Optimal Homology Basis for Arbitrary Complexes – summary

After a short introduction some background information was given about simplicial complexes, because we had not seen them so far. The first part was the definition of a k-dimensional simplex. Formally the simplex is defined as the convex hull of k+1 affinely independent points. For dimension 0 this is just a point, or vertex. A simplex of dimension 1 would be a line segment with two end points, also called edge. For k=2 one would get a triangle and in the third dimension it's a tetrahedron, consisting of four triangular faces.

This now leads to the definition of a simplicial complex. It is a finite collection of simplices K such that the faces of the simplices are also in K and that an intersection of two simplices is always a face of both simplices. A face is in this context a component of smaller dimension, e.g. the two points of an edge. Therefore a triangle with an edge attached to one of its vertices is a valid complex, whereas two triangles partially sharing an edge yet with each triangle having different end points for that edge do not form a valid simplicial complex. One important note is that three edges forming a triangle with nothing to fill it is not the same as a 2-dimensional triangular simplex. Finally, to make working with complexes easier the next definition introduces abstract simplicial complexes. Here the last condition about intersections is left out and one can thereby construct more complexes and does not have to worry about an embedding into the Euclidean space.

To finish the background part there are some notations and remarks left. Firstly, if an abstract simplicial complex has dimension 1, it resembles a graph. Furthermore, if the complex is not abstract, the graph is embedded on a surface. The representation of a simplicial complex is a set of all its faces and their faces recursively – e.g. a triangle {{1},{2},{3}}, consisting of its three vertices, labeled from 1 to 3. The dimension of a face is defined as $\dim(\sigma)=\mathrm{card}(\sigma)-1$ and lastly, the vertex set Vert(K) consists of all faces of dimension 0 in K.

With the basics done the next part will now deal with $\mathbb{Z}_2$-Homology. For that some more definitions are needed, the first being the definition for a d-chain. A d-chain is a linear combination of d-simplices of a simplicial complex K, therefore it is a subset of all d-simplices in K. $C_d(K)$ is the set of all d-chains in K. Every two appearances of a simplex in the chain cancel each other out, such that at the end every simplex appears at most once in the chain. With this knowledge we can now define how the boundary operator works. The operator is denoted by ∂_d and it maps a d-chain $c \in C_d(K)$ to the (d-1)-chain $c' \in C_{d-1}(K)$ that consists of all the faces of dimension d-1 of c. E.g. a triangle would be mapped to its three edges and two edges connected by a shared vertex would be mapped to the two points that are not shared, since the shared point would appear twice and thus cancel out in the resulting set. If applying ∂_d to some c yields the empty set, c is called a cycle.

Hence the set of all cycles $Z_d(K)$ is the kernel of ∂_d. Logically the results of the boundary operator are called boundaries and the set of all boundaries $B_d(K)$ is also the image of the ∂_{d+1} operator. Applying ∂ twice always yields the empty set 0. From that one can conclude that $B_d(K) \subset Z_d(K)$ holds, so every boundary is also a cycle.

With this property one can define the homology group $H_d(K) := Z_d(K) \backslash B_d(K)$. It is the set of homology classes of dimension d and if two cycles belong to the same homology class, they are homologous. After that comes the definition of the distance d(p,q) between two points p and q in a simplex, given as the shortest path from p to q. For this we assume that every edge has a nonnegative weight. Furthermore, the distance from p to a simplex σ is given as the maximum of the distances from p to one of the vertices of Vert(σ). Finally one can define the geodesic ball B_p^r of radius r around the vertex p, that is to say a set with every simplex that has distance r or smaller from p. Now, to measure a homology class $h \in H_d(K)$, its size S(h) will be defined as the minimal radius r such that one can find a vertex p in K and a cycle $z \in h$ with z being a subset of the geodesic ball B_p^r. If β_d is the dimension of $H_d(K)$, then the size of the basis $h_1,...,h_{\beta d}$ for the homology group is the sum $S(h_1)+...+S(h_{\beta d})$. β_d is called the Betti number. Note that the geodesic ball is just an approximation that is not guaranteed to be close to the actual size, since a cycle inside the ball could still wiggle a lot and thus be greater than the ball.

With all the definitions done one can now move on to the actual algorithm to compute the optimal homology basis for a simplicial complex. The algorithm takes K as input and uses it to initialize K^1, while initializing the set for the basis B with the empty set. Then it computes β_d times the smallest homology class h_ℓ in K^ℓ and adds it to B, with ℓ increasing from 1 to β_d with every step and $K^{\ell+1} := K^\ell$ with h_ℓ sealed. Sealing works as follows: Attach a new object to the complex such that it fills the hole of h_ℓ. By doing this the homology class becomes trivial. The newly added cell gets weight ∞ to prevent interference with the size function.

The next step is to explain how the smallest homology class of K^ℓ can be computed. One has to fix a vertex p and consider all geodesic balls with growing radius centering at p. While adding new objects to K we create new cycles. If a cycle is created we say it is born and if it disappears by sealing it we say it dies. Persistent cycles are those which never die. The youngest essential cycle is the one that is still alive and was created before all other living cycles. It is the candidate for the next non-trivial homology class. These operations take time $O(n^4)$, where n is the number of simplices and the output will be the vertex p and the radius for the geodesic ball. Since this step is

repeated β_d times, the overall running time is $O(\beta_d n^4)$.

One can improve this running time by not recomputing everything when moving over the different vertices and by using the property that the birth time of the first non-bounding cycle for neighboring vertices differs by at most 1, assuming that every edge has weight 1. For our search we use breadth-first search, creating a spanning tree and for that we also assume that the complex is connected. It is important to determine whether a subcomplex carries any non-trivial homology of K. Therefore another lemma is needed. For any matrix $A=[A_1\ A_2]$ the equation $\dim(\{A\gamma|A_2\gamma = 0\}) = \text{rank}(A) - \text{rank}(A_2)$ holds, where the matrices represent the set of boundaries of K intersected with the set of cycles of K that are carried by the subcomplex and the set of cycles of K intersected with the set of cycles of K that are carried by the subcomplex. If the ranks are different, the subcomplex carries a non-trivial homology of K.

Lastly, if β_d and d have at most size O(log n), we get an overall running time of $O(\beta_d n^3 \log^2 n)$ with the rank computation. If they are larger one would use the algorithm without improvement, running in time $O(\beta_d n^4)$. That concludes the construction of the algorithm with an optimal running time of $O(\beta_d n^3 \log^2 n)$.