Homology of simplicial complexes in Haskell

Anders Mörtberg

`mortberg@chalmers.se`

June 9, 2011

1 Introduction

The main motivation in the ForMath project for the formalization of the Smith normal form algorithm is to *compute of homology groups of simplicial complexes*. The goal of this document is to explain what this means and show how it can be represented in Haskell. A small application of computing the number of connected components and holes in digital images is shown in the end of the document.

The general theory will be explained but the actual computations will be restricted to the case when the underlying ring is a field which makes things easier as the computations will be reduced to computations over vector spaces and we can then use well known methods and theorems from linear algebra.

2 Simplicial complexes

Simplicial complexes are a combinatorial description of topological spaces suitable for computation. A *vertex set* is a nonempty ordered set.

Definition. A *simplicial complex* X over a vertex set V is a subset of the powerset 2^V such that $A \subseteq B \in X$ implies that $A \in X$.

Note that there is no restriction on the size of X so simplicial complexes can be infinite, but in the rest of this document they are assumed to be finite. Hence they can easily be represented on a computer using lists.
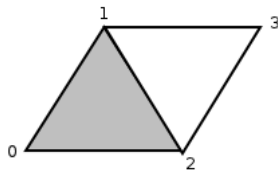
```
-- The type of a simplex, it is assumed that there is an ordering on a type
type Simplex a = [a]

-- Type of simplicial complexes
type SC a = [Simplex a]
```

Example. A simple example of a simplicial complex (with vertex set $\mathbb{N}$) is

$$\mathcal{S} = \{\emptyset, \{0\}, \{1\}, \{2\}, \{3\}, \{0,1\}, \{0,2\}, \{1,2\}, \{1,3\}, \{2,3\}, \{0,1,2\}\}$$

1

It can be visualized as

and represented in HASKELL as

```
ex :: SC Z
ex = [[],[0],[1],[2],[3],[0,1],[0,2],[1,2],[1,3],[2,3],[0,1,2]]
```

The above example should give the geometric intuition of simplicial complexes and show that they can be viewed as higher dimensional graphs.

An element of X is called a *face* and a maximal element with respect to inclusion is called a *facet*.

Example. The facets of $\mathcal{S}$ are $\{1,3\}$, $\{2,3\}$ and $\{0,1,2\}$. This can be computed in HASKELL by

```
> facets ex
[[1,3],[2,3],[0,1,2]]
```

The *n-simplices* of a simplicial complex is the faces of cardinality $n+1$.

Example. The 1-simplices of X are $\{0,1\}$, $\{0,2\}$, $\{1,2\}$, $\{1,3\}$ and $\{2,3\}$.

3   Homology of simplicial complexes

Let X be a simplicial complex and let $C_n(X,R)$ be the free abelian group with basis the set of n-simplices in X with coefficient in R. An element in $C_n(X,R)$ is called a *n-chain*. The n-chains are formal sums $\sum_i n_i A_i$ where $n_i \in R$ and $A_i \in X$ where A_i is an i-simplex. The most common choices for R in algebraic topology are $\mathbb{Z}$, $\mathbb{Q}$ or $\mathbb{Z}_p$ for p prime.

Definition. A *boundary map* is a map $\partial_n : C_n(X,R) \to C_{n-1}(X,R)$ defined by

$$\partial_n(A_n) = \sum_{i=0}^{n} (-1)^i \left(A_n \setminus \{a_i\} \right)$$

where A_n is a n-simplex of X.

Example. An example of a boundary map is $\partial_2 : C_2(\mathcal{S},\mathbb{Z}) \to C_1(\mathcal{S},\mathbb{Z})$. For this map we have

$$\partial_2(\{0,1,2\}) = \{1,2\} - \{0,2\} + \{0,1\}$$

2

One can prove that $\partial_{n-1} \circ \partial_n = 0$ which give that $im(\partial_n) \subseteq ker(\partial_{n-1})$ as if $x \in C_n(X, R)$ then $\partial_n(x) \in C_{n-1}(X, R)$ which is mapped to $0 \in C_{n-2}(X, R)$ by ∂_{n-1} but this simply means that $\partial_n(x) \in ker(\partial_{n-1})$. This has been implemented in HASKELL and tested using QUICKCHECK:

```
propIncidenceMatrix :: Int → SC Z → Property
propIncidenceMatrix n xs =
  n > 0 ⟹ allZero (incidenceMatrix n xs 'mult' incidenceMatrix (n-1) xs)

> quickCheck propIncidenceMatrix
+++ OK, passed 100 tests.
```

This has also been implemented and proved correct using COQ and SSREFLECT by the La'Rioja node of the ForMath project [4].

Hence we get a *chain complex* of n-chains:

$$\ldots \longrightarrow C_n(X, R) \xrightarrow{\partial_n} C_{n-1}(X, R) \longrightarrow \ldots \longrightarrow C_1(X, R) \xrightarrow{\partial_1} C_0(X, R) \xrightarrow{\partial_0} 0$$

Which means that we can define

Definition. The *n:th homology group* $H_n(X, R) = ker(\partial_n)/im(\partial_{n+1})$.

The intuition behind the homology groups is that they measure the difference between the simplices in different dimension. One can prove that $H_0(X, \mathbb{Z})$ measure the number of connected components in X[1]. For higher dimensions the homology groups can be seen as measuring the number of (topological) holes in each dimension.

So how do we compute the homology groups? This is where the Smith normal form algorithm comes into play. If $R = \mathbb{Z}$ (or any other principal ideal domain) then the Smith normal algorithm give a diagonalization process which can be used to compute the free part and torsion part of finitely generated abelian groups as proved in the structure theorem. Using this one obtains an algorithm for computing the homology groups, this is explained in greater detail in [1].

But if we restrict our attention to the case when R is a field k the questions is instead how to compute the kernel and image of linear transformations between vector spaces, which we know how to do from linear algebra.

In order to do this first we need to represent the boundary maps as matrices, these matrices are called *incidence matrices* in [4]. The main idea is that we let the columns represent n-simplices and the rows (n-1)-simplices and let the a_{ij} entry be either 0 or ± 1 depending on the value of the coefficient of the i:th (n-1)-simplex in the boundary map of the j:th n-simplex.

Example. The first incidence matrix, M_1, of $\mathcal{S}$ is given by:

$$\begin{pmatrix} -1 & -1 & 0 & 0 & 0 \\ 1 & 0 & -1 & -1 & 0 \\ 0 & 1 & 1 & 0 & -1 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix}$$

---

[1] See the proof of proposition 2.7 in [3].

as for example $\partial(\{0,1\}) = \{1\} - \{0\}$. This give a chain complex

$$0 \longrightarrow C_2(\mathcal{S}, R) \xrightarrow{M_2} C_1(\mathcal{S}, R) \xrightarrow{M_1} C_0(X, R) \xrightarrow{M_0} 0$$

where M_2, M_1 and M_0 are the incidence matrices of ∂_2, ∂_1 and ∂_0 respectively. The higher n-chains are all zero as there are no simplices in higher dimensions.

This can be computed in HASKELL by

```
> incidenceMatrix 1 ex
[[-1,-1,0,0,0],[1,0,-1,-1,0],[0,1,1,0,-1],[0,0,0,1,1]]
```

Now lets focus on the homology groups of a general chain complex obtained from the n-chains and incidence matrices of a simplicial complex. These are defined by $H_n(X, k) = ker(M_n)/im(M_{n+1})$ where M_i are the incidence matrix associated to the boundary map ∂_i. From linear algebra[2] we know that the quotient space is isomorphic to k^d where $d = dim(ker(M_n)) - dim(im(M_{n+1}))$. First note that the dimension of the image of M_{n+1} is the rank of M_{n+1}. The rank-nullity theorem[3] now give that $dim(ker(M_n)) = c - dim(im(M_n)) = c - rank(M_n)$ where c is the number of columns of M_n, that is, the number of n-simplices of X. Hence we get

$$H_n(X, k) \cong k^{c - rank(M_n) - rank(M_{n+1})}$$

which means that the only thing we need is an algorithm for computing the rank of a matrix for computing the homology groups when the coefficient lie in a field. To do this we can use the Smith normal form algorithm as it is a generalization of Gaussian elimination.

Example. If we fix $k = \mathbb{Z}_2$ we can apply the Smith normal form algorithm to M_1 of $\mathcal{S}$ and get that the rank is 3. Similarly we can compute that the rank of M_2 is 1. Hence we get that

$$H_0(\mathcal{S}, \mathbb{Z}_2) = (\mathbb{Z}_2)^{4-0-3} = \mathbb{Z}_2^1$$

as we have four 0-simplices. This should be interpreted as $\mathcal{S}$ only have component. The next homology group is

$$H_1(\mathcal{S}, \mathbb{Z}_2) = (\mathbb{Z}_2)^{5-3-1} = \mathbb{Z}_2^1$$

As there are five 1-simplices. The interpretation of this is that there are one hole in $\mathcal{S}$. For higher dimensions the homology groups are all zero. This can be computed in HASKELL using a specialized implementation of the Smith normal form algorithm for sparse matrices over $\mathbb{Z}_2$:

---

[2]http://en.wikipedia.org/wiki/Quotient_space_(linear_algebra)
[3]http://en.wikipedia.org/wiki/Rank-nullity_theorem

```
> homology 0 ex
1
> homology 1 ex
1
> homology 2 ex
0
```

Here some details has been swept under the carpet. The homology groups are computed over $\mathbb{Z}_2$ and not $\mathbb{Z}$ which leads to the question why we can draw any conclusion about the number of components and holes at all? The answer to this come from the universal coefficient theorem[4] for homology. As the above example is torsion-free we get that the homology groups computed with coefficient in $\mathbb{Z}_2$ are isomorphic to the homology groups computed with coefficients in $\mathbb{Z}$. The same will hold for the examples of simplicial complexes obtained from digital images in the next section.

4   Application – Digital images

There is a natural way of associating 2D black and white images with simplicial complexes explained in detail in [2]. This has been implemented in Haskell together with a small implementation of terminal based "images" of characters:

```
> drawImage b

###
#  #
###
#  #
###

*Talk> length $ imageToSC b
110
```

So the simplicial complex related to the image showing a "B" has 110 faces. We can now compute the homology groups of the simplicial complex related to the letter "B":

```
> homology 0 $ imageToSC b
1
> homology 1 $ imageToSC b
2
> homology 2 $ imageToSC b
0
```

So there is one component (the letter) and two holes which is what is expected for "B". There is also a small program in which one can write a string and get the number of connected components and holes:

---

[4]http://en.wikipedia.org/wiki/Universal_coefficient_theorem

```
Please type some text:
homology

You wrote:

#   #   ##   #   #   ##   #       ##    ###   #   #
#   #   #  #   ## ##   #   #   #     #   #   #       # #
####   #   #   # # #   #   #   #     #   #   #   ##     #
#   #   #   #   #   #   #   #   #   #     #   #   #   #     #
#   #   ##   #   #   ##   ####   ##    ###      #


Zero homology group: Z^8
Number of connected components: 8

First homology group: Z^3
Number of holes: 3
```

This is just a small quite trivial example but it is of course possible to use the algorithms for analyzing more interesting pictures. The methods also scale to higher dimensions naturally [2].

I will end with an overview of the program:

1. Representation of digital images

2. Compute simplicial complexes from the images

3. Compute incidence matrices

4. Triangulate using Smith normal form

5. Compute homology

References

[1] J.-G. Dumas, F. Heckenbach, B. D. Saunders, and V. Welker. Computing Simplicial Homology Based on Efficient Smith Normal Form Algorithms. In *Algebra, Geometry, and Software Systems*, pages 177–206. Springer, 2003.

[2] ForMath La Rioja node. From Digital Images to Simplicial Complexes: A Report. Technical report, Departamento de Matemáticas y Computación, Universidad de La Rioja, 2011.

[3] A. Hatcher. *Algebraic Topology*. Cambridge University Press, 2002.

[4] J. Heras, M. Poza, M. Denes, and L. Rideau. Incidence simplicial matrices formalized in coq/ssreflect. Lecture Notes in Computer Science, 2011.