Towards a verifiable topology of data

Laureano Lambán (U. La Rioja)
Francisco J. Martín-Mateos (U. Sevilla)
Julio Rubio (U. La Rioja)
José L. Ruiz-Reina (U. Sevilla)

Schedule

- Mathematical background.
- Persistent homology over
  - a field.
  - an elementary divisor ring.
- Formalisation in ACL2.

Mathematical backgroud

Simplicial Complex

V ordered set

Simplicial Complex (over V)

- $K \subseteq P_{fin}(V)$, K closed under the subset relation
- $S_n(K)$, "simplices whose cardinality is n+1"
- $\sigma^i_n : S_n(K) \rightarrow S_{n-1}(K)$,

  $\sigma^i_n(v_0,..,v_n)=(v_0,..,v_{i-1},v_{i+1},..v_n)$

Mathematical backgroud

Chain Complex

R ring

Chain Complex (over R)

$\mathbf{C}_* = \{(C_n, d_n) \ / \ n \epsilon Z\}$

- C_n , R-module
- d_n: $C_n \rightarrow C_{n-1}$, morphism of R-modules, $\forall n$
- $d_n \, d_{n+1} = 0$, $\forall n$

Chain Complex associated with a simplicial complex K

K simplicial complex.

$C_*(K)$ is given by:

- $C_n(K)$, free R-module over $S_n(K)$
- $d_n = \sum_{i=0}^{n}(-1)^i \sigma^i_n$

Homology

K simplicial complex.

The **n-th homology group of K**

$$H_n(K) = \frac{Z_n}{B_n}$$

($d^2 = 0$ implies $\mathrm{Im}\ d_{n+1} = B_n \leq Z_n = \mathrm{Ker}\ d_n$)

Filtration

- Filtration of a simplical complex K

$$K^1 \subseteq K^2 \subseteq \dots \subseteq K^d = K$$

(K is a filtered simplicial complex)

- Filtration of chain complexes

$$C^0_* \xrightarrow{f_0} C^1_* \xrightarrow{f_1} C_*^2 \xrightarrow{f_2} \dots$$

(f_i is a morphism of chain complexes)

Mathematical backgroud

Induced filtration

A filtration of simplicial complexes

$$K^0 \to K^1 \to K^2 \ldots \to K^d = K$$

$\downarrow$ induces

$$C_*(K^0) \to C_*(K^1) \to \ldots \to C_*(K^d)$$

a filtration of chain complexes

(arrows are inclusion maps)

Persistent homology

A filtration of simplicial complexes

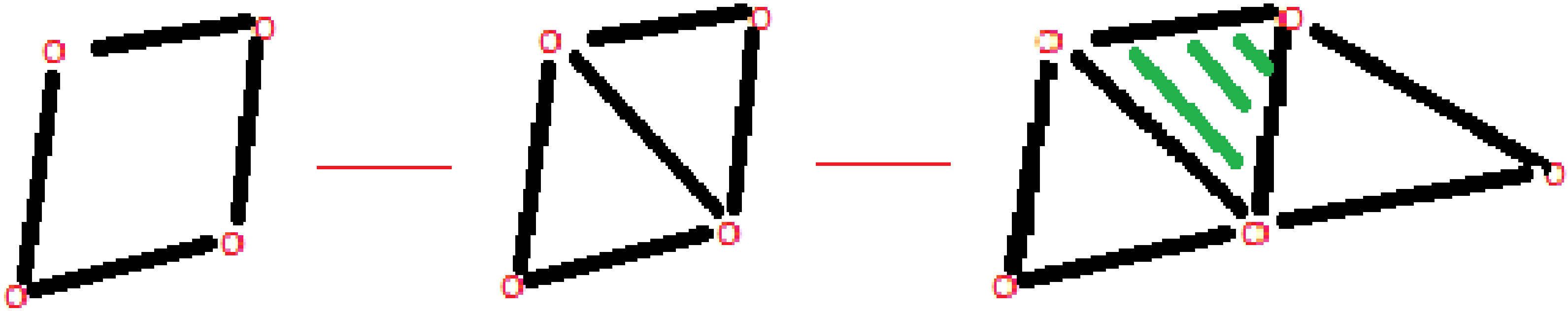
$$K^0 \to K^1 \to K^2 \ldots \to K^d = K$$

The **p-persistent n-dimensional homology group of K^j** is given by

$$H_n^{j,p} = Z_n^j / B_n^p \cap Z_n^j \qquad (j \leq p)$$

$$(Z_n^j \leq Z_n^p \leq C_n(K^p) \text{ and } B_n^p \leq C_n(K^p))$$

An example



K^3 K^4 K^5

$H_1^{4,4} = Z+Z$ $H_1^{4,5} = Z$ $H_1^{5,5} = Z+Z$

Mathematical basis

- Artin-Rees Theorem implies

  "F-modules of persistence ≡ F[x]-modules"

  (finite type)

- F is a field, F[x] is a PID

- To computing the persistent homology of a filtration of F-modules is equivalent to computing the homology of its (Artin-Rees) associated F[x]-module.

Algorithm

Variant of the gaussian elimination algorithm

- Polynomial time
- Persistence intervals
- Barcodes

Computing persistent homology (over an ED-ring R)

Mathematical basis

Two essential bricks:

- Echelon forms for matrices
  (effective Bezout domains)

- Smith normal form
  (elementary divisor rings)

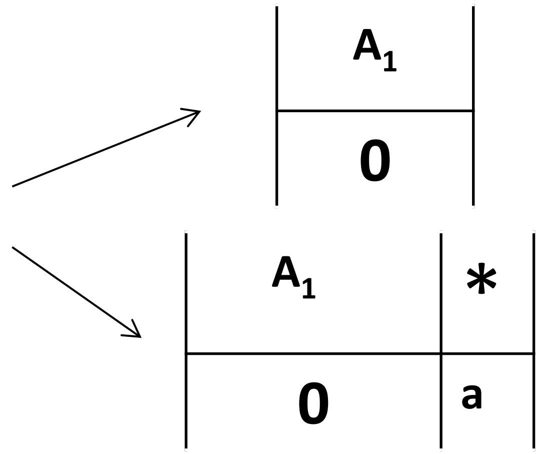
Echelon form

An echelon form

($A \in M(nxm,R)$)

$n=1 \longrightarrow A \longrightarrow (0\ 0\ 0 \ldots a), \quad a \in R$

$n>1 \longrightarrow A$

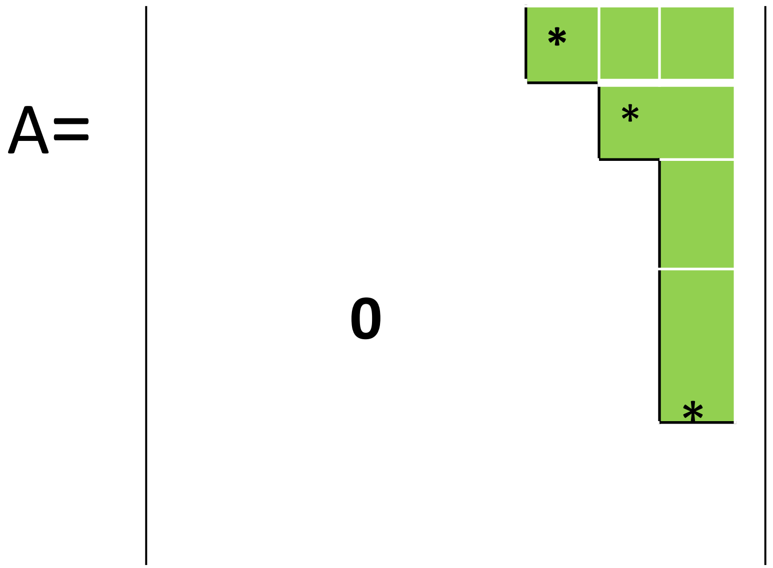


($a \neq 0$) and echelonForm(A_1)

Computing persistent homology (over an ED-ring R)

EchelonForm

echelonForm(A)



(the width of the steps is 1)

Existence of echelon form

R is an effecctive Bezout domain.

Th1: Existence of echelon form ($A \in M(nxm,R)$)

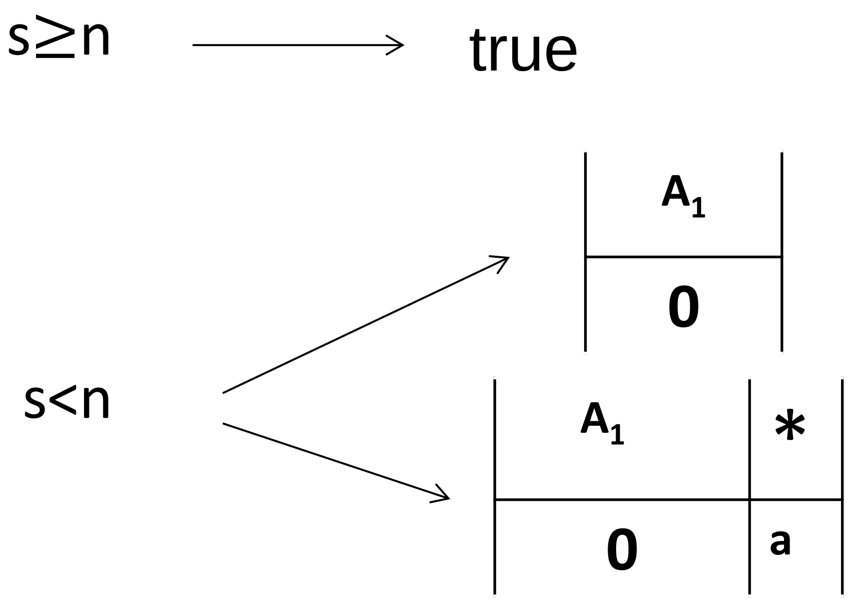
"There exists P invertible s.t. echelonForm(AP)"

(P is a sequence of Bezout and permutation "elementary operations")

Computing persistent homology (over an ED-ring R)

Generalized echelon form

echelonForm(A,s) $A \in M(n\mathrm{x}m,R)$

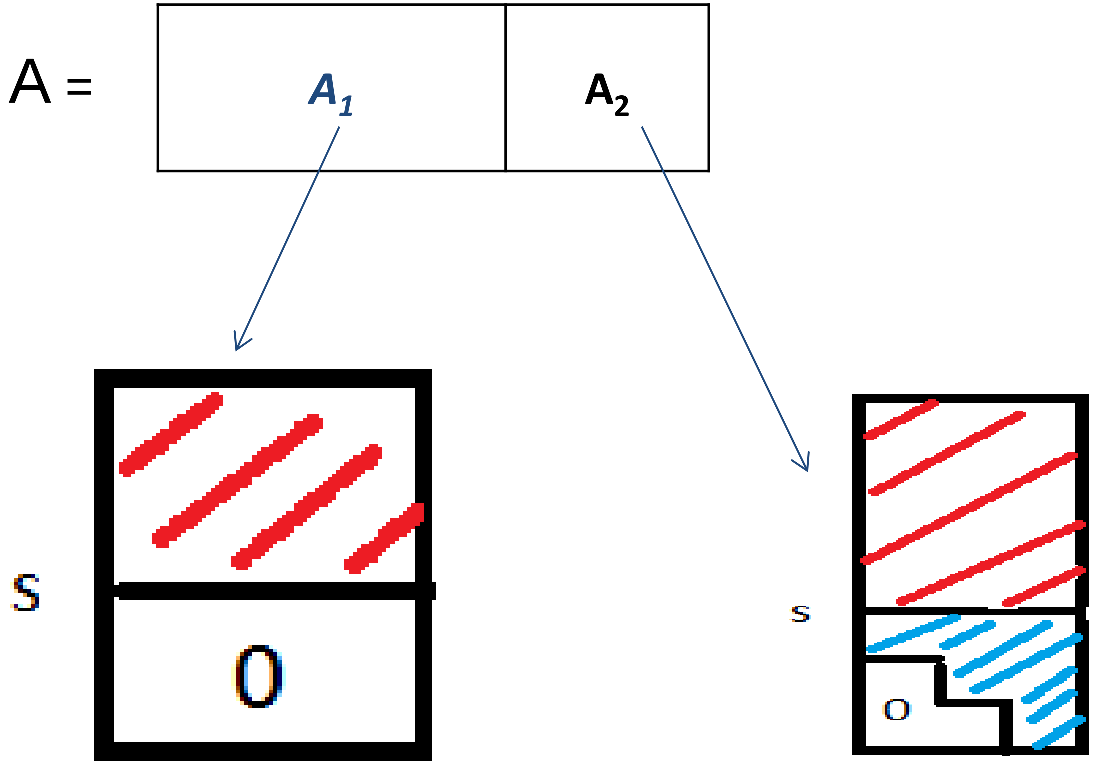


$(a \neq 0)$ and echelonForm(A_1,s)

Computing persistent homology (over an ED-ring R)

Generalized echelon form

echelonForm (A,*s*)

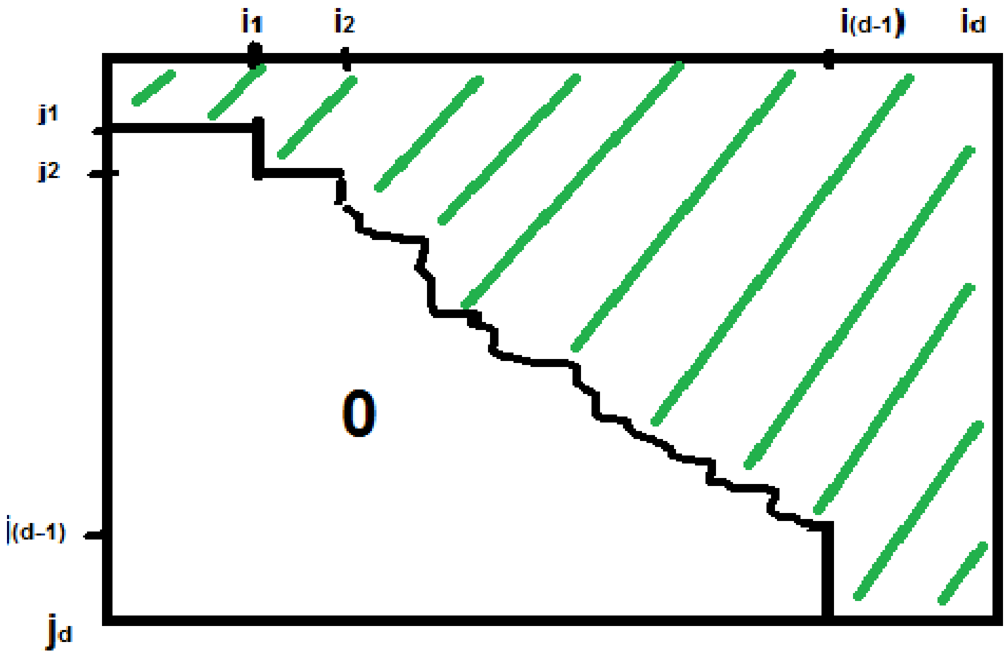


Computing persistent homology (over an ED-ring R)

Graded matrices

Graded matrix

$A \in M(nxm, R)$



$0 \leq j_1 \leq j_2 \leq \ldots.. \leq j_d = n$

$0 \leq i_1 \leq i_2 \leq \ldots.. \leq i_d = m$

Graded matrices

A filtration of simplicial complexes

$$K^0 \to K^1 \to K^2 \ldots \to K^d = K$$

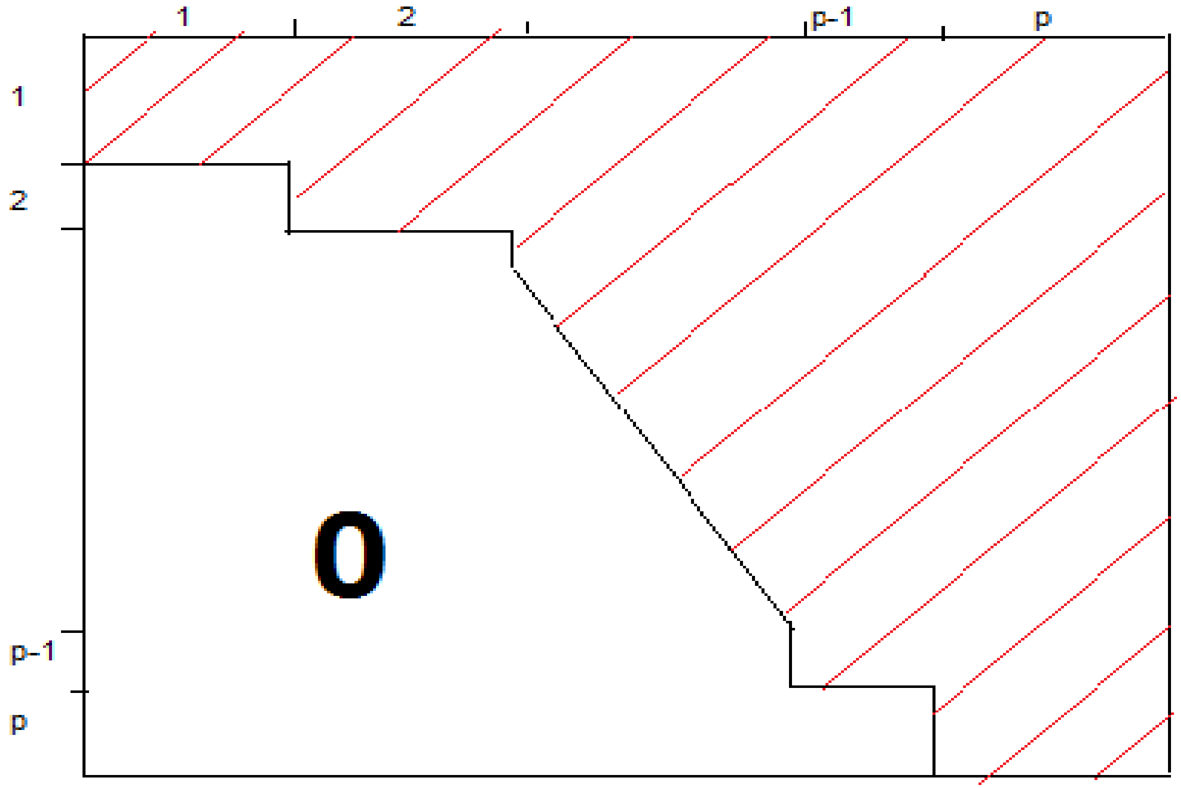
The standard matrix representation of the differential map

$$d_n = \{d_n^i : C_n(K^i) \to C_{n-1}(K^i)\}$$

is a graded matrix.

Graded matrices

A =



i=1,2,…,p represent the filtration index

Graded echelon form

R is an effecctive Bezout domain.

Th2: “Existence of graded echelon form”

Input: (A,s,t)

- A ∈ M(nxm,R), graded matrix
- $0 \leq s \leq t \leq d$

Graded echelon form

R is an effecctive Bezout domain.

Th2: “Existence of graded echelon form”

Input: (A, s, t)

Output: $(P_s, P_{s+1},..,P_t)$

- P_i are invertible matrices
- echelonForm(E_i , s), where $E_i = A_i P_i$, and…

Computing persistent homology (over an ED-ring R)

Graded echelon form

$$E_i = \left[\begin{array}{c|c} E_i^1 & E_i^2 \end{array}\right]$$

$$E_i = A_i P_i$$

$$E_i^1 \in M(\_x\ b_i , R)$$

$$P_i = \left[\begin{array}{c|c} P_{i-1}^{1..b(i-1)} & \\ \hline 0 & * \end{array}\right]$$

Smith normal form

R is an elementary divisor ring.

Th3: "Existence of Smith normal form"

Input: A $(A \in M(nxm,R))$

Output: $((d_1,\ldots,d_r), P, R)$

- P, Q invertible matrices
- $d_i | d_{i+1}$ (i=1..r-1)
- PAQ=DiagonalMatrix $(d_1,\ldots,d_r,0,..,0)$

Algorithm

Algorithm

Input:

- $K^0 \to K^1 \to K^2 \ldots \to K^d = K$
  (a filtered complex K)
- $n \geq 0$
- r, t such that $0 \leq r \leq t \leq d$

Output:

- $H_n^{r,j}(K)$ $r \leq j \leq t$

Computing persistent homology (over an ED-ring R).

Algorithm (Step 1)

Let A be the standard matrix representation of

$$d_n=\{d_n^{\,i}: C_n(K^i) \to C_{n-1}(K^i) \mid i=1..d\}$$

Step 1: gradedEchelonForm (A, 0, d)

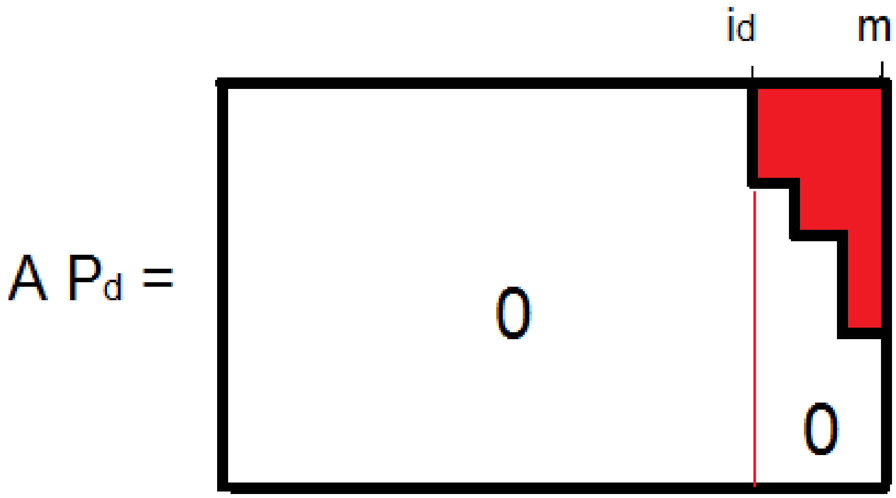
$$\downarrow$$

invertible matrices $(P_1,\ldots.P_d)$

Computing persistent homology (over an ED-ring R).

Algorithm (Step 1)

invertible matrices $(P_1,..,P_d)$ such that



Columns of P's matrices provide us a family of compatible basis $B_1 \leq B_2 \leq .... \leq B_d$

of the kernels $Z_1 \leq Z_2 \leq .... \leq Z_d$

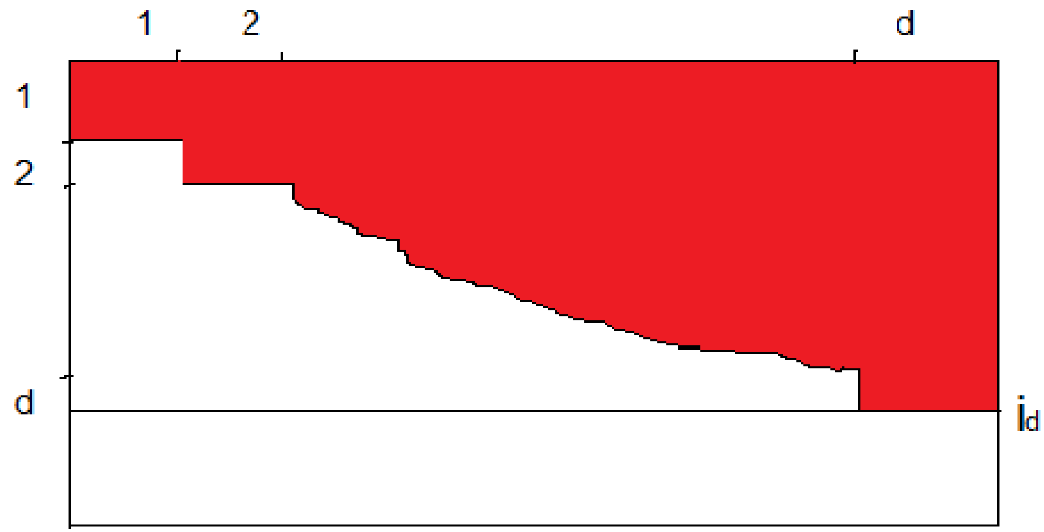
Computing persistent homology (over an ED-ring R).

Algorithm (Step 2)

Let B be the standard matrix representation of d_{n+1}

$$d_{n+1}=\{d_{n+1}^{i}: C_{n+1}(K^{i}) \to C_{n}(K^{i}) \mid i=1..d\}$$

Step 2: $P_d^{-1} B$ =



($\mathrm{Im}\ d_{n+1}^{i} \leq \mathrm{Ker}\ d_{n}^{i} = Z_i \Rightarrow$ last rows of $P_d^{-1} B$ are zero)

Computing persistent homology (over an ED-ring R).

Algorithm (Step 3)

Step 3: gradedEchelonForm $(P_d^{-1} B, r, t)$

$$\downarrow$$

$(Q_{r+1}, .., Q_t)$

echelonForm$((P_d^{-1} B)_j Q_j, r)$ $(r \leq j \leq t)$

Computing persistent homology (over an ED-ring R).

Algorithm (Step 4)

Let $E_j = (P_d^{-1} B)_j Q_j = [E_j^1 \mid E_j^2]$

Step 4: SmithForm (E_j^1)

$\downarrow$

$H_n^{r,j}(K)$

An Smith computation for each filtration index j!

Formalisation in ACL2

- Our goal: implement and formally verify this algorithm in the ACL2 Theorem Prover

- This means to implement and verify

  Echelon form

  Graded echelon form

  Smith normal form

Formalisation in ACL2

- One of our main concerns is efficiency

  Advantage of using ACL2: we verify Common Lisp code.

- But ACL2 is an applicative subset of CL. In principle, we have not CL arrays, only lists.

- Fortunately, we can use stobjs (single-threaded objects), which alllows destructive updates and constant-time accesses, without losing the applicative semantic.

What we have now:

- An array-based version of echelon form algorithm ("efficient" but unverified).
- A list-based version of echelon form and Smith form algorithms (still unverified).
- Infrastructure proving operational equivalence between both approaches.

Our goal: reasoning using the second approach and executing using the first one.