CW-structure on a torus endowed with an acyclic discrete vector field (left) and a non-regular CW-structure (right).

this concept a *marking*.

A *chain* in a discrete vector field is a sequence of arrows

$$\ldots, s_1 \to t_1, s_2 \to t_2, s_3 \to t_3, \ldots$$

where s_{i+1} lies in the boundary of t_i for each i. A chain is said to be a *circuit* if it is of finite length with source s_1 of the initial arrow $s_1 \to t_1$ lying in the boundary of the target t_n of the final arrow $s_n \to t_n$. A discrete vector field is said to be *admissible* if it contains no circuits and no chains that extend infinitely to the right. We are only concerned with finite CW-spaces; in this context a discrete vector field is admissible if it contains no circuits. In the context of finite CW-spaces we use the term *acyclic* as a synonym for *admissible*. We say that an admissible discrete vector field is *maximal* if it is not possible to add an arrow while retaining admissibility. A cell in X is said to be *critical* if it is not involved in any arrow.

Theorem 1 *[11, 12, 35] If X is a regular CW-space with admissible discrete vector field then there is a homotopy equivalence*

$$X \simeq Y$$

where Y is a CW-space whose cells are in one-one correspondence with the critical cells of Y.

The CW-space Y in this theorem is determined (up to some choice in its cell attaching maps) by X and the discrete vector field on X. In Forman's paper [11] the cellular chain complex $C_*(Y)$ is referred to as the *Morse complex* of X. We shall refer to the space Y as the *Morse CW-complex* of X. The following example is presented in such a way that it serves as an informal constructive proof of Theorem 1. This constructive proof underlies our algorithm for computing a small presentation of $\pi_1(X)$. We plan to present a more rigorous account of this proof in [2].

Example 2 Consider the acyclic discrete vector field on the regular CW-decomposition of the torus X shown in Figure 4 (left). It has one critical 0-cell, two critical 1-cells and one critical 2-cell. We shall derive a homotopy equivalence $h\colon X \to Y$ where Y is a CW-space with one 0-cell, two 1-cells and one 2-cell as illustrated in Figure 4 (right). To this end we let X^n denote the n-skeleton of X; we let T^n denote the union of X^n with all those $(n+1)$-cells that occur as the target of some arrow on X; we denote by V^{n+1} the union of T^n with all the critical $(n+1)$-cells of X. We take V^0 to be the collection of critical 0-cells. There are inclusions

$$V^n \subseteq X^n \subseteq T^n \subseteq V^{n+1}.$$

The spaces T^0, V^1 and T^1 are illustrated in Figure 5.

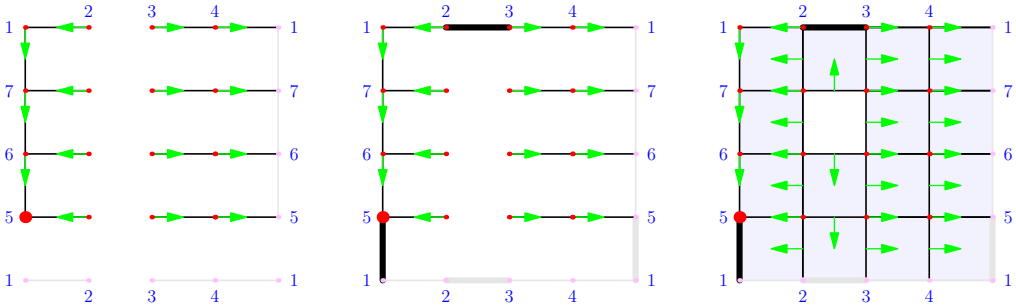

Fig. 5 The spaces T^0, V^1 and T^1 for Example 2 with critical cells shown in bold.

We construct the CW-space Y recursively, defining its 0-skeleton Y^0 to consist of the critical 0-cells. So $V^0 = Y^0$. Suppose now that we have constructed a homotopy equivalence $\nu\colon V^n \to Y^n$ where the cells of Y^n are in one-one correspondence with the critical k-cells of X for $k \leq n$.

In the spirit of our discussion of deformation retracts in Section 2.3 we can establish that V^n is a deformation retract of T^n. To do this we note that since the discrete vector field is admissible there must be some n-cell e^n in T^n that is the source of the initial arrow $e^n \to e^{n+1}$ of some chain in the vector field on T^n. Let $T^n_1 \subset T^n$ denote the regular CW-space obtained from T^n by removing e^n and e^{n+1}. This constitutes an *elementary collapse* $T^n \searrow T^n_1$ and *elementary expansion* $T^n_1 \nearrow T^n$ in the sense of simple homotopy theory [6]; in particular, T^n_1 is a deformation retract of T^n. Applying the same procedure to T^n_1 we obtain a deformation retract $T^n_2 \hookrightarrow T^n_1$ with $V^n \subseteq T^n_2$. Repeating the procedure we obtain a sequence of deformation retracts

$$V^n \nearrow T^n_k \nearrow T^n_{k-1} \nearrow \cdots \nearrow T^n_1 \nearrow T^n.$$

Therefore there exists a map $\tau\colon T^n \to V^n$ which is a homotopy equivalence.

For each critical $(n+1)$-cell e^{n+1} in X we have a characteristic map $D^{n+1} \to X$ which resiticts to $\phi\colon S^n \to X^n$. The composite map

$$\phi'\colon S^n \xrightarrow{\phi} X^n \subseteq T^n \xrightarrow{\tau} V^n \xrightarrow{\nu} Y^n$$

can be used to form the union $Y^n \cup e^{n+1}$ with CW-structure in which ϕ' extends to a characteristic map for e^{n+1}. We take Y^{n+1} to be the space obtained by attaching to Y^n all the critical $(n+1)$-cells of X in this manner. The construction of Y^{n+1} is such that the homotopy equivalence $\nu\colon V^n \to Y^n$ extends to a homotopy equivalence $\nu\colon V^{n+1} \to Y^{n+1}$.

Let V, Y denote the CW-spaces with skeleta V^n, Y^n. For $n = \dim(X)$ we have the desired homotopy equivalence $h\colon X = V^n \to Y$.

Our computer representation of a discrete vector field on a finite n-dimensional regular CW-space K is a sequence of lists $V^0, V^1, \ldots, V^{n-1}$, where the jth term of the list $V^k = \{v^k_1, v^k_2, \ldots\}$ is either unbound (*i.e.* empty) or else records an arrow wth source the j-th cell of dimension k and

target cell number v_j^k of dimension $k+1$. For algorithmic efficiency we also record the sequence of lists $U^1, U^2, ..., U^n$, where the jth term of list $U^k = \{u_1^k, u_2^k, ...\}$ is either unbound or records an arrow with target the j-th cell of dimension k and source equal to cell number u_j^k of dimension $k-1$.

There are a number of approaches to constructing discrete vector fields, some of which are based on the following obvious result.

Lemma 1 *Let X be a regular CW-space endowed with an admissible discrete vector field. Suppose that there exist a pair of critical cells s, t in X such that:* $\dim(t) = \dim(s)+1$*; s lies in the boundary of t; any other cell of dimension* $\dim(s)+1$ *containing s in its boundary is critical. Then the vector field on X can be extended by adding the arrow $s \to t$ and the resulting discrete vector field is admissible.*

One procedure for constructing a discrete vector field using Lemma 1 is given in Algorithm 2.1 There are various possibilities for Step 1 of the algorithm. The cells could be partially ordered in

Algorithm 2.1 Discrete vector field on regular CW-space

Require: A finite regular CW-space X
Ensure: A maximal admissible discrete vector field on X.
1: **procedure**
2: Partially order the cells of X in any fashion.
3: Initially deem all cells of X to be critical.
4: Furthermore, deem all critical cells to be *inessentially critical* and none to be *essentially critical*.
5: **while** there exists an inessentially critical cell **do**
6: **while** there exists a pair of inessentially critical cells s, t such that: $\dim(t) = \dim(s) + 1$; s lies in the boundary of t; no other inessentially critical cell of dimension $\dim(s) + 1$ contains s in its boundary; **do**
7: Choose such a pair (s, t) with s minimal in the given partial ordering.
8: Add the arrow $s \to t$ and deem s and t to be non critical.
9: **end while**
10: **if** there exists an inessentially critical cell **then**
11: Choose a minimal inessentially critical cell and deem it to be essentially critical.
12: **end if**
13: **end while**
14: **end procedure**

some way that ensures any cell of dimension k is less than all cells of dimension $k+1$. This partial ordering guarantees that the resulting discrete vector field on a path-connected regular CW-space K will have a unique critical 0-cell. A natural alternative that seems to give good results is to partially order the cells of K in some fashion such that any cell of dimension $k+1$ is less than all cells of dimension k.

Example 3 The implementation of Algorithm 2.1 available in [8] was applied to the 5-dimensional regular CW-space C'' of Example 1. With $(k+1)$-cells ordered less than k-cells the implimentation produces a discrete vector field with one critical 0-cell, two critical 1-cells, two critical 2-cells and no critical cells in higher dimensions. This output corresponds to a representation of the homotopy type of the knot complement as a CW-space with just five cells. With k-cells ordered less than $(k+1)$-cells the implementation produces one critical 0-cell, five critical 1-cells, five critical 2-cells and no critical cells in higher dimensions. When the algorithm, with same ordering, is applied directly to the much larger homotopy equivalent 3-dimensional CW-space C' it returns one critical 0-cell, three critical 1-cells, three critical 2-cells and no critical cells in higher dimensions.

There are other possible algorithms for constructing discrete vector fields. For example, two algorithms which are similar but slightly different to the one presented above are described in [17, 18]. We plan to discuss and compare a range of algorithms in a subsequent article [2].

3 Computing fundamental groups of cellular spaces

A CW-space is said to be *reduced* if it has just one 0-cell. It is well-known that the cellular structure of the 2-skeleton of a reduced CW-space Y is encoded by a corresponding presentation for its fundamental group in terms of a generating set $\underline{x}$ for a free group $F = F(\underline{x})$ and a set $\underline{r} \subset F$ of relators. The generators are in bijection with the 1-cells of Y, and the relators are in bijection with the 2-cells. A precise statement of this assertion can be given using the language of free crossed modules and is due to Whitehead [33,34]. It is well-known that if Y is any path-connected CW-space then, by contracting any maximal tree in its 1-skeleton, the CW-structure on the space Y can be modified to that of a reduced one.

The above is very standard material covered, for instance, in [14]. We now summarize details of how to construct a presentation $\langle \underline{x} \mid \underline{r} \rangle$ for the fundamental group $\pi_1(Y)$ of the CW-space Y of Theorem 1 that corresponds to a regular CW-space K endowed with admissible discrete vector field.

We say that a cell in K is *terminal* if it is either critical or the target of an arrow. Otherwise, we say that the cell is *initial*. Note that a 0-cell is terminal if and only if it is critical.

Any 0-cell e^0 in K can be associated with a unique terminal 0-cell $H(e^0)$ by recursively defining

$$H(e^0) = \begin{cases} e^0 & \text{if } e^0 \text{ is terminal,} \\ H(e'^0) & \text{if there exists an arrow } e^0 \to e^1 \text{ with } e'^0 \neq e^0 \text{ a boundary of the } 1-\text{cell } e^1 \ . \end{cases}$$

By an *orientation* on a 1-cell e^1 in K we mean that its two distinct vertices have been labelled as $\partial^- e^1_i$ and $\partial^+ e^1_i$. We say that $\partial^- e^1_i$ is the *first* vertex of e^1_i and that $\partial^+ e^1_i$ is the *second* vertex. Any oriented 1-cell e^1 in K can be associated with a unique path $H(e^1)$ of terminal 1-cells $(e^1_1, e^1_2, \ldots, e^1_n)$ in K. By a *path* we mean that each edge is endowed with an orientation such that $\partial^+ e^1_i = \partial^- e^1_{i+1}$ for $1 \leq i \leq n-1$ and $\partial^- e^1_i = \partial^+ e^1_{i-1}$ for $2 \leq i \leq n$. Once we have defined $H(e^1)$ we will be able to define, for any path of 1-cells $f = (f^1_1, f^1_2, \ldots, f^1_m)$, the path $H(f)$ to be the concatenation of the ordered sequence of paths $H(f^1_1), H(f^1_2), \ldots, H(f^1_m)$.

For any oriented initial 1-cell e^1 we have an associated 2-cell $e^1 \to e^2$. The boundary of e^2 specifies a path $f^1_1, f^1_2, \ldots, f^1_m$ where the first vertex of f^1_1 is the first vertex of e^1, and the second vertex of f^1_m is the second vertex of e^1. We denote this path by $\leftarrow e^1 \rightarrow$. The path involves all 1-cells of the boundary of e^2 except the 1-cell e^1.

We define $H(e^1)$ recursively by

$$H(e^1) = \begin{cases} e^1 & \text{if } e^1 \text{ is terminal,} \\ H(\leftarrow e^1 \rightarrow) & \text{otherwise.} \end{cases}$$

Note that in the recursive definitions of $H(e^0)$ and $H(e^1)$ the recusrion will terminate because of the admissibility condition on the discrete vector field.

We can now describe the CW-space Y of Theorem 1. The 0-skeleton Y^0 consists of one 0-cell for each critical 0-cell e^0 in K; we denote by ρe^0 the 0-cell of Y corresponding to e^0. The 1-skeleton Y^1 consists of Y^0 together with one 1-cell ρe^1 for each critical 1-cell e^1 of K. The 1-cell ρe^1 is attached to the 0-cells $\rho H(e^0)$, $\rho H(e'^0)$ where e^0 and e'^0 are the 0-cells in the boundary of e^1. The 2-skeleton Y^2 consists of Y^1 together with one 2-cell ρe^2 for each critical 2-cell e^2 in K. The 2-cell ρe^2 is attached via the path $(\rho(e^1_1), \rho(e^1_2), \ldots, \rho(e^1_n))$ where the sequence $(e^1_1, e^1_2, \ldots, e^1_n)$ is obtained by applying H to the entire boundary ∂e^2 of e^2 and deleting any non-critical 1-cells from the resulting path $H(\partial e^2)$. The boundary ∂e^2 is a closed path of 1-cells and it does not matter at which 1-cell the closed path is deemed to start.

Example 4 In the discrete vector field of Figure 4 the CW-space Y has one 0-cell, two 1-cells and one 2-cell. If we denote the arbitrarily oriented 1-cells of Y by x and y, and denote their opposite orientations by x^{-1} and y^{-1}, then the 2-cell is attached via the path $xyx^{-1}y^{-1}$. Using the well-known correspondence between reduced 2-dimensional CW-spaces and group presentations we obtain the presentation $\pi_1(Y) \cong \langle x, y \mid xyx^{-1}y^{-1} \rangle$.

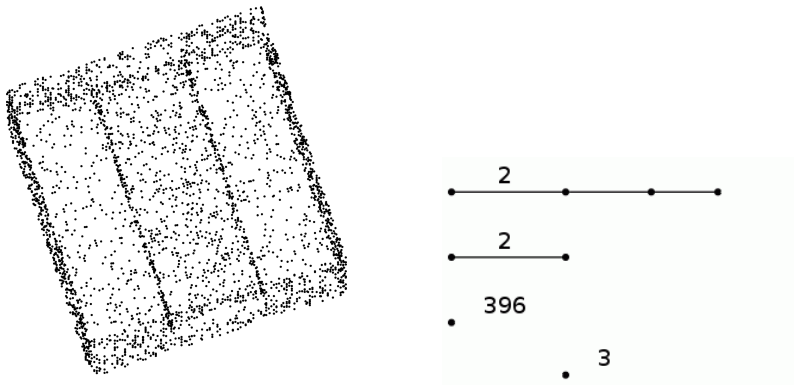

Fig. 6 Sample of 3397 points from a surface in $\mathbb{R}^3$ and its β_1 barcode.

The CW-space Y of Theorem 1 will often not be reduced. However, if it is connected then we can construct a maximal tree in its 1-skeleton and contract the tree to a point in order to obtain a reduced CW-structure on Y. We read a presentation for the fundamental group of a reduced CW-space Y directly from its 2-skeleton.

Example 5 The above construction of a group presentation from an admissible discrete vector field on a regular CW-space has been implemented in [8]. Applying this implementation to the first vector field of Example 3 yields the presentation $\langle x, y \mid y^{-1}x^{-1}y^{-1}xyx, y^{-1}x^{-1}yxyx^{-1}\rangle$ for the fundamental group of the complement of the knot in Figure 3. On applying GAP's Tietze operation procedure this presentation simplifies to $\langle x, y \mid y^{-1}x^{-1}y^{-1}xyx\rangle$.

Example 6 Figure 6 (left) shows a set X of 3397 points sampled from an unkown surface $K \subset \mathbb{R}^3$. The surface K lies inside the cube of side 100 and the points in X have been chosen to have integer coordinates. We let K_0^X denote the pure cubical complex consisting of one unit 3-cube centred at the integer vector x for each vector $x \in X$. We define K_i^X recursively to consist of the union of those unit 3-cubes centred at integer vectors that intersect non-trivially with K_{i-1}^X. As a means of inferring infomation about K we can follow the standard persistent homology approach to data analysis and investigate the induced homology maps $\alpha_n^{ij}\colon H_n(K_i^X, \mathbb{Q}) \to H_n(K_j^X, \mathbb{Q})$ for $i \leq j$. These are maps of vector spaces and thus determined by their rank $\beta_n^{i,j} = \mathrm{rank}(\alpha_n^{ij})$. The degree 1 homology maps

$$H_1(K_2^X, \mathbb{Q}) \to H_1(K_6^X, \mathbb{Q}) \to H_1(K_{10}^X, \mathbb{Q}) \to H_1(K_{14}^X, \mathbb{Q})$$

are described by the barcode of Figure 6 (right) which was produced using the persistent homology implementation in [8]. The number of vertices in each column of the barcode, counted by multiplicity, equals $\beta_1^{i,i} = \beta_1(K_i^X) = \mathrm{rank}(\mathrm{H}_1(\mathrm{K}_\mathrm{i}^\mathrm{X}, \mathbb{Q}))$. So $\beta_1^{2,2} = 400$, $\beta_1^{6,6} = 7$, $\beta_1^{10,10} = 2$, $\beta_1^{14,14} = 2$. The number of lines from the i-th column to the j-th column equals $\beta_1^{i,j}$. Thus $\beta_1^{2,6} = 4$, $\beta_1^{2,10} = 2$, $\beta_1^{2,14} = 2$ etc. The long line corresponding to $\beta_1^{2,14} = 2$ implies that two 1-dimensional homology classes in K_2^X persist as distinct homology classes in K_{14}^X.

The barcode is consistent with X having been sampled from a manifold K with $\beta_1(K) = 2$. The corresponding β_0 and β_2 barcodes (which are not shown) suggest that $\beta_0(K) = 1$, $\beta_2(K) = 1$. These Betti numbers are consistent with K being a torus $K = S^1 \times S^1$ or a wedge of one 2-sphere with two 1-spheres $K = S^2 \vee S^1 \vee S^1$.

Up to this point in the example we have applied standard techniques from persistent homology to the topological data analysis of our point cloud (*c.f.* [7]). At this point it is useful to take the novel step of computing the fundamental group of K_{14}^X. This pure cubical complex has 1321192 cubes of dimension 3. A ziz-zag deformation retract $L \simeq K_{14}^X$ can be constructed with just 664 3-cubes. The space L and fundamental group presentation $\pi_1(L) \cong \langle x, y \mid xyx^{-1}y^{-1}\rangle$ were computed using [8]. The computation involved the construction of a discrete vector field on L with one critical

0-cell, two critical 1-cells, one critical 2-cell and no other critical cells. The presentation suggests that X was sampled from a torus.

The method of Example 6 could be directly applied to the topological analysis of 3-dimensional digital images and thus provides an additional tool in this area (*cf.* [15]). For higher-dimensional data a variant of Example 6 would be to choose an increasing sequence of real numbers $\epsilon_1, \epsilon_2, \ldots$ and take K_i^X to be the simplicial *clique complex* (also called the *Vietoris-Rips* complex) having one vertex for each point in X, and one k-simplex for each subset of vertices $\{x_1, \ldots, x_{k+1}\} \subset X$ where the distance between each pair of points in the subset satisfies $d(x_i, x_j) \leq \epsilon_k$. The distance d could be Euclidean distance or any other distance. This variant would use only a knowledge of the distances between each pair of points in X.

A notable feature of Examples 5 and 6 is the relatively uncomplicated nature of the presentations for the fundamental groups which were produced using the implementation in [8]. Above we have stated that our fundamental group algorithm returns a presentation which is *typically small*. This is an informal statement with the terms *typically* and *small* intentionally left undefined. One standard measure of the size of a presentation $\mathcal{P} = \langle \underline{r} \mid \underline{x} \rangle$ of a group G involves its *deficiency* $\mathrm{def}(\mathcal{P}) = |\underline{r}| - |\underline{x}|$, defined as the difference between the number of relators and the number of generators. One has the inequalities

$$\mathrm{def}(\mathcal{P}) \geq d(H_2(G, \mathbb{Z})) - \mathrm{rank}(H_1(G, \mathbb{Z})) \tag{1}$$

$$|\underline{x}| \geq \mathrm{rank}(H_1(G, \mathbb{Z})) \tag{2}$$

where $H_n(G, \mathbb{Z})$ is the integral homology of G, *rank* is the torsion-free rank, and $d()$ denotes the minimum number of generators of a group (see for instance [3]). Inequalities (1) and (2) are equalities for the presentations computed in Examples 5 and 6 and thus the number of generators and relations in these examples is as small as possible. Of course, in other examples the output presentation will not be minimal in this sense and it seems a difficult task to quantify the term *typically* in our informal statemnent. We should also mention that there exist finite groups G due to Swan and torsion free groups G due to Lustig (see [3] for references) for which (1) is a strict inequality for all presentations of G.

In keeping with the spirit of persistent homology, we consider the homomorphisms $\pi_1^{ij} \colon \pi_1(K_i^X) \to \pi_1(K_j^X)$ induced by inclusion. In Example 6 we readily compute that $\pi_1(K_6^X)$ is the free group on seven generators, and that $\pi_1(K_{10}^X)$ is the free abelian group on two generators. In this example the theoretical isomorphism $\pi_1(K_i^X)_{ab} \cong H_1(K_i^X, \mathbb{Z})$ of first integral homology as the abelianization of the fundamental group, together with the given computations implies that $\pi_1^{6,14}$ is a surjection and that $\pi_1^{10,14}$ is an isomorphism.

In general we could attempt to compute $\pi_1^{ij} \colon \pi_1(K_i^X) \to \pi_1(K_j^X)$ by computing a maximal discrete vector field on K_i^X and a possibly unrelated maximal discrete vector field on K_j^X. To simplify the discussion let us suppose that K_i^X, K_j^X are path-connected and that we have chosen to construct a vector field on K_i^X with a unique critical 0-cell in K_i^X. This vector field determines a tree $T_i \subset K_i^X$ in which all 1-cells are targets of arrows. Each generator $x \in \pi_1(K_i^X)$ corresponds to some critical 1-cell e_x^1 in K_i^X. This critical 1-cell e_x^1 determines a loop γ_x of 1-cells representing x with all but one of the 1-cells of γ_x lying in T_i. Each 1-cell in the loop γ_x is also a 1-cell in K_j^X. For each 1-cell f in the loop γ_x let $H(f)$ denote the sequence of 1-cells in K_j^X defined above with respect to the disrete vector field on K_j^X. Then $\pi_1^{ij}(x)$ is represented by the concatenation of the sequences $H(f)$ as f runs through the edges of the path γ.

This computation of homomorphisms of finitely presented groups applies to the homomorphism $\pi_1(K) \to \pi_1(L)$ induced by any inclusion $K \subseteq L$ of regular CW-spaces. We give a computed example in Section 4.

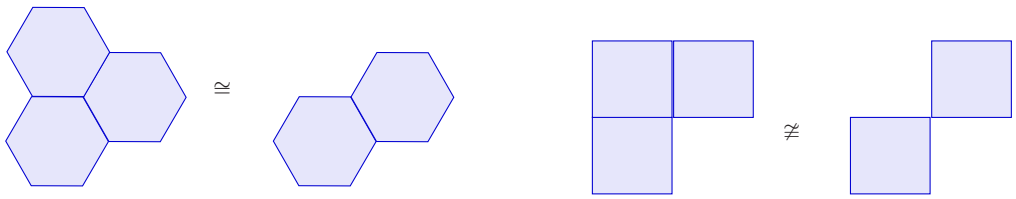

Fig. 7 Permutahedral retracts are homeomorphic whereas cubical retracts may not be.

4 Cellular manifolds and a complete knot invariant

An n-dimensional pure permutahedral complex K has the property that any two distinct n-cells e^n, e'^n with non-trivially intersecting boundaries share exactly one common boundary cell of dimension $n-1$. This implies that K is a topological manifold (i.e. locally homeomorphic to $\mathbb{R}^n$). Furthermore it implies that for any "elementary" deformation retract $\overline{K \setminus \overline{e^n}} \subset K$ there exists a homeomorphism $\overline{K \setminus \overline{e^n}} \cong K$ (cf. Figure 7). Since any zig-zag deformation retract is just a sequence of such elementary deformations it follows that any zig-zag deformation retract of a pure permutahedral complex K is actually homeomorphic to K.

A second important feature of elementary deformation retracts in the pure permutahedral and pure cubical settings is that they preserve ambient isotopy type. That is, if K' is a zig-zag deformation retract of K then there exists a continous map $F\colon \mathbb{R}^n \times [0,1] \to \mathbb{R}^n$ such that F_0 is the identity, F_t is a homeomorphism, and F_1 maps K to K'.

These properties make 3-dimensional pure permutahedral complexes a good setting for computational knot theory.

Example 7 The implementation of permutahedral complexes in [8] was used to construct a 3-dimensional pure permutahedral complex K ambient isotopic to a small closed neighbourhood of the knot in Figure 3 (left). The complex K contained 1606 3-dimensional permutahedra. A contractible compact manifold $R \subset \mathbb{R}^3$ was constructed so that K lies in the interior of R. The homeomorphism type of the complement closure $\overline{R \setminus K}$ was realized as a 3-dimensional pure permutahedral complex C involving 4793 3-dimensional permutahedral cells. The manifold C was then represented as a regular CW-space. The manifold boundary $T = \partial C$ was computed as a 2-dimensional pure CW-space. The total number of cells in C and T was 207553 and 171374 respectively. As theory predicts, the complex T had two connected components, one homeomorphic to a torus and the other homeomorphic to a 2-sphere corresponding to the boundary of R. Choosing a common base-point for T and C in the torus component of T, we then used the the fundamental group procedure described above to compute the induced homomorphism $\pi_1(T) \to \pi_1(C)$ as the homomorphism of finitely presented groups

$$\langle u, v \mid uvu^{-1}v^{-1} \rangle \longrightarrow \langle x, y \mid xyxy^{-1}x^{-1}y^{-1} \rangle, u \mapsto x^{-3}yx^2yx, v \mapsto x \ .$$

By results of Waldhausen [32] and Gordon and Luecke [16] this homomorphism is known to completely determine the ambient isotopy of the knot K.

The following GAP code was used to construct $\pi_1(T) \to \pi_1(C)$.

```
gap> K:=ReadPDBfileAsPurePermutahedralComplex("1V2X.pdb");
Pure permutahedral complex of dimension 3.

gap> C:=ComplementOfPureComplex(K);
Pure permutahedral complex of dimension 3.

gap> C:=ZigZagContractedPureComplex(C);
Pure permutahedral complex of dimension 3.

gap> Y:=PermutahedralComplexToRegularCWComplex(C);;
Regular CW-space of dimension 3

gap> i:=BoundaryPairOfPureRegularCWComplex(Y);
Map of regular CW-spaces
```

```
gap> CriticalCellsOfRegularCWComplex(Source(i));
[ [ 2, 1 ], [ 2, 1331 ], [ 1, 9951 ], [ 1, 31415 ],
  [ 0, 22495 ], [ 0, 25646 ] ]

gap> phi:=FundamentalGroup(i,22495);
[ f1, f2 ] -> [ f1^-3*f2*f1^2*f2*f1, f1 ]

gap> RelatorsOfFpGroup(Source(phi));
[ f1*f2^-1*f1^-1*f2 ]

gap> RelatorsOfFpGroup(Target(phi));
[ f1^-1*f2^-1*f1*f2*f1*f2^-1 ]
```

Let K be an arbitrary manifold endowed with the cell structure of a regular CW-space. We mentioned above that any k-cell e^k lying in the boundary of exactly one cell e^{k+1} of dimension $k+1$ yields a deformation retract $K \setminus (e^{k+1} \cup e^k) \hookrightarrow K$. In general this deformation retract will not be a manifold. There is an alternative basic operation that can sometimes be applied to the regular CW-structure of the manifold K to yield a homeomorphic regular CW-manifold with two fewer cells. Suppose that K contains a k-cell e^k lying in the boundary of precisely two $(k+1)$-cells e_1^{k+1}, e_2^{k+1} with identical coboundaries (i.e. e_1^{k+1} lies in the boundary of a cell e^{k+2} if and only if e_2^{k+1} lies in the boundary of e^{k+2}.) Then the three cells $e^k, e_1^{k+1}, e_2^{k+1}$ can be removed and replaced by a single cell of dimension $k+1$. The topological space K is unchanged; only its CW-structure changes. The resulting CW-structure will not in general be regular. However, it will be regular if the sets of vertices V_0, V_1, V_2 lying in the boundaries of $e^k, e_1^{k+1}, e_2^{k+1}$ respectively are such that $V_0 = V_1 \cap V_2$. Several applications of this operation are illustrated in Figure 8. A simplification procedure for regular CW-manifolds, based on this operation, is implemented in [8]. (A related simplification procedure for arbitrary finitely generated chain complexes is described in [21].)

5 Computations with low-index subgroups

Novikov [26] and Boone [1] have shown that there exists a finitely presented group for which it is impossible to find an algorithm for deciding equality between group elements given by products of powers of the generators. Nevertheless, there are algorithms for determining some basic isomorphism invariants of finitely presented groups. Moreover, the GAP system contains implementations

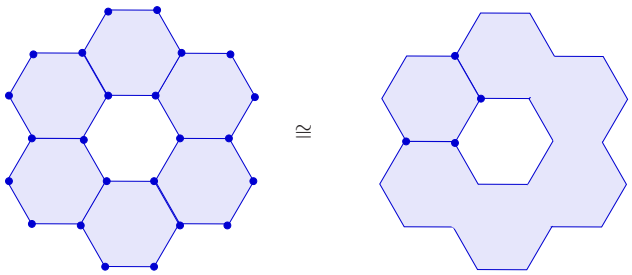

Fig. 8 Simplification of cell structure of regular CW-spaces

of these which seem to be quite practical for many groups G involving just a few generators and a few short relators. The algorithms we have in mind compute the following invariants of G. We refer the reader to GAP's extensive manual pages [13] for further information on these algorithms and implementations.

- The invariants of the abelian quotient $G/[G,G]$.
- For a given integer n, a list of finite presentations for the subgroups $S < G$ of finite index at most n.
- For a given integer c, a finite presentation and normal form procedure for the nilpotent group $G/\gamma_{c+1}G$ where terms of the lower central series are defined as $\gamma_1 G = G$, $\gamma_{i+1} = [\gamma_i G, G]$. The normal form procedure provides an algorithm for determining equality between group elements.
- For given integers c and p, a finite presentation and normal form procedure for the finite p-group $G/\gamma^p_{c+1}G$ where terms of the lower p-central series are defined as $\gamma^p_1 G = G$, $\gamma^p_{i+1} = [\gamma^p_i G, G]\,(\gamma^p_i G)^p$.
- For given integers c, p and m, the invariants of the integral homology groups $H_m(G/\gamma_{c+1}G, \mathbb{Z})$ and $H_m(G/\gamma^p_{c+1}G, \mathbb{Z})$.

In view of the availability of the above algorithms we consider the following invariants of a group G and group homomorphism $\phi\colon G \to G'$.

Definition 2 For a group G and for integers $m, n, c \geq 1$ we define the set of homology groups

$$I^{[n,c,m]}(G) = \{H_m(S/\gamma_{c+1}S, \mathbb{Z}) \ : \ S \leq G, |G:S| \leq n\}.$$

For a homomorphism $\phi\colon G \to G'$, prime integer $p > 1$ and integers $m, c \geq 1$ we define the canonical abelian group homomorphism

$$J^{[p,c,m]}(\phi) : \ H_m(S/\gamma_c S, \mathbb{Z}) \to H_m(S'/\gamma_{c+1}S', \mathbb{Z})$$

for $S = [G,G]\,G^p$, $S' = [G',G']\,G'^p$.

Both invariants in Definition 2 can, in principle, be computed in GAP for any finitely presented group and any homomorphism of finitely presented groups. In practice, the computation is not practical for groups with large generating sets or large sets of relators, or small sets of relators where the word length of individual relators is large. It is difficult to quantify 'large' as it depends to some extent on the nature of the presentations involved.

Example 8 The following GAP code computes the invariant $I^{[5,2,3]}(G)$ for G the complement of the protein knot shown in Figure 3. The set $I^{[5,2,3]}(G)$ is printed as a sorted list of abelian group invariant lists. From the printout we see that G has nine subgroups $S < G$ of index $|G:S| \leq 5$. Timings, shown in milliseconds, were obtained on a Lenovo ThinkPad W530 Linux laptop.

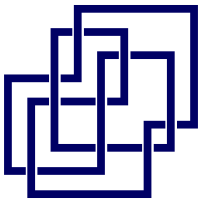

Fig. 9 A 3-dimensional pure cubical complex representation of knot 49 on nine crossings.

```
gap> K:=ReadPDBfileAsPurePermutahedralComplex("1V2X.pdb");;time;
7901
gap> G:=KnotGroup(K);;time;
25722
gap> L:=LowIndexSubgroupsFpGroup(G,5);;time;
0
gap> Apply(L,F->Range(IsomorphismFpGroup(F)));;time;
8
gap> Apply(L,S->NilpotentQuotient(S,2));;time;
4
gap> Apply(L,Q->GroupHomology(Q,3));;time;
952
gap> Print(SortedList(L));
[ [  ], [  ], [ 2, 2, 2, 4, 4, 4, 4 ], [ 3, 3 ],
  [ 3, 3 ], [ 3, 3 ], [ 4 ], [ 4 ], [ 9 ] ]
```

The KnotInfo website [5] includes a list of all prime knots that can be presented as a planar diagram with at most 11 crossings. For each knot the site links to an arc presentation of the knot given in the Knot Atlas [23]. A table of these arc presentations has been imported into [8] and used to produce a function that returns the arc presentation as a small 3-dimensional pure cubical complex K such as that in Figure 2 for the trefoil knot. Each such cubical complex has been constructed as a topological manifold. Mirror images have been excluded from the table. Figure 9 illustrates the pure cubical complex K representing the 49th knot on nine crossings.

For each prime knot with eleven or fewer crossings, excluding mirror images, we have computed the invariant $I^{[6,1,1]}(G)$ for G the fundamental group of the complement of our 3-dimensional pure cubical complex representation of the knot. Mirror images are excluded because it is well-known that the fundamental group of a knot complement does not distinguish between the knot and its mirror image. The computation, which took about 15 minutes on a laptop, yielded the following.

Theorem 2 *For a knot $K\colon S^1 \to \mathbb{R}^3$ define $G(K) = \pi_1(\mathbb{R}^3 \setminus K)$. The knot invariant*

$$I^{[6,1,1]}(G(K)) = \{S_{ab} \ : \ S \leq G(K), |G(K) : S| \leq 6\}$$

distinguishes, up to mirror image, between ambient isotopy classes of all prime knots that admit planar diagrams with eleven or fewer crossings.

A procedure for computing the classical Alexander polynomial, from a presentation of the group $G(K)$, has been implemented in [8]. There are 801 prime knots with 11 or fewer crossings (excluding mirror images and the trivial knot). The Alexander polynomial attains 550 distinct values on these 801 knots.

References

1. W.W. Boone. The word problem. *Ann. of Math. (2)*, 70:207–265, 1959.
2. P. Brendel, P. Dłotko, G. Ellis, M. Juda, and M. Mrozek. An algorithm for the fundamental group of a regular cw-space (provisional title). *in preparation*, 2014.
3. M. R. Bridson and M. Tweedale. Deficiency and abelianized deficiency of some virtually free groups. *Math. Proc. Cambridge Philos. Soc.*, 143(2):257–264, 2007.
4. The CAPD Group. *CAPD::RedHom - Reduction Homology Algorithms, the topological part of CAPD*, 2013. (`http://redhom.ii.uj.edu.pl/` and `http://capd.ii.uj.edu.pl/`).
5. J.C. Cha and C. Livingston. *KnotInfo: Table of Knot Invariants.* KnotInfo web page, 2013. (`http://www.indiana.edu/ knotinfo`).
6. M.M. Cohen. *A course in simple homotopy theory.* Graduate texts in mathematics. Springer, New York, 1973.
7. H. Edelsbrunner and J. Harer. *Computational Topology - an Introduction.* American Mathematical Society, 2010.
8. G. Ellis. *HAP – Homological Algebra Programming, Version 1.10.13*, 2013. (`http://www.gap-system.org/Packages/hap.html`).
9. G. Ellis and F. Hegarty. Computational homotopy of finite regular cw-spaces. *Journal of Homotopy and Related Structures*, pages 1–30, 2013.
10. Jeff Erickson and Kim Whittlesey. Greedy optimal homotopy and homology generators. In *Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1038–1046 (electronic). ACM, New York, 2005.
11. R. Forman. Morse theory for cell complexes. *Adv. Math.*, 134(1):90–145, 1998.
12. R. Forman. A user's guide to discrete Morse theory. *Sém. Lothar. Combin.*, 48:Art. B48c, 35, 2002.
13. The GAP Group. *GAP – Groups, Algorithms, and Programming, Version 4.5.6*, 2013. (`http://www.gap-system.org`).
14. R. Geoghegan. *Topological methods in group theory*, volume 243 of *Graduate Texts in Mathematics*. Springer, New York, 2008.
15. Rocío González-Díaz and Pedro Real. On the cohomology of 3D digital images. *Discrete Appl. Math.*, 147(2-3):245–263, 2005.
16. C. McA. Gordon and J. Luecke. Knots are determined by their complements. *J. Amer. Math. Soc.*, 2(2):371–415, 1989.
17. S. Harker, K. Mischaikow, M. Mrozek, and V. Nanda. Discrete morse theoretic algorithms for computing homology of complexes and maps. *Foundations of Computational Mathematics*, 2013.
18. S. Harker, K. Mischaikow, M. Mrozek, V. Nanda, H. Wagner, M. Juda, and P. Dłotko. The efficiency of a homology algorithm based on discrete morse theory and coreductions. *Proceedings of the 3rd International Workshop on Computational Topology in Image Context, Chipiona, Spain, November 2010*, 1:41–47, 2010.
19. D.W. Jones. A general theory of polyhedral sets and the corresponding T-complexes. *Dissertationes Math. (Rozprawy Mat.)*, 266:110, 1988.
20. T. Kaczynski, K. Mischaikow, and M. Mrozek. *Computational homology*, volume 157 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 2004.
21. T. Kaczynski, M. Mrozek, and M. Ślusarek. Homology computation by reduction of chain complexes. *Computers and Mathematics with Applications*, 34(4):59–70, 1998.
22. Junho Kim, Miao Jin, Qian-Yi Zhou, Feng Luo, and Xianfeng Gu. Computing fundamental group of general 3-manifold. In G. Bebis, R. Boyle, B. Parvin, D. Koracin, P. Remagnino, F. Porikli, J. Peters, J. Klosowski, L. Arns, Y. Chun, T.-M. Rhyne, and L. Monroe, editors, *Advances in Visual Computing*, volume 5358 of *Lecture Notes in Computer Science*, pages 965–974. Springer Berlin Heidelberg, 2008.
23. Knot Atlas. *Knot Atlas*, 2013. (`http://katlas.math.toronto.edu/wiki/`).
24. D. Letscher. On persistent homotopy, knotted complexes and the Alexander module. In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*, ITCS '12, pages 428–441, New York, NY, USA, 2012. ACM.
25. W.S. Massey. *A basic course in algebraic topology*, volume 127 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1991.
26. P. S. Novikov. *Ob algoritmičeskoĭ nerazrešimosti problemy toždestva slov v teorii grupp.* Trudy Mat. Inst. im. Steklov. no. 44. Izdat. Akad. Nauk SSSR, Moscow, 1955.
27. D. Nureki, K. Watanabe, S. Fukai, R. Ishii, Y. Endo, H. Hori, and S. Yokoyama. Deep knot structure for construction of active site and cofactor binding site of trna modification enzyme. *Structure*, 12, 593, 2004. (`http://www.rcsb.org/pdb/explore/explore.do?structureId=1V2X`).
28. J.H. Palmieri et al. *Finite Simplicial Complexes.* Sage v5.10, 2009. (`http://www.sagemath.org/doc/reference/homology/sage/homology/simplicial_complex.html`).
29. S. Rees and L.H. Soicher. An algorithmic approach to fundamental groups and covers of combinatorial cell complexes. *J. Symbolic Comput.*, 29(1):59–77, 2000.
30. E.H. Spanier. *Algebraic Topology, corrected reprint of the 1966 original.* Springer, New York, 1981.
31. W. A. Stein et al. *Sage Mathematics Software (Version 5.10).* The Sage Development Team, 2013. (`http://www.sagemath.org`).
32. F. Waldhausen. On irreducible 3-manifolds which are sufficiently large. *Ann. of Math. (2)*, 87:56–88, 1968.
33. J. H. C. Whitehead. Combinatorial homotopy. I. *Bull. Amer. Math. Soc.*, 55:213–245, 1949.
34. J. H. C. Whitehead. Combinatorial homotopy. II. *Bull. Amer. Math. Soc.*, 55:453–496, 1949.
35. J. H. C. Whitehead. Simple homotopy types. *Amer. J. Math.*, 72:1–57, 1950.