1 REVIEW: STANDARD ALGORITHM TO COMPUTE PERSISTENT HOMOLOGY CYCLE BASES, BY MATRIX DECOMPOSITION

The first step in Algorithms 1 and 2 is to compute a persistent homology cycle basis. The standard method to compute such a basis invokes a so-called $R = DV$ decomposition of the boundary matrices $\partial_n$ [Cohen-Steiner et al. 2006]. Here we provide a brief review of this process; further details may be found in Cohen-Steiner et al. [2006]; de Silva et al. [2011].

To begin, we must place a total order on each set $S_n(K)$, in ascending order of birth. This naturally allows us to regard $\partial_n$ as an element of $G^{[S_{n-1}(K)] \times [S_n(K)]}$. The low function on a matrix $A \in G^{k \times l}$ is defined by

$$\text{low} : \{ j : A[:,j] \neq 0 \} \rightarrow \mathbb{Z}, \quad j \mapsto \max \{ i : A[i,j] \neq 0 \}.$$ 

We say that $A$ is reduced if low is injective. An $R = DV$ decomposition is a matrix equation where $R$ is reduced and $V$ is invertible and upper triangular.

Suppose that $R_n = \partial_n V_n$ is such a decomposition for each $n$. Let low$^n$ be the low function of $R_n$, and let $\Gamma_n = \{(j, \text{low}(j)) : R[:,j] \neq 0\}$ be the graph low$^n$. It can then be shown (Cohen-Steiner et al. [2006]; de Silva et al. [2011]) that

1. Each set $S_n(K)$ partitions into three disjoint subsets: $B_n \sqcup B_n^* \sqcup H_n$, where $B_n$ is the image of low$^{n+1}$ and $B_n^*$ is the domain of low$^n$.
2. If $B_n^*(t)$ denotes the set of simplices in $B_n$ born by time $t$, then $\partial_n[:,B_n^*(t)]$ is a basis for the space of boundaries $B_n(K_t)$.
3. Let $\hat{\Gamma}_n$ denote the subset of $\Gamma_n$ consisting of those pairs $(\sigma, \tau)$ such that Birth$(\sigma) \neq$ Birth$(\tau)$, and let $E_n = \{ \tau : (\sigma, \tau) \in \hat{\Gamma} \} \cup H_n$. Then $V_n[:,H_n] \cup R_n[:,E_n]$ is a persistent homology cycle basis. The lifespan of $V_n[:,\sigma]$ is $[\text{Birth}(\sigma), \infty)$ and the lifespan of $R_n[:,\tau]$ is $[\text{Birth}(\sigma), \text{Birth}(\tau)]$, where $\sigma$ is the unique simplex such that $(\sigma, \tau) \in \Gamma$.
4. In particular, the barcode of $K_\bullet$ may be read off from the $R = DV$ decompositions.

2 CORRECTNESS OF ALGORITHMS 1 AND 2

Here we provide proofs of correctness for Algorithms 1 and 2. As the details are primarily technical in nature, the exposition is fairly terse. The arguments are primarily self-contained, however we begin by recalling one result from Henselman and Ghrist [2016], which will be used in the proof of Algorithm 1.

2.1 Review: characterization of persistent homology bases

We will invoke a result from Henselman and Ghrist [2016] and Henselman-Petrusek [2017] which requires one new definition and one new notational convention. For notation, let $n$ be given, fix $\epsilon_i < \epsilon_j$, and set

$$X^{i,j} := Z_n(K_{\epsilon_i}) \cap B_n(K_{\epsilon_j}) \quad Y^{i,j} := Z_n(K_{\epsilon_i-1}) + B_n(K_{\epsilon_j-1}) \quad Q^{i,j} := \frac{X^{i,j}}{X^{i,j} \cap Y^{i,j}}.$$
We then define $q^{i,j}$ as the quotient map

$$q^{i,j} : X^{i,j} \rightarrow Q^{i,j}$$

**Definition 2.1.** A subset $E \subseteq X^{i,j}$ is an $(i,j)$-basis if $q^{i,j}|_E$ is injective and $q^{i,j}(E)$ is a basis of $Q^{i,j}$.

**Theorem 2.2** (Henselman and Ghrist [2016]; Henselman-Petrusek [2017]). A family of $n$-cycles $E$ is a persistent homology cycle basis iff

$$E^{i,j} := E \cap (X^{i,j} - Y^{i,j})$$

is an $(i,j)$ basis for all $i$ and $j$.

**Remark 2.3.** A special case of Theorem 2.2 (for simplex-wise filtrations) also appeared in [Wu et al., 2017, Theorem 1]. This construction is also closely related to that of pair groups, c.f. Edelsbrunner and Harer [2008].

### 2.2 Correctness of Algorithm 1

We restate Algorithm 1 for ease of reference.

**Algorithm 1** Edge-loss persistent cycle minimization

1. Compute a persistent homology basis $B$ for homology in dimension 1, with coefficients in $\mathbb{Q}$, using the standard matrix decomposition procedure described in the Supplementary Material. Arrange the elements of $B$ into an ordered sequence $Z^0 = (z^{1,1}, \ldots, z^{1,m})$.
2. for $j = 0, \ldots, m - 1$ do
3.   Solve Program (14) to optimize the $j + 1$th element of $Z^j$. Let $x$ denote the solution to this problem, and define $Z^{j+1}$ by replacing the $j + 1$th element of $Z^j$ with $x$. Concretely, $z^{j+1,1} = x$, and $z^{j+1,k} = z^{j,k}$ for $k \neq j$.
4. end for
5. Return $B^* := \{z^{m,1}, \ldots, z^{m,m}\}$, the set of elements in $Z^m$.

Recall that Program (14) optimizes the $j$th element of an ordered sequence of cycle representatives $Z = (z^1, \ldots, z^m)$. In particular, it seeks to minimize $x^{\text{Orig}} := z^j$. To define this program, we first construct a matrix $A$ such that $A[i,i] = z^i$ for $i = 1, \ldots, m$. We then define three index sets, $\mathcal{P}, \mathcal{Q}, \mathcal{R}$ such that

$$\mathcal{P} = \{i : \text{Birth}(z^i) \leq \text{Birth}(x^{\text{Orig}}), \text{Death}(z^i) \leq \text{Death}(x^{\text{Orig}}), i \neq j\}$$

$$\mathcal{Q} = \{\sigma \in S_2(K) : \text{Birth}(\sigma) \leq \text{Birth}(x^{\text{Orig}})\}$$

$$\mathcal{R} = \{\sigma \in S_1(K) : \text{Birth}(\sigma) \leq \text{Birth}(x^{\text{Orig}})\}$$

Program (14) can then be defined as follows.
\[ \minimize \|Wx\|_1 = \sum_{i=1}^{N} (x_i^+ + x_i^-) \]

subject to \( (x^+ - x^-) = x^{\text{Orig}}[R] + \partial_2[R, Q]q + A[R, P]p \)
\( p \in \mathbb{Q}^P \)
\( q \in \mathbb{Q}^Q \)
\( x \in G^R \)
\( x^+, x^- \geq 0 \)

\[ (14) \]

**Theorem 2.4.** For each \( k \), the family of cycles \( \{z_i^1, \ldots, z_i^m\} \) constructed in Algorithm [7] is a persistent homology cycle basis. Moreover, lifespans are preserved, in the sense that
\[ L(z_0^0, l) = L(z^k_l) \quad (S1) \]
for all \( k \) and \( l \).

**Proof.** We proceed by induction on \( k \), the base case \( k = 0 \) begins clear. Assume the desired conclusion holds for \( k \). For ease of reference, put
\[ Z := (z_1^1, \ldots, z_m^m) = (z_i^1, \ldots, z_i^m) = Z^k \]
\[ x^{\text{Orig}} := z_{k+1} \quad [\epsilon_i, \epsilon_j] := L(x^{\text{Orig}}) \]

We may then partition \( P \) as the disjoint union \( S \sqcup T \), where
\[ S = \{l \in P : \text{Birth}(z^l) < \text{Birth}(z_{k+1}) \text{ or } \text{Death}(z^l) < \text{Death}(z_{k+1})\} \]
\[ T = \{l \in P : \text{Birth}(z^l) = \text{Birth}(z_{k+1}), \text{Death}(z^l) = \text{Death}(z_{k+1}), \; l \neq k + 1\} \]

An optimal solution to Program (14) can then be expressed in form
\[ x = z_{k+1} + \partial_2[R, Q]q + A[R, P]q = z_{k+1} + \underbrace{\partial_2[R, Q]q + A[R, S]q[S]}_{u} + \underbrace{A[R, T]q[T]}_{v} + w \]
where
\[ u \in B(K_{i_j}) \subseteq X_{i,j} \cap Y_{i,j} \quad v \in X_{i,j} \cap Y_{i,j} \quad w \in \text{span} \left( \{z^t : t \in T\} \right) \]

Now put
\[ F := \{z^l : t \in T\} \sqcup \{z_{k+1}\} \]
\[ F' := \{z^l : t \in T\} \sqcup \{z_{k+1} + w\} \]
\[ F'' := \{z^l : t \in T\} \sqcup \{z_{k+1} + u + v + w\} = \{z^l : t \in T\} \sqcup \{x\} \]
Since $w \in \text{span}\{z^t : t \in T\}$, it is easily argued that $\text{span}(F) = \text{span}(F')$. Thus $\text{span}(q^{i,j}(F)) = \text{span}(q^{i,j}(F')) = Q^{i,j}$. Dimension counting thus implies that $q^{i,j}(F'')$ is an $(i,j)$-basis.

Given this observation, it is straightforward to verify that \( \{z^{k+1,1}, \ldots, z^{k+1,m}\} \) is a bona-fide cycle basis and \( \mathcal{L}(x) = \mathcal{L}(x^{\text{Orig}}) \). The desired conclusion follows.

This establishes our primary objective:

**Theorem 2.5.** The set \( B^* \) returned by Algorithm is a bona fide persistent homology cycle basis of the filtered simplicial complex \( K^\bullet \).

### 2.3 Correctness of Algorithm 2

Recall that [Obayashi 2018] defines a *persistent volume* for a birth-death pair \((\sigma_b, \sigma_d)\) as an \((n+1)\) chain \( v \in C^{n+1}(K_{d_i}) \) such that

\[
v = \sigma_d + \sum_{\sigma_k \in F_n} \alpha_k \sigma_k \tag{S2}
\]

\[
(\partial_{n+1} v)_\tau = 0 \quad \forall \tau \in F_n \tag{S3}
\]

\[
(\partial_{n+1} v)_{\sigma_b} \neq 0, \tag{S4}
\]

where

\[
F_l := \{ \sigma_k \in S_l(K) : b_i < k < d_i \} \tag{S5}
\]

is the family of \( l \)-simplices whose birth time lies strictly between \( b_i \) and \( d_i \). The linear program associated to \((\sigma_b, \sigma_d)\) in [Obayashi 2018] can then be summarized as

\[
\text{minimize } \text{loss}(v) \text{ subject to } (S2), (S3), (S4) \quad v \in C^{n+1}(K_{d_i}) \tag{10}
\]

Let us restate Algorithm 2 and the corresponding optimization problem, Program (15), for ease of reference.

Recall that we refer to Program (15) as the *general triangle-loss problem*.

\[
\text{minimize } \|Wv\|_1 = \sum_{i=1}^N (v^+_i + v^-_i) \text{ subject to } \partial_{n+1}[\sigma, \hat{F}_{n+1}]v \neq 0 \quad \partial_{n+1}[F_n, \hat{F}_{n+1}]v = 0 \quad v_\tau = 1 \quad v^+, v^- \geq 0 \quad v^+, v^- \in G_{\hat{F}_{n+1}} \tag{15}
\]
Algorithm 2 Triangle-loss persistent cycle minimization

1. Place a filtration-preserving linear order $\preceq(l)$ on $S_l(K)$ for each $l$.
2. Compute an $R = \partial_{l+1}V$ decomposition as described in [Cohen-Steiner et al. 2006] and the 
   Supplementary Material. We then obtain a set $\Gamma$ of birth/death pairs $(\sigma, \tau)$.
3. For each $(\sigma, \tau) \in \Gamma$ such that $\text{Birth}(\sigma) < \text{Birth}(\tau)$, put
   
   $$F_n := \{ \sigma' \in S_n(K) : \text{Birth}(\sigma') \leq \text{Birth}(\tau), \ \sigma \preceq^{(n)} \sigma' \}$$
   $$F_{n+1} := \{ \tau' \in S_{n+1}(K) : \text{Birth}(\sigma) \leq \text{Birth}(\tau'), \ \tau' \preceq^{(n+1)} \tau \}$$
   
   and $\hat{F}_{n+1} := F_{n+1} \cup \{ \tau \}$. Compute a solution to the corresponding Program (15), and denote this 
   solution by $x^{\sigma, \tau}$.
4. Put $\hat{D} := \{ \partial_{n+1}(x^{\sigma, \tau}) : (\sigma, \tau) \in \Gamma \text{ and } \text{Birth}(\sigma) < \text{Birth}(\tau) \}$ and let $\hat{D}' := \{ z \in M : \text{Death}(z) = \infty \}$, where $M$ is a persistent homology cycle basis calculated by the standard $R = DV$ method.
5. Return $D := \hat{D} \cup \hat{D}'$.

To verify that Algorithm 2 returns a bona-fide persistent homology cycle basis, let us begin by placing 
the elements of $K$ into a sequence $(\sigma_1, \ldots, \sigma_{|K|})$ by ordering simplices first by birth time, second (that 
is, breaking ties when birth times agree) by dimension, and finally (breaking ties when birth times and 
dimensions agree) by the chosen linear orders $\preceq(l)$. It is simple to verify that this rule defines a unique 
linear order on $K$, and that the filtration $K_\bullet$ defined by

$$K_1 := \{ \sigma_1, \ldots, \sigma_1 \}$$

is a simplex-wise refinement of $K_\bullet$.

**Theorem 2.6.** Let $(\sigma, \tau)$ be a birth-death pair, and choose $b_i, d_i$ such that $(\sigma_{b_i}, \sigma_{d_i}) = (\sigma, \tau)$. Then the 
sets $F_n$ and $F_{n+1}$ defined in Algorithm 2 both satisfy (S5). Consequently, Program (15) is a special case 
of Program (10).

**Proof.** The proof is a straightforward exercise in definition checking.

Now let $\hat{B}$ be a set containing the boundary of one optimal solution $x^{\sigma, \tau}$ to Program (15) for each 
birth-death pair $(\sigma, \tau) = (\sigma_{b_i}, \sigma_{d_i})$ (even if $\text{Birth}(\sigma) = \text{Birth}(\tau)$). Let $\mathcal{N}$ be a persistent homology cycle basis for $K_\bullet$, and let $\hat{B}' = \{ z \in \mathcal{N} : \text{Death}(z) = \infty \}$ be the collection of cycle representatives in $\mathcal{N}$ that 
never die. Then, by combining Theorem 2.6 with Theorem 5 of [Obayashi 2018], we find that $B := \hat{B} \cup \hat{B}'$ 
is a persistent homology cycle basis of $K_\bullet$.

If we assume that the bounding volumes used to obtain $\hat{B}$ are the same as those used to obtain $\hat{D}$ in 
Algorithm 2 and, likewise, that $\hat{B}' = \hat{D}'$ (this will be true provided we use the order $\preceq(l)$ when ordering 
columns of $\partial_l$ for the $K_\bullet$ calculation) then

$$D = \{ z \in B : \text{Birth}(z) < \text{Death}(z) \}$$

where Birth and Death are the birth and death functions of $K_\bullet$, not $K_\bullet$.

From here, it remains only to verify that $D$ is a bona fide persistent homology cycle basis $K_\bullet$. This 
follows from the following general observation.
THEOREM 2.7. Let $L_\bullet'$ be a refinement of a filtration $L_\bullet$ on a simplicial complex $L$, and let $B$ be a persistent homology cycle bases for $L_\bullet'$. If Birth and Death are the birth and death functions of $L_\bullet$ and $D := \{z \in B : \text{Birth}(z) < \text{Death}(z)\}$, then $D$ is a persistent homology cycle basis of $L_\bullet$.

PROOF. Recall that, by definition, a set $E$ is a persistent homology cycle basis of $K_\bullet$ iff two criteria hold: (i) $L(z)$ must be nonempty for each $z \in E$, and (ii) $\{[z] \in H_n(K_{\epsilon_i}; G) : \epsilon_i \in L(z)\}$ is a basis for $H_n(K_{\epsilon_i}; G)$ for each $i$. Since $B$ is a persistent homology cycle basis for $L_\bullet'$, it is a straightforward exercise in definition checking to verify that $D$ is a persistent homology cycle basis for $L_\bullet$.

This establishes our primary objective:

THEOREM 2.8 (Correctness of Algorithm 2). Algorithm 2 returns a bona-fide persistent homology cycle basis of $K_\bullet$.

3 SUPPLEMENTARY TABLES AND FIGURES
Figure S1: Computation time of the GLPK linear solver (red) and the Gurobi linear solver (green) to solve the uniform/length-weighted edge-loss minimal problems in Algorithm 1. We perform experiments on 90 data sets, 10 for each dimension 2-10, generated from the normal distribution. The horizontal axis is the dimension of the data set, and the vertical axis is the time it takes to solve an optimization problem. We observe that the Gurobi solver is consistently faster than the GLPK solver and that computation time seems fairly constant across dimension.
Table S1. Classifying the coefficients of the optimal cycles for all of the real-world data discussed in Section 5.1 as well as all of the synthetic sets discussed in Section 5.2. The rows are labeled by the coefficient type of the cycle representatives: “Integral” means the coefficients for the cycle representative $x$ are in $\mathbb{Z}$ and “$\text{In } \{-1, 0, 1\}$” means the coefficients for the representative $x$ are in $\{-1, 0, 1\}$. For the columns, $x$ represents the optimal representative with its superscript indicating the type of optimization problem: $I$ for integer programming and $NI$ for linear programming, and its subscript indicating the type of optimal cycle: $E$-Len, $E$-Unif, $T$-Unif refer to edge loss length-weighted minimal (minimizing total length of 1-simplices), edge loss uniform (minimizing total number of 1-simplices), and triangle loss uniform (minimizing the number of 2-simplices a cycle representative bounds), respectively.

| Coefficient Type | Randomly Generated Data Sets | Real-World Data Sets |
|------------------|-----------------------------|----------------------|
| $x_{E\text{-Len}}^I$ | $100\%$ | $100\%$ |
| $x_{E\text{-Unif}}^I$ | $100\%$ | $100\%$ |
| $x_{E\text{-Unif}}^{NI}$ | $100\%$ | $100\%$ |

| Coefficient Type | Randomly Generated Data Sets | Real-World Data Sets |
|------------------|-----------------------------|----------------------|
| $x_{T\text{-Unif}}^I$ | $100\%$ | $100\%$ |
| $x_{T\text{-Area}}^I$ | $99.99\%$ | $100\%$ |
| $x_{T\text{-Area}}^{NI}$ | $100\%$ | $100\%$ |

REFERENCES

Cohen-Steiner D, Edelsbrunner H, Morozov D. Vines and vineyards by updating persistence in linear time. *Proceedings of the twenty-second annual symposium on Computational geometry* (2006), 119–126.

de Silva V, Morozov D, Vejdemo-Johansson M. Dualities in persistent (co)homology. *Inverse Problems* **27** (2011) 124003, 17. doi:10.1088/0266-5611/27/12/124003.

Henselman G, Ghrist R. Matroid Filtrations and Computational Persistent Homology. *ArXiv e-prints* (2016).

Henselman-Petrusek G. *Matroids and canonical forms: theory and applications*. Ph.D. thesis, University of Pennsylvania (2017).

Wu P, Chen C, Wang Y, Zhang S, Yuan C, Qian Z, et al. Optimal topological cycles and their application in cardiac trabeculae restoration. *Information Processing in Medical Imaging, IPMI* (2017), 80–92.

Edelsbrunner H, Harer J. Persistent homology—a survey. *Discrete & Computational Geometry - DCG* **453** (2008). doi:10.1090/conm/453/08802.

Obayashi I. Volume-optimal cycle: Tightest representative cycle of a generator in persistent homology. *SIAM Journal on Applied Algebra and Geometry* **2** (2018) 508–534.