Introduction to Computational Topology

Yaroslav Bazaikin

Sobolev Institute of Mathematics, Novosibirsk

Online Summer School on Geometry and Topology, II
08.07.2021

Let for every integer p we have $n_p = rank(C_p)$ be the number of simplices of dimension p. Fix some order of simplices σ_j, $j = 1, \dots, n_p$. Represent homomorphism $\partial : C_p \to C_{p-1}$ by matrix with respect to this order: $\partial(\sigma_i^p) = \sum_j a_j^i \sigma_j^{p-1}$, $\partial = (a_j^i)_{j=1,\dots,n_{p-1}}^{i=1,\dots,n_p}$.

If $c = \sum_i a_i \sigma_i^p$ then

$$\partial_p(c) = \begin{pmatrix} a_1^1 & a_1^2 & \dots & a_1^{n_p} \\ a_2^1 & a_2^2 & \dots & a_2^{n_p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n_{p-1}}^1 & a_{n_{p-1}}^2 & \dots & a_{n_{p-1}}^{n_p} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_{n_p} \end{pmatrix}$$

(every value $a_j^i = 0$ or 1). Simple linear algebra implies

$$rank(B_{p-1}) = rank(\partial_p),$$

$$rank(Z_p) = n_p - rank(B_{p-1}) = n_p - rank(\partial_p).$$

Up to this moment we use natural basis of C_p formed by simplices. Let consider change of basis in C_p and C_{p-1}, independently. We will perform change of basis as consequence of two elementary operations:

1) exchanging of two basis vectors:

$$e_1, \dots, e_i, \dots, e_j, \dots, e_n \to e_1, \dots, e_j, \dots, e_i, \dots, e_n$$

2) add some basis vector to other:

$$e_1, \dots, e_i, \dots, e_j, \dots, e_n \to e_1, \dots, e_i, \dots, e_i + e_j, \dots, e_n$$

Remark that $rank(\partial_p)$ (and ranks of Z_p, B_p) does not depend of change of basis.

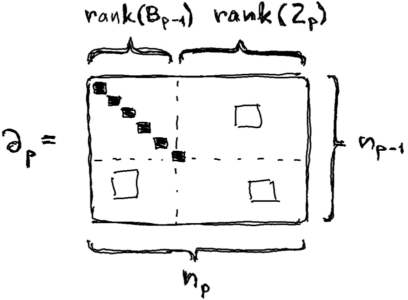
In what extent we can simplify matrix ∂_p by combination of such elementary operations?

Elementary operations for matrix:

1) C_p: exchanging of columns; C_{p-1}: exchanging of rows.

2) C_p: adding one column to other; C_{p-1}: adding one row to other.

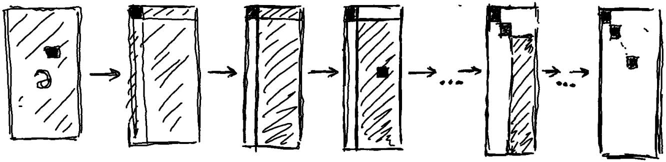
Modification of classical Gauss method allows to reduce matrix ∂_p by elementary operations to the following normal form:



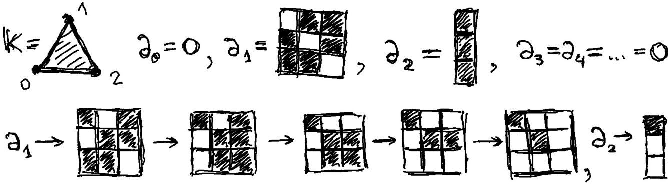
Start Reduce(1) for $N_p = \partial_p$. Finish with N_p in normal form.

```
void Reduce(x)
  if there exists k ≥ x, l ≥ x with N_p[k, l] = 1 then
    exchange rows x and k; exchange columns x and l;
    for i = x + 1 to n_{p-1} do
      if N_p[i, x] = 1 then
        add row x to row i;
      endif
    endfor
    for j = x + 1 to n_p do
      if N_p[x, j] = 1 then
        add column x to column j;
      endif
    endfor
    Reduce(x+1);
  endif
```

Illustration of work of the above algorithm:



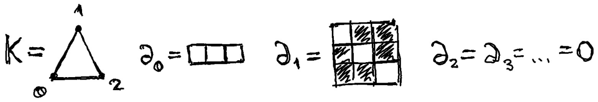
Example 1. $|K| = D^2$, K consists of one 2-dimensional simplex.



Then $rank(B_0) = 2$, $rank(B_1) = 1$, $rank(B_2) = 0$, $rank(Z_0) = 3$, $rank(Z_1) = 1$, $rank(Z_2) = 0$ and

$$b_0(K) = 1, H_0(K) = \mathbb{Z}_2, b_1(K) = 0, H_1(K) = 0,$$

Example 2. $|K| = S^1$, K consists of three segments:



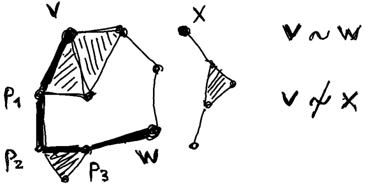
In this case $rank(B_0) = 2$, $rank(B_1) = 0$, $rank(B_2) = 0$, $rank(Z_0) = 3$, $rank(Z_1) = 1$, $rank(Z_2) = 0$ and

$$b_0(K) = 1, H_0(K) = \mathbb{Z}_2, b_1(K) = 1, H_1(K) = \mathbb{Z}_2.$$

By the way, Examples 1 and 2 prove Brouwer Foxed Point Theorem (we need to consider H_1 as topological invariant).

Look more precisely at homology group H_0 and zero Betti number b_0. They have very clear geometric interpretation.

Let K be simplicial space. We say that two vertices v and w from K are connected if there exists collection of 1-dimensional simplices $[vp_1], [p_1p_2], \dots, [p_nw]$ (piecewise linear path connecting vertexes v and w). It is easy to see that property to be connected is the equivalence relation on the set of all 0-dimensional simplices (vertexes) in K.



Equivalence classes are *connected components* of K. The number of connected components evidently is a topological invariant.

Theorem *The number of connected components of simplicial complex K is equals to zero Betti number $b_0(K)$.*

Proof. The main observation is: if v and w are equivalent then

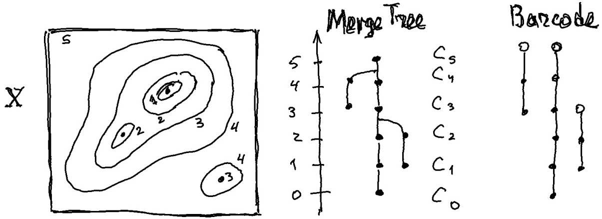
$$\partial_1([vp_1] + [p_1 p_2] + \ldots + [p_k w]) = v + w$$

Therefore $v + B_0 = w + B_0$ in Z_0 if and only if v is connected to w.

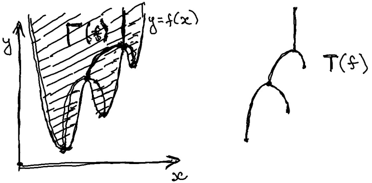
Persistent Homology

Let X be topological space and $f : X \to \mathbb{R}$ be continuous function. Excursion set of function f is the subset $X_a = \{p \in X | f(p) \leq a\}$ for some a. Define C_a be the set of connected components of X_a. If $a \leq b$ then $X_a \subset X_b$ and we obtain ta map

$$f_a^b : C_a \to C_b.$$



More formally, in the space $\Gamma(f) = \{(p, a) \in X \times \mathbb{R} | f(p) \leq a\}$ consider equivalence relation: we say that $(p, a) \sim (q, b)$ if $a = b$ and p, q belongs to the same component of C_a. The topological space $T(f) = \Gamma(f)/\sim$ is a tree and is called *merge tree* of f



To produce barcode $B(f)$ we need to cutoff branches of $T(f)$ in the merge points according to rule:

Elder Rule. *The older of merged branches continues, younger ones end.*

Let K be simplicial complex and $f : K \to \mathbb{R}$ be some function. We say that f is monotonic if for each face τ of every simplex σ from K inequality $f(\tau) \leq f(\sigma)$ holds. If f is monotonic then $K(a) = f^{-1}(-\infty, a]$ is again a simplicial complex (subcomplex in K).

Let m be the number of simplexes in K and let $a_1 < a_2 < \ldots < a_n$ are all the function values. Then denoting $a_0 = -\infty$ and $K_i = K(a_i)$ we obtain increasing sequence of simplicial cmplexes:

$$\emptyset = K_0 \subset K_1 \subset \ldots \subset K_n = K.$$

This sequence is called *filtration*. Filtration describes the construction of K in n steps by adding several simplexes at a time. The question arises: how does the topological complexity increase by these steps?

We have corresponding chain of homomorphisms of homology groups:

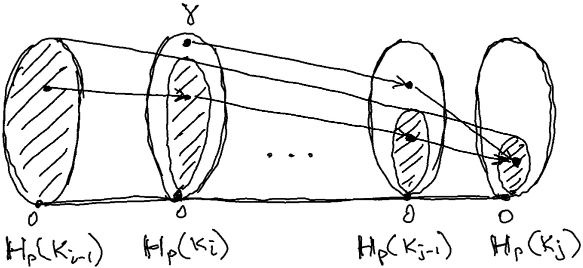
$$0 = H_p(K_0) \to H_p(K_1) \to \ldots \to H_p(K_n) = H_p(K).$$

What happens on every step from $H_p(K_{i-1}$ to $H_p(K_i)$? There are two possibility: some new classes born; and some old classes die or merge with each other. Let $f_p^{i,j} : H_p(K_i) \to H_p(K_j)$ be some fragment of above chain, $i \leq j$.

The *p-th persistent homology groups* are $H_p^{i,j} = im(f_p^{i,j})$ for $1 \leq i \leq j \leq n$. The corresponding *p-th persistent Betti numbers* are the ranks of these groups, $b_p^{i,j} = rank(H_p^{i,j})$.

It is obvious that $H_p^{i,i} = H_p(K_i)$ and $b_p^{i,i} = b_p(K_i)$.

Formulate Elder Rule in this situation. Letting $\gamma \in H_p(K_i)$, we say it is born at K_i if $\gamma \notin H_p^{i-1,i}$. Furthermore, if γ is born at K_i then it dies entering K_j if it merges with an older class as we go from K_{j-1} to K_j, that is, $f_p^{j,j-1}(\gamma) \notin H_p^{i-1,j-1}$ but $f_p^{i,j}(\gamma) \in H_p^{i-1,j}$.



Consider number $\mu_p^{i,j}$ of p-dimensional classes which born at K_i dying entering K_j:

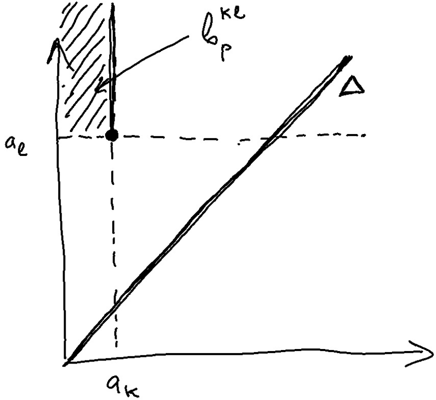
$$\mu_p^{i,j} = (b_p^{i,j-1} - b_p^{i,j}) - (b_p^{i-1,j-1} - b_p^{i-1,j}), i < j.$$

The *p-th persistence diagram* of the filtration is the subset of extended real plane $\bar{\mathbb{R}}^2$ consisting of points with multiplicities:

$$Dgm_p(f) = \{\mu_p^{i,j} \cdot (a_i, a_j) | 1 \leq i < j \leq n\} \cup \{\infty \cdot (a, a) | a \in \mathbb{R}\}$$

Theorem. *For any pair of indices* $0 \leq k \leq l \leq n$

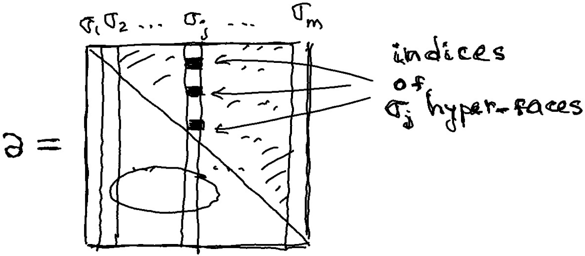
$$b_p^{k,l} = \sum_{i \leq k} \sum_{j > l} \mu_p^{i,j}.$$



Let f be monotonic simplicial complex. Order all simplexes: $K = \{\sigma_1, \ldots, \sigma_m\}$. We say that ordering is compatible with f, if $f(\sigma_i) < f(\sigma_j)$ implies $i < j$. It is clear that compatible ordering exists: we can order firstly simplexes with f-value a_1, then a_2 and so on ...

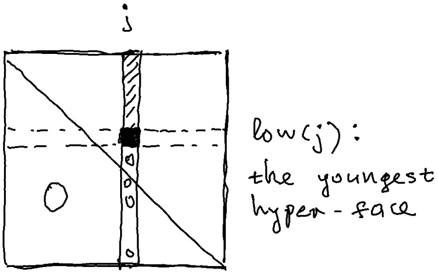
Compatible ordering has following property: for every k, $\{\sigma_1, \sigma_2, \ldots, \sigma_k\}$ is a simplicial subcomplex in K. We use compatible ordering to construct boundary matrix which store all information about all boundary operators in all dimensions:

$$\partial[i,j] = \begin{cases} 1, & \text{if } \sigma_i \text{ is a co-dimension one face of } \sigma_j; \\ 0, & \text{otherwise.} \end{cases}$$



As far as ordering is compatible with monotonic function, matrix ∂ is upper triangular.

Let $low(j)$ be the index of the lowest non-zero element in column with number j (if entire column is zero then $low(j)$ is not defined). We say that 0-1 matrix R is reduced if $low(j_1) \neq low(j_2)$ for every non-zero columns $j_1 \neq j_2$.

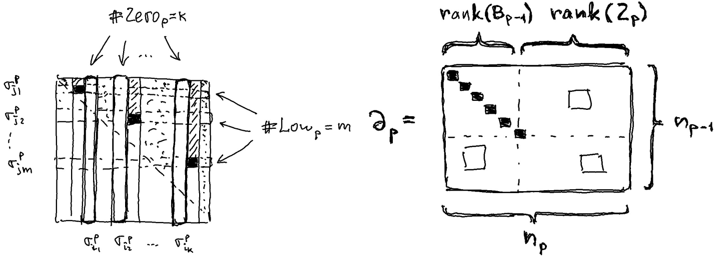


Algorithm (we use adding operation only because we can not independently exchange columns and raws):

R $= \partial$
for $j = 1$ **to** m **do**
  while there exists $j_0 < j$ with $low(j_0) = low(j)$ **do**
    add column j_0 to column j;
endfor

It is clear that R again is upper triangular.

$\#Zero_p(R)$ is the number of p-dimensional zero columns;
$\#Low_p(R)$ is the number of p-dimensional rows containing $low(j)$.



Comparing matrix R with normal form of boundary operator matrix we can conclude:

$$\#Zero_p(R) = rank(Z_p), \#Low_p(R) = rank(B_p),$$

$$b_p = \#Zero_p(R) - \#Low_p(R).$$

Relation $i = low(j)$ is a pairing between $(p-1)$ simplex σ_i and p-simplex σ_j.

Lemma. *Above defined pairing does not depend of way or reducing matrix ∂ to matrix R*

Let column j of R at some stage of algorithm has its final form. There are two possibilities:

1) Column j of R is zero. We can think that we add simplex σ_j to simplicial complex consisting of $\sigma_1, \ldots, \sigma_{j-1}$ and this addition generates new cycle (with zero boundary). We call σ_j *positive* simplex.

2) Column j of R is non-zero. Simplexes in column j form boundary of simplex σ_j and therefore represent cycle. This cycle becomes trivial in homology group (dies) after adding σ_j. We call σ_j *negative* simplex.

So, cycle represented by simplexes in column j dies at step number j (in opposite case we would find combination of columns with indices less than j with zero sum - a contradiction with the fact that R is reduced up to number j).

When this cycle was born? It born in moment $i = low(j)$! In opposite case we find column j_0 with nonzero row $i = low(j)$ - a contradiction to algorithm.

Theorem. *Point* (a_i, a_j) *$(i, j > 0)$ is a point with positive multiplicity in* $Dgm_p(f)$ *if and only if* $i = low(j)$ *and* σ_i *is a simplex of dimension* p*. Point* (a_i, ∞) *is a point with positive multiplicity in* $Dgm_p(f)$ *if and only if column* i *is zero but row* i *does not contain the lowest element.*

The latter case corresponds to cycles which born at some moment a_i but never die in K.

Thank you for attention!