Spectral sequences for computing persistent homology of digital images[*]

Ana Romero, Gadea Mata, Julio Rubio
University of La Rioja (Spain)

Jónathan Heras
University of Dundee (UK)

Francis Sergeraert
University Joseph Fourier (France)

`ana.romero@unirioja.es`

Abstract

Persistent homology and spectral sequences are two Algebraic Topology tools which are defined by means of a filtration and can be applied to study topological properties of a space at different stages. Both concepts are deeply related, and this relation allows us to use some previous programs developed for computing spectral sequences of filtered complexes to determine now persistent homology. In particular, spectral sequences can be applied to compute persistent homology of digital images, which will allow us to determine relevant features, that will be long-lived on contrast with the "noise" which will be short-lived.

Keywords
Persistent homology, digital images, spectral sequences.

1 Introduction

Persistent homology [4] is an algebraic method for measuring topological features of shapes and functions, with many recent applications such as point cloud data, sensor networks, optical character recognition and protein classification. More concretely, this technique consists in identifying homological features that persist within the different stages of a filtration. On the other hand, spectral sequences [7] are a tool for computing homology groups by taking successive approximations. Both concepts are defined by means of a filtration and are deeply related.

In a previous paper [8], we showed that a slight modification of our previous programs for computing spectral sequences [9] is enough to compute also persistent homology. By inheritance from our spectral sequence program, we obtained for free persistent homology programs applicable to spaces not of finite type (provided they are spaces with effective homology) and with $\mathbb{Z}$-coefficients (significantly generalizing the usual presentation of persistent homology over a field). Moreover, our calculations made it possible to detect an error in [4]: the so called "Spectral sequence theorem" [4, p. 171], which shows the relation between spectral sequences and persistent homology, includes a formula which is not correct (see [8] for details).

In this work, we use our spectral sequence programs to compute persistent homology of digital images. This allows us to determine relevant features, that will be long-lived – in the sense that they persist over a certain parameter range – on contrast with the "noise" which will be short-lived. In order to reduce the time of calculations, we can use the combinatorial notion of *Discrete Vector Field* [5]. As a test case, our programs could be applied on a fingerprint database.

---

[*] Partially supported by Ministerio de Educación y Ciencia, project MTM2009-13842-C02-01, and by the European Union's 7th Framework Programme under grant agreement nr. 243847 (ForMath).

2 Preliminaries

Definition 2.1. Let K be a simplicial complex. A (finite) *filtration* of K is a nested sequence of subcomplexes $K^i \subseteq K$ such that $\emptyset = K^0 \subseteq K^1 \subseteq K^2 \subseteq \cdots \subseteq K^m = K$.

For every $i \leq j$ we have an inclusion map on the canonically associated chain complexes $\mathrm{inc}^{i,j} : C(K^i) \hookrightarrow C(K^j)$ and therefore we can consider the induced homomorphisms $f_n^{i,j} : H_n(K^i) \to H_n(K^j)$, for each dimension n. The filtration produces then for each dimension n a sequence of homology groups connected by homomorphisms:

$$0 = H_n(K^0) \to H_n(K^1) \to \cdots \to H_n(K^m) = H_n(K)$$

Definition 2.2. The n-th *persistent homology groups* of K, denoted by $H_n^{i,j}(K) \equiv H_n^{i,j}$, are the images of the homomorphisms $f_n^{i,j}$:

$$H_n^{i,j} = \operatorname{Im} f_n^{i,j}, \text{ for } 0 \leq i \leq j \leq m$$

The group $H_n^{i,j}$ consists of the n-th homology classes of K^i that are still alive at K^j. A class $\gamma \in H_n(K^i)$ is said to *be born* at K^i if $\gamma \notin H_n^{i-1,i}$. It is said to *die* entering K^j if it merges with an older class as we go from K^{j-1} to K^j, that is, $f_n^{i,j-1}(\gamma) \notin H_n^{i-1,j-1}$ but $f_n^{i,j}(\gamma) \in H_n^{i-1,j}$. If γ is born at K^i and dies entering K^j, the difference $j - i$ is called the *persistence index* of γ, denoted $\operatorname{pers}(\gamma)$. If γ is born at K^i but never dies then $\operatorname{pers}(\gamma) = \infty$.

If the homology is computed with field coefficients, each group $H_n^{i,j}$ is a vector space which is determined up to isomorphism by its dimension, and this allows one to represent all persistent homology groups by means of a *barcode* diagram [4]. However, in the integer case one can face extension problems. In order to solve this difficulty, we introduced in [8] a generalization of persistent homology with $\mathbb{Z}$-coefficients. This can be done by means of a double filtration which leads to a new (more general) definition of barcode.

Definition 2.3. Let R be a ring, a *spectral sequence* $E = (E^r, d^r)_{r\geq 1}$ is a sequence of bigraded R-modules $E^r = \{E^r_{p,q}\}_{p,q\in\mathbb{Z}}$, each provided with a differential $d^r = \{d^r_{p,q} : E^r_{p,q} \to E^r_{p-r,q+r-1}\}_{p,q\in\mathbb{Z}}$ of bidegree $(-r, r-1)$ (satisfying $d_{p-r,q+r-1} \circ d_{p,q} = 0$) and with isomorphisms $H(E^r, d^r) \cong E^{r+1}$ for every $r \geq 1$. Since each $E^{r+1}_{p,q}$ is a subquotient of $E^r_{p,q}$, one can define the *final groups* $E^\infty_{p,q}$ of the spectral sequence as the groups which remain after the computation of all successive homologies.

Theorem 2.4. *[7, p.327] Let C be a chain complex with a filtration. There exists a spectral sequence $E \equiv E(C) \equiv (E^r, d^r)_{r\geq 1}$, defined by*

$$E^r_{p,q} = \frac{Z^r_{p,q} + C^{p-1}_{p+q}}{d_{p+q+1}(Z^{r-1}_{p+r-1,q-r+2}) + C^{p-1}_{p+q}}$$

where $Z^r_{p,q}$ is the submodule $Z^r_{p,q} = \{a \in C^p_{p+q} | \, d_{p+q}(a) \in C^{p-r}_{p+q-1}\} \subseteq C^p_{p+q}$, and $d^r_{p,q} : E^r_{p,q} \to E^r_{p-r,q+r-1}$ is the morphism induced on these subquotients by the differential map $d_{p+q} : C_{p+q} \to C_{p+q-1}$. This spectral sequence converges *to the homology groups of C, that is, there are natural isomorphisms*

$$E^\infty_{p,q} \cong \frac{H^p_{p+q}(C)}{H^{p-1}_{p+q}(C)}$$

where $H^p_*(C)$ is the filtration on the homology groups $H_*(C)$ induced by the filtration of C.

3 Computing persistent homology by means of spectral sequences

There are some works in the literature which include some comments on the relation between spectral sequences and persistent homology (see for instance [12] and [3]), but the only reference where we have found an explicit formula which relates them is the book "Computational Topology: An Introduction" by Herbert Edelsbrunner and John Harer [4]. Given a filtered simplicial complex K, the so called "Spectral sequence theorem" ([4, p. 171]) claims that:

The total rank of the groups of dimension $p+q$ in the level $r \geq 1$ of the associated spectral sequence equals the number of points in the $(p+q)$-th persistence diagram whose persistence is r or larger, that is,

$$\sum_{p=1}^{m} \operatorname{rank} E^r_{p,q} = \operatorname{card}\{a \in \operatorname{Dgm}_{p+q}(f) \mid \operatorname{pers}(a) \geq r\}$$

where $\operatorname{Dgm}_{p+q}(f)$ is an appropriate persistence diagram (see [4, Chap. VII]) and where in the left side q decreases as p increases so that the dimension $p+q$ remains constant.

However, we have detected that the formula in [4] is erroneous because in the spectral sequence side (the left side) there can be more elements than in the persistence (right) side; the formula should be therefore an inequality. To illustrate the error in [4], it suffices to consider as a counterexample a simplicial complex K generated by the interval ab, with the filtration given by $K^1 = \{a, b\}$ and $K^2 = K$; in dimension 1 one has $E^1_{2,-1} = \mathbb{Z}$ but there are no classes of persistence at least 1 since the unique element of dimension 1 is not a cycle.

The correct relation between persistent homology and spectral sequences can be expressed by the following theorem:

Theorem 3.1. *[8] The total rank of the images of the differential maps in the level $r \geq 1$ of the spectral sequence equals the number of points in the $(p+q)$-th persistence diagram whose persistence is r:*

$$\sum_{p=1}^{m} \operatorname{rank} A^r_{p,q} = \operatorname{card}\{a \in \operatorname{Dgm}_{p+q}(f) \mid \operatorname{pers}(a) = r\}$$

where $A^r_{p,q} = \operatorname{Im}(d^r_{p+r,q-r+1} : E^r_{p+r,q-r+1} \to E^r_{p,q}) \subseteq E^r_{p,q}$.

This theorem gives us an algorithm for computing the rank of the persistent homology groups of a filtered simplicial complex from the associated spectral sequence. Let us emphasize that this information about ranks determines (up to isomorphism) the groups $H^{i,j}_n$ when one works with coefficients over a field F. Therefore, if we know the groups $E^r_{p,q}$ and the differential maps $d^r_{p,q}$ of the spectral sequence of a filtered simplicial complex, thanks to the formula introduced in Theorem 3.1 we can also easily determine the persistent homology groups of K. If we work with coefficients over $\mathbb{Z}$, the previous information about the ranks relating spectral sequences and persistent homology is not sufficient to determine the groups $H^{i,j}_n$; however, we will see later that one can express the groups $H^{i,j}_n$ in terms of some subgroups appearing in the definition of the spectral sequence, which will allow us to determine $H^{i,j}_n$ also in the integer case.

In a previous work [9], we developed a set of programs computing spectral sequences associated with filtered chain complexes. These programs were implemented in Common Lisp as a new module for the Kenzo system [2], a computer algebra program developed by the last author of this paper and some coworkers which implements the *effective homology* theory [11] and has made it possible to determine homology and homotopy groups of complicated (infinite) spaces. The new programs for spectral sequences use also the effective homology technique and allow the Kenzo user to determine the different components of spectral sequences of filtered complexes even in some cases where the chain complex has infinite type. Using our programs, and thanks to Theorem 3.1, one can determine in this way the ranks of the groups $H^{i,j}_n$.

In fact the computation of the groups $H^{i,j}_n$ can be directly obtained by a small modification of our algorithms without doing the complete process of computing the corresponding groups and differential maps of the spectral sequence. Let us recall that a group $E^r_{p,q}$ in the spectral sequence is given by the formula:

$$E^r_{p,q} = \frac{Z^r_{p,q} + C^{p-1}_{p+q}}{d_{p+q+1}(Z^{r-1}_{p+r-1,q-r+2}) + C^{p-1}_{p+q}}$$

We can observe that each class in $E^r_{p,q}$ is generated by an "almost" cycle of dimension $p+q$ (a chain whose boundary in $K^p - K^{p-r}$ is empty but which may have non-empty boundary in K^{p-r}), and the elements of $E^r_{p,q}$ given by a real cycle x (that is, $d(x) = 0$), correspond to classes of $H_{p+q}(K^p)$ which are born at K^p and are still alive at K^{p+r-1}, and then the persistence indexes of these classes are at least r.

It is not difficult to observe then that the groups $H^{i,j}_n$ can also be described as a quotient:

$$H^{i,j}_n = \frac{\operatorname{Ker} d_n \cap C^i_n}{d_{n+1}(Z^{j-i}_{j,n-j+1})} = \frac{Z^i_{i,n-i}}{d_{n+1}(Z^{j-i}_{j,n-j+1})}$$

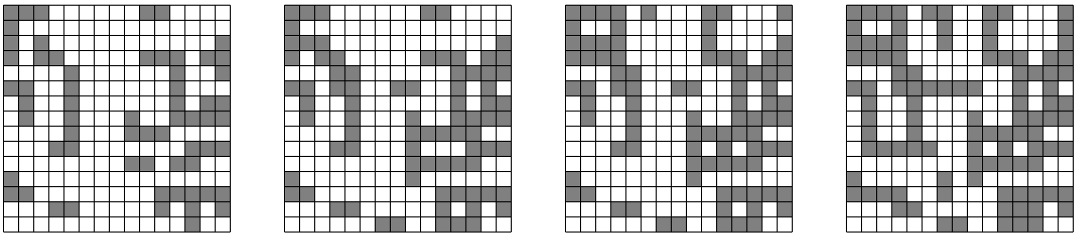

Figure 1: Filtered digital image.

If K is a finite filtered simplicial complex, then our programs determine the different elements of the associated spectral sequence by means of some elementary operations on matrices. More concretely, the programs determine in particular the subgroups $Z^r_{p,q}$, C^{p-1}_{p+q} and $d_{p+q+1}(Z^{r-1}_{p+r-1,q-r+2})$ which appear in the formula of Theorem 2.4 (which can be determined if K is finite), and then calculate the desired quotient. The groups $H^{i,j}_n$ are determined in terms of similar subgroups and then it has been very easy to adapt our programs in order to compute also $H^{i,j}_n$ for finite (filtered) simplicial complexes. It is important to remark that this is also valid in the integer case and this makes it possible to solve the possible extension problems. Our programs can also be applied in the infinite case, where the effective homology method can be used to determine the groups $H^{i,j}_n$ by means of a *reduction* of the initial chain complex C to an auxiliary chain complex of finite type (see [8] for details).

4 Persistent homology of digital images

Given a digital image, we can naturally associate a simplicial complex K and compute its homology groups in dimensions 0 and 1 which show respectively the number of connected components and holes that the image contains. If the image is *filtered* (for example, it comes from a stack of images), one can also determine the persistent homology groups which will allow us to determine relevant features, that will be long-lived – in the sense that they persist over a certain parameter range – on contrast with the "noise" which will be short-lived.

Let us consider the filtered image of Figure 1. The final homology groups are $H_0 = \mathbb{Z}^7$ and $H_1 = \mathbb{Z}^4$. We can see the *evolution* of the corresponding homology classes along the four filtration steps by using our programs for computing persistent homology groups based on spectral sequences. For example, $H^{1,4}_0 = \mathbb{Z}^4$, which means that in dimension 0 there are 4 classes which are born at the first step and are still alive at (the last) step 4:

```
> (prst-hmlg-group K 1 4 0)
Persistent Homology H^{1,4}_0
Component Z
Component Z
Component Z
Component Z
```

Similarly, $H^{2,4}_1 = \mathbb{Z}^2$ means that there are 2 holes at stage 2 which are still alive at step 4:

```
> (prst-hmlg-group K 2 4 1)
Persistent Homology H^{2,4}_1
Component Z
Component Z
```

These same results have been also obtained by a *certified* program, executed inside the Coq proof assistant (this kind of *verified* programs have been developed in the frame of the ForMath European project [1], and have been documented in [6]).

For bigger digital images, we can reduce the time of calculations by using the combinatorial notion of *Discrete Vector Field*, which is an essential component of Forman's Discrete Morse Theory [5], adapted to the algebraic setting in [10]. As explained in [10], given a digital image, an admissible discrete vector field can be constructed by means of some elementary operations on the differential matrices of the associated chain complex. This vector field produces a *reduction* from the initial (big) chain complex to a (much) smaller one whose homology groups are explicitly isomorphic to the homology groups of the image, so that the computation of these homology groups can be done in a more efficient way.

If we are interested in computing persistent homology groups, we can follow a similar process to construct a discrete vector field and reduce the initial (big) chain complex. In this case the discrete vector field must be compatible with the filtration, which can be done applying the same elementary methods of [10] to the differential submatrices corresponding to each step of the filtration. The vector field so obtained is of course smaller than the non-filtered one, but it usually decreases significantly the number of generators. This vector field produces again a reduction, which in this case is compatible with the given filtration, which implies that the persistent homology groups of the initial image are isomorphic to the persistent homology groups of the reduced one (see [8] for details). Applying now our programs for computing persistent homology to the small chain complex, we can compute the persistent homology groups of big images in an efficient way.

Computation of persistent homology groups of digital images could be applied to study fingerprints. Given a fingerprint image, we could filter it taking at the first step some initial horizontal lines, adding at each stage of the filtration some additional lines and ending with the whole image. This filtration would produce some persistent homology groups. A similar process could be done in the vertical direction, taking successively the columns of the image, producing in that way different persistent homology groups. It seems natural that given two (different) fingerprint images corresponding to the same person, the so obtained persistent homology groups should be similar. Persistent homology could help in this way for fingerprint recognition.

References

[1] *ForMath: Formalisation of Mathematics*, `http://wiki.portal.chalmers.se/cse/pmwiki.php/ForMath/ForMath`.

[2] X. Dousson, J. Rubio, F. Sergeraert, and Y. Siret, *The Kenzo program*, Institut Fourier, Grenoble, 1999, `http://www-fourier.ujf-grenoble.fr/~sergerar/Kenzo/`.

[3] H. Edelsbrunner and J. Harer, *Persistent homology a survey*, Contemporary Mathematics (2008), 1–26.

[4] H. Edelsbrunner and J. Harer, *Computational topology: An introduction*, Applied mathematics, American Mathematical Society, 2010.

[5] R. Forman, *Morse theory for cell complexes*, Advances in Mathematics **134** (1998), 90–145.

[6] J. Heras, T. Coquand, A. Mörtberg, and V. Siles, *Computing Persistent Homology within Coq/SSReflect*, To appear in ACM Transactions on Computational Logic, 2013.

[7] S. MacLane, *Homology*, vol. 114, Springer, 1963.

[8] A. Romero, J. Heras, J. Rubio, and F. Sergeraert, *Defining and computing persistent $\mathbb{Z}$-homology in the general case*, Preprint, 2013.

[9] A. Romero, J. Rubio, and F. Sergeraert, *Computing spectral sequences*, Journal of Symbolic Computation **41** (2006), no. 10, 1059–1079.

[10] A. Romero and F. Sergeraert, *Discrete Vector Fields and fundamental Algebraic Topology*, Preprint. `http://arxiv.org/abs/1005.5685v1`, 2010.

[11] J. Rubio and F. Sergeraert, *Constructive Algebraic Topology*, Bulletin des Sciences Mathématiques **126** (2002), no. 5, 389–412.

[12] A. Zomorodian and G. Carlsson, *Computing persistent homology*, Discrete and Computational Geometry **33** (2005), no. 2, 249–274.