Persistent Homology

David Gu

Yau Mathematics Science Center Tsinghua University Computer Science Department Stony Brook University

gu@cs.stonybrook.edu

August 16, 2023

Persistent Homology

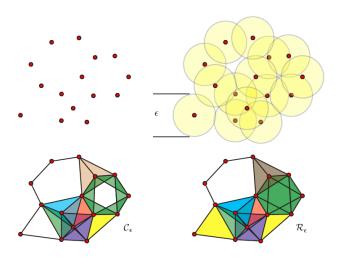


Figure: Cech complex.

Definition (Cech Complex)

Given a set of points $\{x_{\alpha}\}$ in Euclidean space \mathbb{R}^n , the Cech complex (also known as the nerve), $\mathcal{C}_{\varepsilon}$, is the abstract simplicial complex where a set of k+1 vertices spans a k-simplex whenever the k+1 corresponding closed $\varepsilon/2$ -ball neighborhoods have nonempty intersection.

Definition (Vietoris-Rips Complex)

Given a set of points $\{x_{\alpha}\}$ in Euclidean space \mathbb{R}^n , the Vietoris-Rips complex, $\mathcal{R}_{\varepsilon}$, is the abstract simplicial complex where a set S of k+1 vertices spans a k-simplex whenever the distance between any pair of points in S is at most ε .

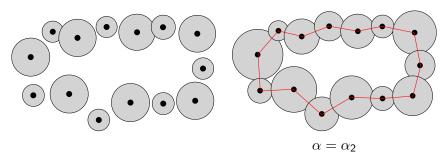


Figure: Cech complex.

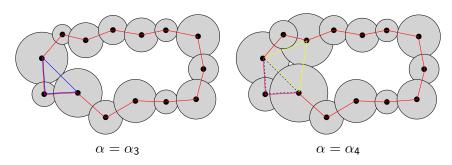


Figure: Cech complex.

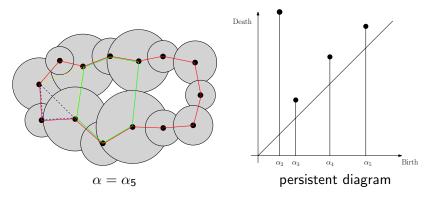


Figure: Cech complex.

Filtration

Definition (filtration)

A filtration of a simplicial complex $\ensuremath{\mathbb{K}}$ is a nested sequence of complexes,

$$\emptyset = \mathbb{K}_{-1} \subset \mathbb{K}_0 \subset \mathbb{K}_1 \subset \cdots \subset \mathbb{K}_n = \mathbb{K}.$$

Example

Suppose ${\mathbb K}$ is a simplicial complex, we sort all the simplices in a sequence

$$\sigma_1^0, \sigma_2^0, \cdots, \sigma_{n_0}^0, \sigma_1^1, \sigma_2^1, \cdots, \sigma_{n_1}^1, \sigma_1^2, \sigma_2^2, \cdots, \sigma_{n_2}^2.$$

where σ_i^k is the *i*-th *k*-simplex in \mathbb{K} . Then we relabel all the simplices as

$$\sigma^0, \sigma^1, \sigma^2, \cdots,$$

We define \mathbb{K}_i as the union of $\sigma^0, \sigma^1, \dots, \sigma^i$.



Homology

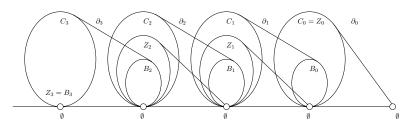


Figure: Chain, cycle, boundary groups and their images under the boundary operators.

$$H_k(\mathbb{K}, \mathbb{Z}_2) = \frac{\mathsf{Ker} \partial_k}{\mathsf{Img} \partial_{k+1}} = \frac{Z_k}{B_k}.$$

Persistent Homology

The inclusion map $f: \mathbb{K}_{i-1} \hookrightarrow \mathbb{K}_i$ defined by f(x) = x induces a homomorphism $f_*: H_p(\mathbb{K}_{i-1}) \to H_p(\mathbb{K}_i)$. The nested sequence of complexes corresponds to a sequence of homology groups connected by the induced maps,

$$0 = H_p(\mathbb{K}_{-1}) \to H_p(\mathbb{K}_0) \to \cdots \to H_p(\mathbb{K}_n) = H_p(\mathbb{K})$$

Persistent homology studies how the homology groups change over the filtration.

Generator and Killer

Definition (positive simplex)

Given a filtration of \mathbb{K} , suppose $\mathbb{K}_i - \mathbb{K}_{i-1} = \sigma_i$, where σ_i is a (k+1)-simplex. We call σ_i is positive if it belongs to a (k+1)-cycle in \mathbb{K}_i and negative otherwise.

A positive simplex is also called a generator, a negative simplex a killer.

Generator and Killer

Definition (Betti Number)

Given a complex K, the i-th Betti number β_i is the rank of $H_i(K)$,

$$\beta_i = \mathsf{Rank} H_i(K, \mathbb{Z}_2)$$

Suppose the number of positive k-simplexes is pos_k , and the number of negative k-simplexes is neg_k , then

$$\beta_k = \mathsf{pos}_k - \mathsf{neg}_{k+1}$$

Persistent Homology

Definition (Persistent Homology)

Define Z_k^I, B_k^I be the K-th cycle group and k-th boundary group respectively, of the I-complex K^I in a filtration. The p-persistent k-th homology group K^I is

$$H_k^{l,p} := \frac{Z_k^l}{B_k^{l+p} \cap Z_k^l}.$$

The *p*-persistent *k*-th Betti number $\beta_k^{I,p}$ of K^I is the rank of $H_k^{I,p}$.

Lemma

Consider the homomorphism $\eta_k^{l,p}: H_k^l \to H_k^{l+p}$, then

$$\mathit{img}\ \eta_k^{\mathit{I},\mathit{p}} \cong \mathit{H}_k^{\mathit{I},\mathit{p}}$$



Generator

Lemma

For each positive k-simplex σ^i , there exists a non-exact k-cycle c^i , c^i contains σ^i but no other positive k-simplices.

Proof.

Start with an arbitrary a k-cycle that contains σ^i and remove other positive k-simplices by adding their corresponding k-cycles. This method succeeds because each added cycle contains only one positive k-simplex by inductive assumption.

We use σ^i to represent c^i , and in turn the homologous class $[c^i] = c^i + B_k$.

$$\sigma^i \to c^i \to [c^i] = c^i + B_k. \quad \sigma^i \sim c^i$$

We add $[c^i]$ to the basis of $H_k(\mathbb{K}^i)$.



Generator

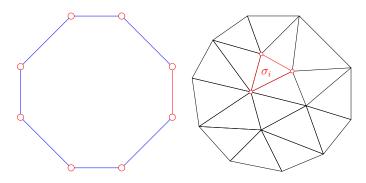


Figure: Generator, positive simplex.

Killer

For each negative (k+1)-simplex σ^j , its boundary $d=\partial_{k+1}\sigma^j$ is a k-cycle, and can be represented as the linear combination of the basis of $H_k(\mathbb{K}_{j-1})$,

$$[d] = \sum_{g} [c^g], \{c^g\} \text{ basis } H_k(\mathbb{K}_{j-1}),$$

each $[c^g]$ is represented by a positive k-simplex σ^g , g < j, that is not yet paired. The collection of positive non-paired k-simplices is denoted as $\Gamma = \Gamma(d)$,

$$\Gamma(d) := \left\{ \sigma^{\mathbf{g}} : [d] = \sum_{\mathbf{g}} [c^{\mathbf{g}}], \quad \sigma^{\mathbf{g}} \sim c^{\mathbf{g}} \right\}$$

Suppose the youngest positive simplex in $\Gamma(\partial_{k+1}\sigma^j)$ is σ^i , then we form the pair (σ^i,σ^j) , and remove $[c^i]$ from $H_k(\mathbb{K}_j)$. $[c^i]$ is created by σ^i and killed by σ^j , the persistence life of the k-cycle $[c^i]$ is j-i-1.

Example Filtration



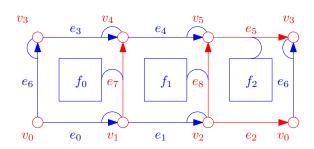


Figure: Generators and killers.

Filtration

$$v_0, v_1, v_2, v_3, v_4, v_5, e_0, e_1, e_2, e_3, e_4, e_5, f_0, f_1, f_2$$

Relabel them as

$$\sigma^{0}, \sigma^{1}, \sigma^{2}, \sigma^{3}, \sigma^{4}, \sigma^{5}, \sigma^{6}, \sigma^{7}, \sigma^{8}, \sigma^{9}, \sigma^{10}, \sigma^{11}, \sigma^{12}, \sigma^{13}, \sigma^{14}, \sigma^{15}$$

Example Generators

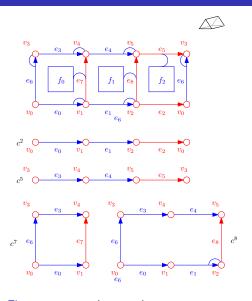


Figure: c_k contains a unique generator e_k .

Example Killers

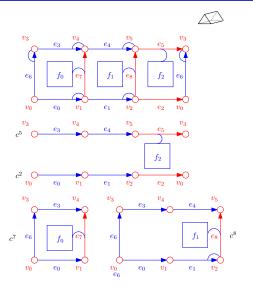


Figure: Killers.



Example Pairing



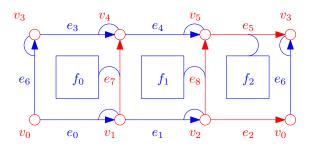


Figure: Generators and killers.

$$\partial_2 f_2 = \frac{e_2}{e_2} + \frac{e_5}{e_5} + e_6 + e_8 = (\frac{e_5}{e_5} + 2e_4 + 2e_3) + (\frac{e_2}{e_2} + 2e_1 + 2e_0) + e_6 + e_8$$

$$= (\frac{e_5}{e_5} + e_4 + e_3) + (\frac{e_2}{e_2} + e_1 + e_0) + \partial_2 (f_0 + f_1)$$

$$= c_5 + c_2 + \partial_2 (f_0 + f_1)$$

Key Lemma

Definition (Collision Free Cycle)

A collision free cycle is one where the youngest positive simplex has not been paired (killed).

Lemma (Collision)

Given a filtration, $\mathbb{K}_j - \mathbb{K}_{j-1} = \sigma^j$, σ^i is the youngest positive simplex in $\Gamma(\partial_{k+1}\sigma^j)$. Let e be a collision free k-cycle in \mathbb{K}_{j-1} homologous to $\partial_{k+1}\sigma^j$. Suppose the youngest positive simplex in e is σ^g , then

$$\sigma^i = \sigma^g$$
.

 $\max \Gamma(\partial_{k+1} \sigma^j) = \max(e) \quad \forall e \text{ collision free}, [e] = [\partial \sigma^j].$



Key Lemma

Proof.

Let f be the sum of the basis cycles, homologous to $d = \partial_{k+1} \sigma^j$. By definition, f's youngest positive simplex is σ^i , namely the youngest simplex in $\Gamma(\partial_{k+1} \sigma^j)$,

$$\sigma^{i} = \max \Gamma(\partial_{k+1}\sigma^{j}).$$

This implies that there are no cycles homologous to d in \mathbb{K}_{i-1} or earlier complexes. Let σ^g be the youngest positive simplex in e. [e] = [d], therefore $g \geq i$.

If g > i, then e = f + c, where c bounds in \mathbb{K}^{j-1} . $\sigma^g \notin f$, implies $\sigma^g \in c$, and as σ^g is the youngest in e, it is also the youngest in c.

Key Lemma

continued.

Since e is collision free, the cycle created by σ^g , denoted as c^g , is still a non-boundary cycle in \mathbb{K}_{j-1} . Hence c^g can't be c, and can't be homologous to c when c becomes a boundary. Namely, when c is killed, σ^g is not paired yet.

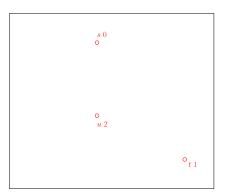
It follows that the negative (k+1)-simplex that kills c must pair a positive k-simplex in c, which is younger than σ^g , a contradiction.

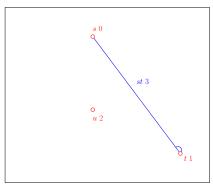
This lemma shows, when σ^j is added to \mathbb{K}_{j-1} , we need to find any collision free cycle e homologous to $\partial_{k+1}\sigma^j$, and pair σ^j with the youngest positive simplex of e.

Pair Algorithm

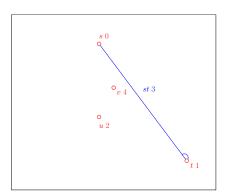
$Pair(\sigma)$

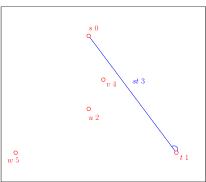
- **2** τ is the youngest positive (p-1)-simplex in c.
- **3** while τ is paired and c is not empty **do**
- find (τ, d) , d is the p-simplex paired with τ ;
- $c \leftarrow \partial_{p} d + c$
- Update τ to be the youngest positive (p-1)-simplex in c
- end while
- **3 if** *c* is not empty **then**
- σ is negative p-simplex and paired with τ
- else
- σ is a positive *p*-simplex
- endif

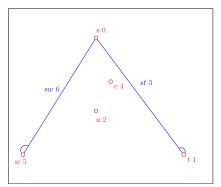


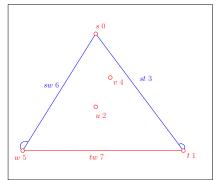


3. $\partial st = s + t$, (t_1, st_3)





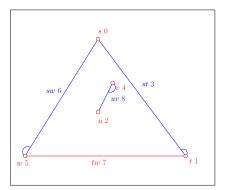


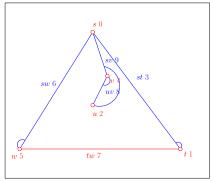


6.
$$\partial sw = s + w$$
, (w_5, sw_6)

7:
$$\partial tw = w_5 + t_1 = w_5 + t_1 + \partial sw$$

= $w_5 + t_1 + (s_0 + w_5)$
= $t_1 + s_0 + \partial st = t_1 + s_0 + (t_1 + s_0) = 0$.



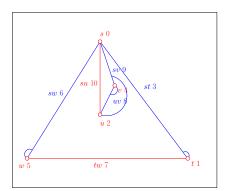


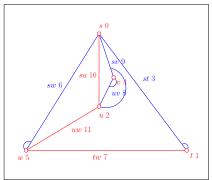
8.
$$\partial uv = u + v$$
, (v_4, uv_8)

9.
$$(u_2, sv_9)$$

9.
$$\partial sv = s_0 + v_4 = s_0 + v_4 + \partial uv$$

= $s_0 + v_4 + (u_2 + v_4)$
= $s_0 + \boxed{u_2}$





10. *su*

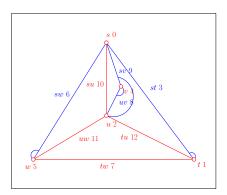
11. *uw*

10.
$$\partial su = s_0 + u_2 = s_0 + u_2 + \partial sv$$

 $= s_0 + u_2 + (s_0 + v_4)$
 $= u_2 + v_4 = u_2 + v_4 + \partial uv$
 $= 0$.

11.
$$\partial uw = u_2 + w_5 = u_2 + w_5 + \partial sw$$

 $= u_2 + w_5 + (s_0 + w_5)$
 $= s_0 + u_2 = s_0 + u_2 + \partial sv$
 $= s_0 + u_2 + (s_0 + v_4)$
 $= u + v = u_2 + v_4 + \partial uv$
 $= u_2 + v_4 + (u_2 + v_4)$
 $= 0$.

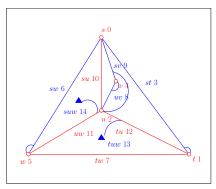


12. tu

13. $tuw,(tu_{12}, tuw_{13})$

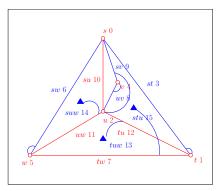
12.
$$\partial tu = t_1 + u_2 = t_1 + u_2 + \partial sv$$

 $= t_1 + u_2 + (s_0 + v_4)$
 $= t_1 + u_2 + s_0 + v_4 + \partial uv$
 $= t_1 + u_2 + s_0 + v_4 + (u_2 + v_4)$ 13. $\partial tuw = tu + uw + wt$
 $= t_1 + s_0$ (tuw, tu)
 $= s_0 + t_1 + \partial st$
 $= s_0 + t_1 + (t_1 + s_0)$
 $= 0$.



14.
$$\partial suw = uw + su + sw$$

 (uw_{11}, suw_{14})

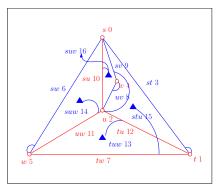


15. stu, (tw_7, stu_{15})

15.
$$\partial stu = su_{10} + tu_{12} + st_3$$

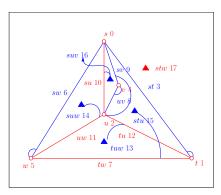
 $= su_{10} + st_3 + tu_{12} + \partial tuw$
 $= su_{10} + st_3 + tu_{12} + (tu_{12} + uw_{11} + tw_7)$
 $= su_{10} + st + uw_{11} + tw_7$
 $= su_{10} + st + uw_{11} + tw_7 + \partial suw$
 $= su_{10} + st + uw_{11} + tw_7 + (sw + su_{10} + uw_{11})$
 $= st + tw_7 + sw$

Hence we obtain the pair (stu_{15}, tw_7) .



16.
$$\partial suv = su + uv + sv$$

 (su_{10}, suv_{16})

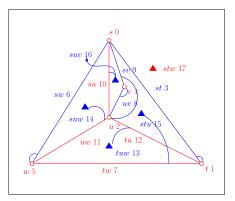


17. *stw*

17.
$$\partial stw = tw + sw + st$$

 $= sw + st + tw + \partial stu$
 $= sw + st + tw_7 + (st + tu_{12} + us_{10})$
 $= sw + st + tw_7 + (st + tu_{12} + us_{10}) + \partial tuw$
 $= sw + st + tw_7 + (st + tu_{12} + us_{10}) + (tu_{12} + uw_{11} + tw_7)$
 $= sw + us_{10} + uw_{11}$
 $= sw + us_{10} + uw_{11} + \partial suw$
 $= sw + us_{10} + uw_{11} + (su_{10} + uw_{11} + ws)$
 $= 0$.

Example



Creater	Killer
t_1	st ₃
u ₂	<i>SV</i> 9
<i>V</i> 4	uv ₈
W ₅	sw ₆
tw ₇	stu ₁₅
<i>su</i> ₁₀	suv ₁₆
uw ₁₁	suw ₁₄
<i>tu</i> ₁₂	tuw ₁₃

The unpaired creaters are s_0 and stw_{17} .

Incidence Matrix

Assuming an ordering of the (p-1) simplices and of the p-simplices, the boundary of a p-chian can be obtained by multiplication of the corresponding vector with the incidence matrix,

$$\partial(c_p)=D_pc_p.$$

The incidence matrix is defined as

$$D_{p}[i,j] = \begin{cases} 1 & \sigma_{i}^{p-1} \in \sigma_{j}^{p} \\ 0 & \sigma_{i}^{p-1} \notin \sigma_{j}^{p} \end{cases}$$

Incidence Matrix and Betti Number

A classic algorithm computes the Betti numbers of K by reducing its incidence matrices to Smith normal form. It uses row and column operations to zero out all entries except along an initial portion of the diagonal.

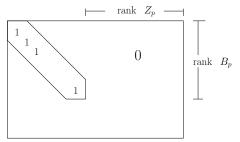


Figure: Smith norm of incidence matrix in \mathbb{Z}_2 .

The Betti number

$$\beta_p = \operatorname{rank} Z_p - \operatorname{rank} B_p.$$

39 / 64

Pairing Algorithm

Definition (Monotonous Filtering)

A filtering is monotonous, if in the ordering of K, any simplex σ is proceeded by its faces.

An algorithm computes the persistence diagrams by pairing the simplices, which uses column operator to reduce D to another 0-1 matrix R. Let $low_R(j)$ be the row index of the last 1 in column j of R, and (undefined if the column is zero).

Definition (Reduced Matrix and Pairing)

We call R reduced and lowR a pairing function, if

$$low_R(j) \neq low_R(j'),$$

whenever $j \neq j'$ specify two non-zero columns.



Pairing Algorithm

Algorithm: Incidence matrix reduction

- R ← D
- **2** for j = 1 to n do
- while $\exists j' < j$ with $low_R(j') = low_R(j)$ do
- add column j' to column j
- endwhile
- o endfor.

The pairing is given by

$$(\sigma_i, \sigma_j) \iff i = low_R(j)$$

 σ_i is positive, it generates a homology class; σ_j is negative, it kills a homolog class.



Killer



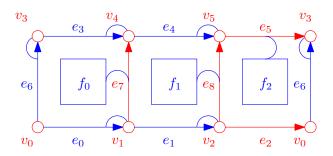


Figure: Generators and killers.

Boundary operator ∂_1 , incidence matrix D_1 ,

	e_0	e_1	e_2	<i>e</i> ₃	<i>e</i> ₄	<i>e</i> ₅	<i>e</i> ₆	e ₇	<i>e</i> ₈
v_0	1	0	1	0	0	0	1	0	0
v_1	1	1	0			0	0	1	0
v_2	0		1	0	0	0		0	1
<i>v</i> ₃	0	0	0	1	0	1	1	0	0
V4	0	0	0	1	1	0	0	1	0
<i>V</i> 5	0	0	0	0	1	1	0	0	1

$$1+2$$
, $4+5$, $3+7$, $4+8$

	<i>e</i> ₀	e_1	e_2	<i>e</i> ₃	<i>e</i> ₄	<i>e</i> ₅	<i>e</i> ₆	e ₇	<i>e</i> ₈
<i>v</i> ₀	1	0	1	0	0	0	1	0	0
v_1	1	1	1	0	0	0	0	1	0
<i>v</i> ₂	0	1	0	0	0		0	0	1
<i>V</i> 3	0	0	0	1	0		1		0
<i>V</i> ₄	0	0	0	1	1	1			
<i>V</i> 5	0	0	0	0	1	0	0	0	0

$$1+2$$
, $3+5$, $3+8$

	<i>e</i> ₀	e_1	<i>e</i> ₂	<i>e</i> ₃	<i>e</i> ₄	<i>e</i> ₅	<i>e</i> ₆	e ₇	<i>e</i> ₈
<i>v</i> ₀	1	0	0	0	0	0	1	0	0
v_1	1	1	0	0	0	0	0	1	0
<i>v</i> ₂	0	1	0	0	0	0	0	0	1
<i>V</i> 3	0	0	0	1	0	0	1	1	1
<i>V</i> ₄	0	0	0	1	1	0	0	0	0
<i>V</i> 5	0	0	0	0	1				0

$$6+7$$
, $6+8$

	<i>e</i> ₀	e_1	e_2	<i>e</i> ₃	<i>e</i> ₄	<i>e</i> ₅	<i>e</i> ₆	e ₇	<i>e</i> ₈
<i>v</i> ₀	1	0	0	0	0	0	1	1	1
v_1	1	1	0	0	0	0	0	1	0
<i>v</i> ₂	0	1	0	0	0	0	0	0	1
<i>V</i> 3	0	0	0	1	0	0	1	0	0
<i>V</i> ₄	0	0	0	1	1	0	0	0	0
<i>V</i> 5	0	0	0	0	1	0	0		0

$$0+7,1+8,0+8$$

	e_0	e_1	e_2	<i>e</i> ₃	<i>e</i> ₄	<i>e</i> ₅	<i>e</i> ₆	e ₇	<i>e</i> ₈
<i>v</i> ₀	1	0	0			0			
v_1	1	1	0	0	0	0	0	0	0
<i>v</i> ₂	0	1	0	0			0	0	0
<i>V</i> 3	0	0	0	1	0				
<i>V</i> 4	0	0	0	1	1			0	
<i>V</i> ₅	0	0	0	0	1			0	0

Generators e_2 , e_5 , e_7 , e_8 , corresponding to 0 columns. Killers corresponds to non-zero columns. Pairing

$$(e_0, v_1), (e_1, v_2), (e_3, v_4), (e_4, v_5), (e_6, v_3)$$



The pairing is

$$(f_0, e_7), (f_1, e_8), (f_2, e_5)$$



Topological Annulus



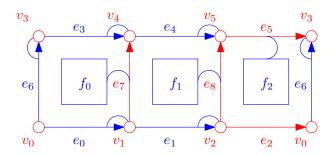
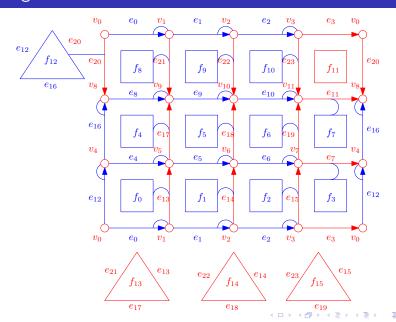
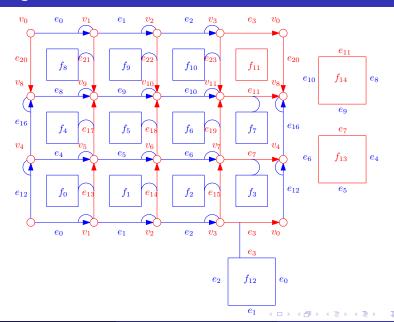


Figure: Topological Annulus. The unpaired creator is v_0 and e_2 .





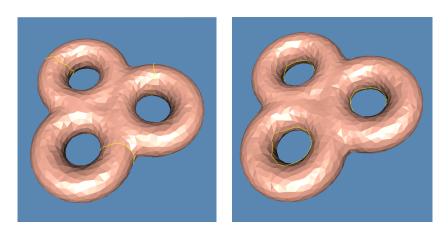


Figure: Handle and tunnel loops.

Handle Loop and Tunnel Loop

- **1** The simplices on the surface M are added into the filtration in any arbitrary order. Since $H_1(M)$ is of rank 2g, the algorithm Pair generates 2g number of unpaired positive edges.
- The simplices up to dimension 2 in I are added into the filtration. Since H₁(I) of rank g, half of 2g positive edges generated in step 1 get paired with the negative triangles in I. Each pair correponds to a killed loop, these g loops are handle loops.
- **3** Or the simplices up to dimension 2 in O are added into the filtration. Since $H_1(O)$ of rank g, half of 2g positive edges genrated in step 2 get paried with the negative triangles in O. Each pair corresponds to a killed loop, these g loops are tunnel loops.

Handle Loops, Tunnel Loops

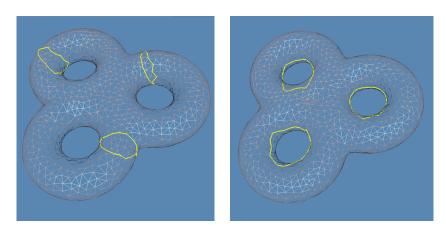
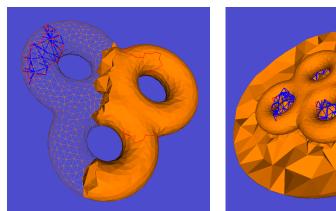


Figure: Handle and tunnel loops.



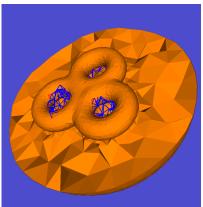


Figure: Interior and exterior volumes.

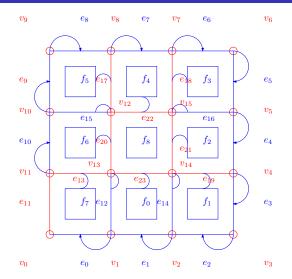
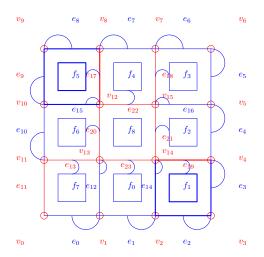
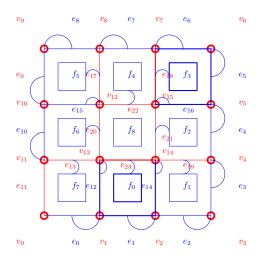


Figure: Quadrilateral example.

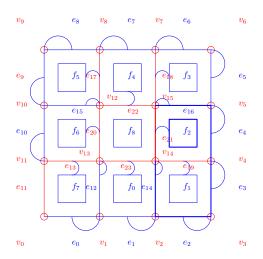


$$\partial f_5 = c_{17} = e_{15} + e_9 + e_8 + e_{17}, \ \partial f_1 = c_{19} = e_3 + e_2 + e_{14} + e_{19}.$$



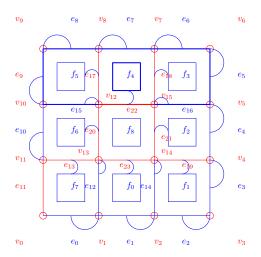
$$\partial f_3 = c_{18} = e_{16} + e_5 + e_6 + \frac{e_{18}}{18}, \ \partial f_0 = c_{23} = e_{12} + e_{14} + e_1 + \frac{e_{23}}{18}.$$



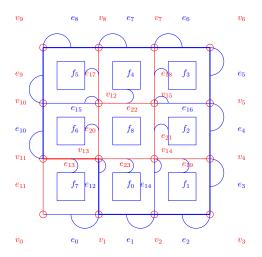


$$c_{21} = e_{16} + e_4 + e_3 + e_2 + e_{14} + \frac{e_{21}}{2}$$
, $\partial f_2 = c_{19} + c_{21}$



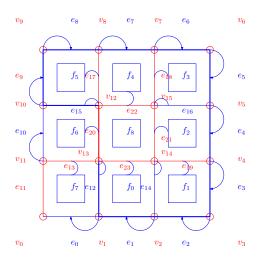


$$c_{22} = e_{15} + e_4 + e_{16} + e_9 + e_5 + e_8 + e_7 + e_6 + \frac{e_{22}}{2}, \ \partial f_4 = c_{22} + c_{18} + c_{17}$$

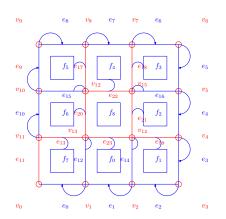


$$c_{13} = e_{12} + e_{10} + e_9 + e_8 + e_7 + e_6 + e_5 + e_4 + e_3 + e_2 + e_1 + e_{13}, \partial f_7 = c_{11} + c_{13}$$

$$c_{11} = e_0 + e_1 + e_2 + e_3 + e_4 + e_5 + e_6 + e_7 + e_8 + e_9 + e_{10} + e_{11}$$

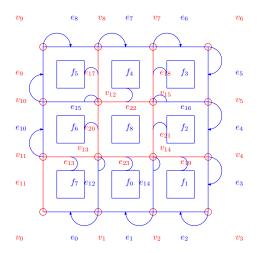


$$c_{20} = e_{15} + e_9 + e_8 + e_7 + e_6 + e_5 + e_4 + e_3 + e_2 + e_2 + e_{12} + \frac{e_{20}}{20}, \partial f_6 = c_{20} + c_{13}.$$



$$\partial f_0 = c_{23}, (e_{23}, f_0),$$

 $\partial f_1 = c_{19}, (e_{19}, f_1),$
 $\partial f_2 = c_{21} + \partial f_1, (e_{21}, f_2),$
 $\partial f_3 = c_{18}, (e_{18}, f_3),$
 $\partial f_5 = c_{17}, (e_{17}, f_5),$
 $\partial f_4 =$
 $c_{17} + c_{18} + c_{22}, (e_{22}, f_4),$
 $\partial f_6 = c_{13} + c_{20}, (e_{20}, f_6),$
 $\partial f_7 = c_{13} + c_{11}, (e_{13}, f_7),$



$$\begin{split} \partial f_8 &\to \partial f_0, \partial f_4, \partial f_2, \partial f_6, \partial f_1, \partial f_3, \partial f_5, \partial f_7 \\ &\to e_0 + e_1 + e_2 + e_3 + e_4 + e_5 + e_6 + e_7 + e_8 + e_9 + e_{10} + e_{11} \end{split}$$