# Computation of persistent homology

Ulrich Bauer

August 13, 2018

## 1 Computing homology

We will now describe a matrix algorithm for computing (persistent) homology with coefficients in a fixed field $\mathbb{F}$. As usual, we assume $\mathbb{F} = \mathbb{Z}_2$ for simplicity, mentioning however when other choices of coefficients require modifications. The basic idea is to find a compatible basis for the subspaces $B_*(K) \subseteq Z_*(K) \subseteq C_*(K)$.

**Remark 1.** Let $\Sigma_U \subset \Sigma_V$ be bases of vector spaces $U \subset V$. Then $(\Sigma_V \setminus \Sigma_U)$ is a set of representatives for a basis of the quotient space $V/U$.

*Proof.* Let $\Sigma_U = \{v_1, \ldots, v_m\}$, $\Sigma_V = \{v_1, \ldots, v_n\}$. We first show that $\Sigma_W = \{v_{m+1} + U, \ldots, v_n + U\}$ spans $V/U$, and then show that $\Sigma_W$ is linearly independent.

Let $v \in V$. Then there are unique coefficients $\lambda_i$ such that $v = \sum_{i=1}^n \lambda_i v_i$. Now

$$v + U = \sum_{i=1}^{n} \lambda_i v_i + U = \sum_{i=m+1}^{n} \lambda_i (v_i + U),$$

and thus $\Sigma_W$ spans $V/U$.

In order to show that $\Sigma_W$ are linearly independent, we need to show that $\lambda_i = 0$ for $i = m + 1, \ldots, n$ whenever $v + U = 0$, i.e., whenever

$$\sum_{i=m+1}^{n} \lambda_i v_i \in U.$$

Since $\Sigma_U$ is a basis for $U$, we get $(\lambda_i)_{i=1}^{m}$ such that

$$\sum_{i=1}^{m} \lambda_i v_i = \sum_{i=m+1}^{n} \lambda_i v_i.$$

But the $v_i$ are linearly independent, and thus $\lambda_i = 0$ for all $i = 1, \ldots, n$. □

Consider a simplicial complex $K$. The *boundary matrix* of $K$ is the matrix $D$ of the boundary map $\partial : C_*(K) \to C_*(K)$ with respect to the basis given by the simplices $\{\sigma_1, \ldots, \sigma_n\}$ of $K$.

**Definition 2.** Let $M$ be a matrix. The *pivot index* of a column $m_i$ of $M$ is

$$\mathrm{pivot}(m_i) = \min\{j \in \mathbb{N} : m_{ik} = 0 \text{ for all } k > j\}.$$

We write

$$\mathrm{pivots}\, M = \{\mathrm{pivot}\, m_j : 1 \leq j \leq n\} \setminus \{0\}.$$

A matrix $M$ is *reduced* if the pivots of non-zero columns are distinct, i.e., $i \neq j$ implies either $\mathrm{pivot}(m_i) \neq \mathrm{pivot}(m_j)$ or $0 = \mathrm{pivot}(m_i) = \mathrm{pivot}(m_j)$.

**Lemma 3.** *The nonzero columns of a reduced matrix $M$ are linearly independent.*

*Proof.* Let $(\lambda_j)_j$ be such that

$$v = \sum_j \lambda_j m_j = 0.$$

Then

$$0 = \operatorname{pivot} v = \max\{\operatorname{pivot} m_j \mid \lambda_j \neq 0\}$$

by the assumption that the pivots are unique. Thus, $m_j \neq 0$ implies $\lambda_j = 0$, meaning that the nonzero columns are linearly independent. □

Homology with field coefficients can be computed using a variant of Gaussian elimination. Specifically, the following algorithm finds an upper triangular matrix $V$ such that $R = D \cdot V$ is reduced.

**Algorithm 4** (matrix reduction)**.**

**Require:** $D$: $m \times n$ matrix
**Ensure:** $R = D \cdot V$ is reduced, $V$ is full rank upper triangular

```
function REDUCE(D)
    R = D, V = I(n)
    while there exist i < j with pivot R_i = pivot R_j do
        add column R_i to column R_j
        add column V_i to column V_j
    end while
    return R, V
end function
```

Note that the reduced matrix $R = D \cdot V$ is not unique, and the algorithm is not deterministic. To obtain a deterministic variant of the algorithm, one may iterate over $j = 1, \ldots, n$. For coefficients other than $\mathbb{Z}_2$, the columns $R_i$ and $V_i$ have to be rescaled so that their pivot entries cancel to 0 in the column addition.

**Example 5.** Consider the empty triangle

$$K = \{\{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}.$$

The boundary matrix is

$$D = \begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

and a matrix reduction is given by

$$R = \begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} = D \cdot \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

We have $\operatorname{pivots} R = \{2, 3\}$; the non-zero column indices of $R$ are $N = \{4, 5\}$, the zero column indices are $P = \{1, 2, 3, 6\}$, and the non-pivot zero column indices are $E = P \setminus \operatorname{pivots} R = \{1, 6\}$.

**Theorem 6.** *The homology of K can be computed by reducing the boundary matrix D.*

*(a) The columns*

$$\Sigma_B = \{r_j \neq 0\}$$

*form a basis of $B_*(K)$.*

*(b) The columns*

$$\widetilde{\Sigma}_Z = \{V_i \mid r_i = 0\}$$

*form a basis of $Z_*(K)$.*

*(c) Exchanging each basis cycle*

$$v_i \in \widetilde{\Sigma}_Z \textit{ with } i \in \operatorname{pivots} R$$

*with the basis boundary*

$$r_j \in \Sigma_B \textit{ with } i = \operatorname{pivot} r_j = \operatorname{pivot} v_i$$

*extends $\Sigma_B$ to another basis of $Z_*(K)$,*

$$\Sigma_Z = \Sigma_B \cup \Sigma_E$$

*with*

$$\Sigma_E = \{v_i \mid r_i = 0, i \notin \operatorname{pivots} R\} \subseteq \widetilde{\Sigma}_Z.$$

*(d) The non-bounding basis cycles $\Sigma_E$ generate a basis*

$$\Sigma_H = \{[v_i] = v_i + B_d(K) \mid r_i = 0, i \notin \operatorname{pivots} R, d = \dim \sigma_i\}$$

*for $H_*(K)$.*

*(e) The Betti numbers $\beta_d(K)$ can be read off from R, without knowing V:*

$$\beta_d = \operatorname{card}\{i \in E : \dim \sigma_i = d\}$$

*(f) We can further extend $\Sigma_Z$ to a basis for $C_*(K)$:*

$$\begin{aligned}\Sigma_C &= \Sigma_Z \cup \{v_j \mid r_j \neq 0\} \\ &= \{r_j \mid r_j \neq 0\} \cup \{v_j \mid r_j \neq 0\} \cup \{v_i \mid r_i = 0,\ i \notin \operatorname{pivots} R\}.\end{aligned}$$

*The matrix of $\partial$ with respect to this basis has at most one nonzero entry in each row/column.*

*Proof.* (a) The columns of the boundary matrix $D$ are a generating set for $B_*$ by definition. Since $V$ is regular, the columns of $R = D \cdot V$ are a generating set as well. The columns in $\Sigma_B$ have distinct pivot indices and therefore have to be linearly independent, by Lemma 3. Therefore the columns $\Sigma_B$ form a basis.

(b) Let $z \in Z_*$. Since $V$ is regular, its columns $\{v_i\}_i$ form a basis for $C_*(K)$, and we can write

$$z = \sum_i \lambda_i v_i.$$

Now

$$0 = \partial z = \sum_i \lambda_i \partial v_i = \sum_i \lambda_i r_i$$

implies, with the fact that $\Sigma_B$ is a basis, that

$$\lambda_j = 0$$

for all $j$ with $r_j \neq 0$. Therefore we can write

$$z = \sum_{i: r_i = 0} \lambda_i v_i,$$

and so $\widetilde{\Sigma}_Z$ is a generating set for $Z_*(K)$. The columns $v_i$ have again distinct pivot indices and are thus linearly independent. We conclude that $\widetilde{\Sigma}_Z$ is a basis for $Z_*(K)$.

(c) Replacing all $v_i$ by $r_j$ for $i = \operatorname{pivot} r_j = \operatorname{pivot} v_i$ yields $\operatorname{pivots} \widetilde{\Sigma}_Z = \operatorname{pivots} \Sigma_Z$, so $\Sigma_Z$ also has distinct pivots and is therefore also a maximal linearly independent set of cycles. Thus $\Sigma_Z$ is another basis for $Z_*(K)$.

(d) This follows from Remark 1 applied to $\Sigma_B \subseteq \Sigma_Z$.

(e) The dimension $\beta_d(K)$ of $H_d(K)$ is the cardinality of its basis.

(f) Similarly to before, $\Sigma_C$ is obtained from the columns of $V$ by exchanging vectors with the same pivot index, and therefore $\Sigma_C$ has distinct pivots and is thus a basis. We have $\partial v_j = r_j$ for all $r_j \neq 0$, while all other basis elements are cycles, implying the claim. □

## 2 Persistent Homology

As it turns out, the same algorithm can also be used to compute the persistent homology of a filtration.

Consider a *simplexwise filtration* $K_\bullet$ of a simplicial complex $K$ such that $K_i = K_{i-1} \cup \{\sigma_i\}$ for $1 \leq i \leq n$, where $\sigma_i \in K$ is a simplex. The *boundary matrix* of $K_\bullet$ is the matrix $D$ of the boundary map $\partial : C_*(K) \to C_*(K)$ with respect to the ordered basis $(\sigma_i)_{1 \leq i \leq n}$. In particular, $D$ is the boundary matrix of the final complex $K$.

**Remark 7.** The persistent homology of $K_\bullet$ can be computed by reducing the boundary matrix $D$:

(a) The columns

$$\Sigma_{B,k} = \{r_j \neq 0 \mid j \leq k\}$$

form a basis of $B_*(K_k)$.

(b) The columns

$$\widetilde{\Sigma}_{Z,k} = \{V_i \mid r_i = 0,\ i \leq k\}$$

form a basis of $Z_*(K)$.

(c) Exchanging each basis cycle

$$v_i \in \widetilde{\Sigma}_Z \text{ with } i \leq k \in \text{pivots}\, R$$

with the basis boundary

$$r_j \in \Sigma_B \text{ with } i = \operatorname{pivot} r_j$$

extends $\Sigma_{B,k}$ to another basis of $Z_*(K_k)$,

$$\Sigma_{Z_k} = \Sigma_{B_k} \cup \Sigma_{P_k} \cup \Sigma_{E_k}$$

with

$$\begin{aligned}\Sigma_{P,k} &= \{r_j \neq 0 \mid i = \operatorname{pivot} r_j,\ k \in [i, j)\},\\ \Sigma_{E,k} &= \{v_i \mid r_i = 0,\ i \neq \operatorname{pivots} R,\ k \in [i, \infty)\}.\end{aligned}$$

(d) The additional cycles generate a basis

$$\Sigma_{H,k} = \{[z] \mid z \in \Sigma_{P,k} \cup \Sigma_{E,k}\}$$

for the homology $H_*(K_k)$.

(e) The inclusion-induced map $H_*(K_k) \to H_*(K_l)$ has rank

$$\operatorname{card}(\Sigma_{H_k} \cap \Sigma_{H,l}).$$

**Remark 8.** Persistent homology (of a simplexwise filtration indexed by natural numbers $[n] = \{1, \ldots, n\}$) is a (commutative) diagram of vector spaces $V_i = H_*(K_i)$, indexed by $[n]$, and linear maps $f_{ij} : V_i \to V_j$ $(i \leq j)$ with $f_{ik} = f_{jk} \circ f_{ij}$ $(i \leq j \leq k)$ and $f_{ii} = \mathrm{Id}_{V_i}$. Such a diagram is also called a *persistence module*.

**Definition 9.** The collection of intervals $B = B(H_*(K_\bullet))$,

$$B = \{[i, \infty) \mid r_i = 0,\ i \notin \operatorname{pivots} R\} \cup \{[i, j) \mid i = \operatorname{pivot} r_j\}$$

is the *persistence barcode* of the filtration $K_\bullet$.

**Corollary 10.** *Let $\mathbb{F}$ be a field. For any interval $I = [i, j) \subset \mathbb{N}$ or $I = [i, \infty) \subset \mathbb{N}$, define a diagram $\mathbf{R} \to \mathbf{Vect}$,*

$$\mathbb{F}^I_k = \begin{cases} \mathbb{F} & \text{if } k \in I, \\ 0 & \text{otherwise}, \end{cases}$$

*with all maps $\mathbb{F}^I_k \to \mathbb{F}^I_l$ $(k \leq l)$ having maximal rank. Then*

$$H_*(K_\bullet) \cong \bigoplus_{I \in B(H_*(K_\bullet))} \mathbb{F}^I_\bullet.$$

# 3 Reindexing

While we have seen examples of general filtrations indexed over $\mathbb{R}$, our algorithm requires a simplexwise filtration indexed over $[n] = \{1, \ldots, n\}$. We will now describe a straightforward way of applying this algorithm to the more general setting.

There are two separate issues: transforming the indexing from $\mathbb{R}$ to $[n]$, and transforming a general filtration to a simplex-wise one.

**Definition 11.** We say that a diagram $G : \mathbb{R} \to C$ (in any category $C$) is *essentially finite* if there exists a diagram $F : [n] = \{1, \ldots, n\} \to C$ and a monotonic reindexing map $r : \mathbb{R} \to [n]$ such that $G = F \circ r$.

This way we can interpret $G$ as a diagram $H$ indexed over $[n]$, letting $H_i = G_t$ for any $t \in T_i$. We can also extend the index category to $\mathbb{N}$ by setting $H_m = H_n$ for $m > n$, connected by the identity. Conversely, given the **n**-indexed diagram together with the interval partition, we can recover the **R**-indexed diagram.

**Definition 12.** We say that a simplexwise filtration $L_\bullet : \mathbf{N} \to \mathbf{Simp}$ *refines* a filtration $K_\bullet : \mathbf{N} \to \mathbf{Simp}$ if there is a map $m : \mathbb{N} \to \mathbb{N}$ such that $K_i = L_{m(i)}$ for all $i \in \mathbb{N}$.

Given a filtration $K_\bullet : \mathbf{N} \to \mathbf{Simp}$ of a simplicial complex $K$, we can construct a refinement $L_\bullet$ by extending the partial order induced by the filtration on the simplices of $K$ into a total order, and filter $K$ according to this total order.

Using these two definitions, we can use the matrix reduction algorithm to compute the persistence barcode of any essentially finite filtration.

**Definition 13.** A simplicial filtration $G : \mathbf{R} \to \mathbf{Simp}$ is *finite* if it is essentially finite and the simplicial complex at any index is finite.

A persistence module $M : \mathbf{R} \to \mathbf{Vect}$ has *finite type* if it is essentially finite and the vector space at any index is finite-dimensional.

Let $(F_t)_{t\in\mathbb{R}}$ be a $\mathbb{R}$-indexed filtration of a finite simplicial complex $K$. Then there exists a simplexwise filtration $(K_i)_{1\leq i\leq n}$ and a monotonic reindexing map $r : \mathbb{R} \to [n] = \{1, \dots, n\}$ such that $F_t = K_{r(t)}$: extend the partial order on $K$ induced by $F_t$ and the partial order $\subseteq$ on $K$ (face order) to a single total order on $K$, which defines the simplexwise filtration $(K_i)_i$. Now each complex $F_t$ appears as some $K_i$ in the simplexwise filtration, and this defines the map $r : t \mapsto i$.

**Corollary 14.** *The persistence barcode of the simplexwise filtration $K_\bullet$ determines the persistence barcode of the filtration $F_\bullet$:*

$$H_*(F_\bullet) = \bigoplus_{\substack{I\in B(H_*(K_\bullet))\\ r^{-1}(I)\neq\emptyset}} \mathbb{F}_\bullet^{r^{-1}(I)}$$

Note that the barcode of $F_\bullet$ is a *multiset* of intervals in $\mathbb{R}$ (either of the form $[a, b)$ or of the form $[a, \infty)$. Each interval corresponds to homology in a certain dimension.

Note that the collection $\{U_I : I \in B(H_*(L_\bullet))\}$ may contain empty intervals and multiple copies of the same interval. We may define the barcode of a general finite-type persistence module as a multiset of intervals. These multisets can be formalized either as a function $\mathcal{I}_\mathbb{R} \to \mathbb{N}$ determining the multiplicity of each interval, or as a set of pairs $(I, k) \in \mathcal{I}_\mathbb{R} \times \mathbb{N}$, so that multiple instances of the same interval can be distinguished by different indices $k$.