## Article

# Approximate Triangulations of Grassmann Manifolds 

Kevin P. Knudson (1)<br>Department of Mathematics, University of Florida, Gainesville, FL 32611, USA; kknudson@ufl.edu

Received: 24 June 2020; Accepted: 16 July 2020; Published: 17 July 2020


#### Abstract

We define the notion of an approximate triangulation for a manifold $M$ embedded in Euclidean space. The basic idea is to build a nested family of simplicial complexes whose vertices lie in $M$ and use persistent homology to find a complex in the family whose homology agrees with that of $M$. Our key examples are various Grassmann manifolds $G_{k}\left(\mathbb{R}^{n}\right)$.


Keywords: Grassmannian; persistent homology; Vietoris-Rips complex; witness complex; triangulation
MSC: 55N31; 57Q15

## 1. Introduction

Smooth manifolds admit piecewise-linear triangulations [1]. However, there are many subsequent questions one might ask: How many simplices are required? What is the minimal number of vertices? Is there an algorithm to construct a triangulation?

A great deal of work in algebraic topology has been devoted to these topics. The question of the number of simplices required to triangulate a given manifold is often attacked by sophisticated cohomological methods involving characteristic classes (such arguments also often yield estimates on the minimal embedding dimension for the manifold). Surprisingly, much of this work is very recent [2,3]. A main result in [3] is the following.

Theorem 1 ([3], Theorem 3.10). Every triangulation of the Grassmann manifold $G_{k}\left(\mathbb{R}^{n+k}\right)$ must have at least

$$
[(n+k)(n+k+1)-2 k n] \cdot\left(2^{k n+1}-1\right)
$$

simplices.
For example, any triangulation of the manifold $G_{2}\left(\mathbb{R}^{4}\right)$ must have at least 372 simplices. The Grassmann manifolds will be defined in Section 2.1 below. These are important spaces to study because of their utility in algebraic topology, especially with respect to the study of characteristic classes [4]. Moreover, while the theory of these spaces is rich, there are still many unsolved questions about them. For example, if we pass to the limit as $n \rightarrow \infty$, we obtain an infinite Grassmannian $G_{k}\left(\mathbb{R}^{\infty}\right)$; it is still unknown [5] if these spaces are triangulable for $k \geq 3$. If one could construct compatible triangulations of the various $G_{k}\left(\mathbb{R}^{n}\right)$, then one might be able to answer this question. Our methods here do not apply, however.

Unfortunately, most results along these lines are not constructive; that is, the proofs do not yield an explicit triangulation of the manifold. In fact, if one seeks a triangulation of a Grassmannian $G_{k}\left(\mathbb{R}^{n}\right)$, the end result is usually disappointment. For the smallest nontrivial space, $G_{1}\left(\mathbb{R}^{3}\right)=\mathbb{R} P^{2}$, there are many well-known small triangulations, and even an algorithm to generate a triangulation from any collection of points in general position [6]. Beyond that, however, results are sparse.

In this paper, we develop a procedure to find what we call an approximate triangulation of the manifold $G_{k}\left(\mathbb{R}^{n}\right)$ (Definition 3). The basic idea is to first generate a sample of points on $G_{k}\left(\mathbb{R}^{n}\right)$.

This already leads to technical difficulties involving embeddings of these spaces into a Euclidean space $\mathbb{R}^{N}$, but we are able to solve this. We then build a nested family of simplicial complexes on the point cloud, parametrized by the positive real numbers. The persistent homology of this family is then computed and we identify an interval of parameters for which the $\bmod 2$ homology of the complexes in that range agrees with that of $G_{k}\left(\mathbb{R}^{n}\right)$. Such a complex is then a viable model for the manifold: its vertices lie in $G_{k}\left(\mathbb{R}^{n}\right) \subset \mathbb{R}^{N}$ and it has the correct mod 2 homology. (As a further check, we can compute the mod 3 homology to ensure it is correct as well if we wish.) We then implement this procedure for the following spaces: $\mathbb{R} P^{2} \subset \mathbb{R}^{4}, \mathbb{R} P^{2} \subset \mathbb{R}^{5}, \mathbb{R} P^{3} \subset \mathbb{R}^{9}$, and $G_{2}\left(\mathbb{R}^{4}\right) \subset \mathbb{R}^{16}$. Computational limitations have so far prohibited further calculations; we discuss this in Section 4.

In effect, what we are doing is taking a tool developed for examining unknown objects (persistent homology modules built from point clouds) and using it on spaces that are well-understood (Grassmann manifolds) to build new structures associated with those spaces (approximate triangulations). We have chosen the Grassmannians because their homology is relatively easy to compute and they arise in numerous areas in mathematics (topology, algebraic geometry, combinatorics). These same ideas could be applied to any manifold whose homology groups are known a priori to produce an approximate triangulation, which then might be useful for other applications (e.g., storing the manifold efficiently in a computer).

## 2. Materials and Methods

Further details and proofs of the results in Sections 2.1 and 2.2 may be found in [4].

### 2.1. Grassmann Manifolds

Denote by $\mathbb{R}^{n}$ the Euclidean space of dimension $n$. By a $k$-frame in $\mathbb{R}^{n}$, we mean a $k$-tuple of linearly independent vectors; denote by $V_{k}\left(\mathbb{R}^{n}\right)$ the collection of $k$-frames in $\mathbb{R}^{n}$. This is an open subset of the $k$-fold Cartesian product $\mathbb{R}^{n} \times \cdots \times \mathbb{R}^{n}$.

Definition 1. The Grassmann manifold $G_{k}\left(\mathbb{R}^{n}\right)$ is the set of all $k$-dimensional planes through the origin in $\mathbb{R}^{n}$. It is topologized via the quotient map $V_{k}\left(\mathbb{R}^{n}\right) \rightarrow G_{k}\left(\mathbb{R}^{n}\right)$ which takes a $k$-frame to the $k$-plane it spans.

When $k=1$, we see that $G_{1}\left(\mathbb{R}^{n}\right)$ is the real projective space $\mathbb{R} P^{n-1}$, a manifold of dimension $n-1$. In general, we have the following result.

Lemma 1. The Grassmannian $G_{k}\left(\mathbb{R}^{n}\right)$ is a compact manifold of dimension $k(n-k)$. The map $X \rightarrow X^{\perp}$, which takes a $k$-plane to its orthogonal complement is a diffeomorphism between $G_{k}\left(\mathbb{R}^{n}\right)$ and $G_{n-k}\left(\mathbb{R}^{n}\right)$.

### 2.2. Schubert Cells

Grassmannians have a well-known cell decompostion into Schubert cells. Consider the sequence of subspaces of $\mathbb{R}^{n}: \mathbb{R}^{0} \subset \mathbb{R}^{1} \subset \mathbb{R}^{2} \subset \cdots \subset \mathbb{R}^{n}$, where $\mathbb{R}^{i}$ consists of the vectors of the form $\left(a_{1}, \ldots, a_{i}, 0, \ldots, 0\right)$. Any $k$-plane $X$ gives rise to a sequence of integers

$$
0 \leq \operatorname{dim}\left(X \cap \mathbb{R}^{1}\right) \leq \operatorname{dim}\left(X \cap \mathbb{R}^{2}\right) \leq \cdots \leq \operatorname{dim}\left(X \cap \mathbb{R}^{n}\right)=k
$$

Consecutive integers differ by at most 1 .
Definition 2. A Schubert symbol $\sigma=\left(\sigma_{1}, \ldots, \sigma_{k}\right)$ is a sequence of $k$ integers satisfying

$$
1 \leq \sigma_{1}<\sigma_{2}<\cdots<\sigma_{k} \leq n
$$

Given a Schubert symbol $\sigma$, let $e(\sigma) \subset G_{k}\left(\mathbb{R}^{n}\right)$ denote the set of $k$-planes $X$ such that

$$
\operatorname{dim}\left(X \cap \mathbb{R}^{\sigma_{i}}\right)=i, \operatorname{dim}\left(X \cap \mathbb{R}^{\sigma_{i}-1}\right)=i-1
$$

Each $X \in G_{k}\left(\mathbb{R}^{n}\right)$ belongs to precisely one of the sets $e(\sigma)$.
Lemma 2. $e(\sigma)$ is an open cell of dimension $d(\sigma)=\left(\sigma_{1}-1\right)+\left(\sigma_{2}-2\right)+\cdots+\left(\sigma_{k}-k\right)$.
In terms of matrices, $X \in e(\sigma)$ if and only if it can be described as the row space of a $k \times n$ matrix of the form

$$
\left[\begin{array}{ccccccccccccccc}
* & \cdots & * & 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\
* & \cdots & * & * & * & \cdots & * & 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\
\vdots & & & & & & & & & & & & & & \vdots \\
* & \cdots & * & * & * & \cdots & * & * & * & \cdots & * & 1 & 0 & \cdots & 0
\end{array}\right]
$$

where the $i$-th row has $\sigma_{i}$-th entry positive (say equal to 1 ) and all subsequent entries zero. Equivalently, we could (and do in the sequel) consider the column space of the transpose of this matrix.

For example, the possible Schubert symbols and cells for $G_{2}\left(\mathbb{R}^{4}\right)$ are as follows. Such a symbol has the form $\sigma=\left(\sigma_{1}, \sigma_{2}\right)$ where $1 \leq \sigma_{1}<\sigma_{2} \leq 4$.

| $\sigma$ | $d(\sigma)$ |
| :---: | :---: |
| $(1,2)$ | 0 |
| $(1,3)$ | 1 |
| $(1,4)$ | 2 |
| $(2,3)$ | 2 |
| $(2,4)$ | 3 |
| $(3,4)$ | 4 |

Theorem 2. The $\binom{n}{k}$ sets $e(\sigma)$ form the cells of a $C W$-decomposition of $G_{k}\left(\mathbb{R}^{n}\right)$.
Proposition 1. The number of $r$-cells in $G_{k}\left(\mathbb{R}^{n}\right)$ is equal to the number of partitions of $r$ into at most $k$ integers, each of which is $\leq n-k$.

The $\bmod 2$ homology of $G_{k}\left(\mathbb{R}^{n}\right)$ is easily computed from the Schubert cell decomposition: since the induced boundary maps are all either 0 or multiplication by 2 , the $\bmod 2$ homology has a basis corresponding to the cells.

Continuing the example of $G_{2}\left(\mathbb{R}^{4}\right)$, we have

$$
H_{i}\left(G_{2}\left(\mathbb{R}^{4}\right), \mathbb{Z} / 2\right)= \begin{cases}\mathbb{Z} / 2 & i=0 \\ \mathbb{Z} / 2 & i=1 \\ \mathbb{Z} / 2 \oplus \mathbb{Z} / 2 & i=2 \\ \mathbb{Z} / 2 & i=3 \\ \mathbb{Z} / 2 & i=4\end{cases}
$$

Since we will need it below, we also note the integral homology of $G_{2}\left(\mathbb{R}^{4}\right)$ [7]:

$$
H_{i}\left(G_{2}\left(\mathbb{R}^{4}\right), \mathbb{Z}\right)= \begin{cases}\mathbb{Z} & i=0 \\ \mathbb{Z} / 2 & i=1,2 \\ 0 & i=3 \\ \mathbb{Z} & i=4\end{cases}
$$

Using the Universal Coefficient Theorem, one then quickly deduces that the homology groups $H_{i}\left(G_{2}\left(\mathbb{R}^{4}\right), \mathbb{Z} / 3\right)$ are $\mathbb{Z} / 3$ for $i=0,4$ and 0 , otherwise.

### 2.3. Persistent Homology

Suppose we are given a finite nested sequence of finite simplicial complexes

$$
K_{R_{1}} \subset K_{R_{2}} \subset \cdots \subset K_{R_{p}}
$$

where the $R_{i}$ are real numbers $R_{1}<R_{2}<\cdots<R_{p}$. For each homological degree $\ell \geq 0$, we then obtain a sequence of homology groups and induced linear transformations (homology with $\mathbb{Z} / 2$-coefficients for simplicity)

$$
H_{\ell}\left(K_{R_{1}}\right) \rightarrow H_{\ell}\left(K_{R_{2}}\right) \rightarrow \cdots \rightarrow H_{\ell}\left(K_{R_{p}}\right)
$$

Since the complexes are finite, each $H_{\ell}\left(K_{R_{i}}\right)$ is a finite-dimensional vector space. Thus, there are only finitely many distinct homology classes. A particular class $z$ may come into existence in $H_{\ell}\left(K_{R_{s}}\right)$, and then one of two things happens. Either $z$ maps to 0 (i.e., the cycle representing $z$ gets filled in) in some $H_{\ell}\left(K_{R_{t}}\right), R_{s}<R_{t}$, or $z$ maps to a nontrivial element in $H_{\ell}\left(K_{R_{p}}\right)$. This yields a barcode, a collection of interval graphs lying above an axis parametrized by $R$. An interval of the form $\left[R_{s}, R_{t}\right]$ corresponds to a class that appears at $R_{s}$ and dies at $R_{t}$. Classes that live to $K_{R_{p}}$ are usually represented by the infinite interval $\left[R_{s}, \infty\right)$ to indicate that such classes are real features of the full complex $K_{R_{p}}$.

As an example, consider the boundary of the tetrahedron $T$ with filtration

$$
T_{0} \subset T_{1} \subset T_{2} \subset T_{3} \subset T_{4} \subset T_{5}=T
$$

defined by $T_{0}=\left\{v_{0}, v_{1}, v_{2}, v_{3}\right\}, T_{1}=T_{0} \cup\{$ all edges $\}, T_{2}=T_{1} \cup\left[v_{0} v_{1} v_{2}\right], T_{3}=T_{2} \cup\left[v_{0} v_{1} v_{3}\right]$, $T_{4}=T_{3} \cup\left[v_{0} v_{2} v_{3}\right]$, and $T_{5}=T$ (this is topologically a 2-sphere). The barcodes for this filtration are shown in Figure 1. Note that, initially, there are four components ( $\beta_{0}=4$ ), which get connected in $T_{1}$, when three independent 1-cycles are born $\left(\beta_{1}=3\right)$. These three 1-cycles die successively as triangles get added in $T_{2}, T_{3}$, and $T_{4}$. The addition of the final triangle in $T_{5}$ creates a 2-cycle $\left(\beta_{2}=1\right)$.


Figure 1. The barcodes for a filtration of the boundary of the tetrahedron.
For analyzing point cloud data, one needs a simplicial complex modeling the underlying space. Since it is impossible to know a priori if a complex is "correct", one builds a nested family of complexes approximating the data cloud, computes the persistent homology of the resulting filtration, and looks for homology classes that exist in long sections of the filtration. We discuss two popular methods for doing this in the next subsection.

### 2.4. Vietoris-Rips and Witness Complexes

Now suppose we are given a discrete set $X$ of points in some metric space (typically a Euclidean space $\mathbb{R}^{m}$ ). The standard example of such an object is a sample of points from some geometric object $M$. We would like to recover information about $M$ from the sample $X$, and the first step is to obtain an approximation of $M$ using only the point cloud $X$. There are many such techniques; perhaps the most classical is the Delaunay triangulation of $X$. This is defined as follows. Say $X=\left\{x_{1}, x_{2}, \ldots, x_{r}\right\} \subset \mathbb{R}^{m}$. The Voronoi decomposition of $\mathbb{R}^{m}$ relative to $X$ is the partition of $\mathbb{R}^{m}$ into cells $V\left(x_{i}\right), i=1, \ldots, r$, defined by

$$
V\left(x_{i}\right)=\left\{x \in \mathbb{R}^{m}:\left\|x-x_{i}\right\| \leq\left\|x-x_{j}\right\|, j \neq i\right\}
$$

The corresponding Delaunay triangulation, $\operatorname{Del}(X)$, is the nerve of the Voronoi decomposition; that is, a collection $V\left(x_{i_{0}}\right), \ldots, V\left(x_{i_{\ell}}\right)$ forms an $\ell$-simplex in $\operatorname{Del}(X)$ if $\cap_{j=0}^{\ell} V\left(x_{i_{j}}\right) \neq \varnothing$. One obtains a geometric realization of $\operatorname{Del}(X)$ via the map $V\left(x_{i}\right) \mapsto x_{i}$. See Figure 2 for an example.


Figure 2. (a) a Delaunay triangulation of a collection of points in the plane with the corresponding Voronoi diagram, and (b) two associated witness complexes.

While the Delaunay triangulation provides a good approximation to the underlying space $M$, it has several disadvantages. If the point cloud $X$ is large, there will be a very large number of simplices in $\operatorname{Del}(X)$. In addition, $\operatorname{Del}(X)$ suffers from the "curse of dimensionality;" that is, if the ambient dimension $(m)$ is large, calculating the Voronoi decomposition is computationally expensive.

There are many popular alternatives to the Delaunay triangulation. The one used most often is the Vietoris-Rips complex, which is built as follows. Consider the point cloud $X$ and let $r>0$. The Vietoris-Rips complex with parameter $r$ is the simplicial complex $V R(X, r)$ whose $k$-simplices are

$$
\left\{\left(x_{0}, \ldots, x_{k}\right): d\left(x_{i}, x_{j}\right)<r, i \neq j\right\}
$$

That is, if one imagines a ball of radius $r / 2$ around each point $x \in X$, then we join the points $x_{i}$ and $x_{j}$ with an edge if the balls intersect. Observe that if $r<r^{\prime}$ then there is an inclusion of complexes $V R(X, r) \subset V R\left(X, r^{\prime}\right)$. We therefore have a nested sequence of complexes $\{V R(X, r)\}_{r \geq 0}$ and we may study the persistent homology of this filtration. The corresponding barcodes yield information about the topology of the underlying space $M$.

Many software packages support the calculation of Vietoris-Rips persistence on point clouds. In this paper, we use the Eirene package developed by Gregory Henselman [8]. Other popular programs include Ulrich Bauer's Ripser [9] and Vidit Nanda's Perseus [10].

In Section 3.5, we shall use the witness complexes of de Silva and Carlsson [11]. The idea is to model the Delaunay triangulation on a smaller set of points $L \subset X$, called landmarks, in such a way that the topology of the underlying object is well-approximated. Moreover, the definition makes sense
in any metric space, so assume that $X$ is a metric space with distance function $d$ (e.g., $X$ could be a finite point cloud in $\mathbb{R}^{m}$ with the usual Euclidean distance). Choose a subset $L=\left\{\ell_{1}, \ell_{2}, \ldots, \ell_{n}\right\}$ of $X=\left\{x_{1}, x_{2}, \ldots, x_{N}\right\}$ and let $R \geq 0$ be a real number.

The witness complex $W(X, L, R)$ is defined as follows:

- $\quad$ The vertex set of $W(X, L, R)$ is $L$;
- $\quad \ell, \ell^{\prime} \in L$ span an edge if there exists an $x \in X$, called a witness, such that

$$
d(x, \ell), d\left(x, \ell^{\prime}\right) \leq R+\min \left\{d\left(x, \ell^{\prime \prime}\right): \ell^{\prime \prime} \in L-\left\{\ell, \ell^{\prime}\right\}\right\}
$$

- A collection $\ell_{0}, \ldots, \ell_{p} \in L$ spans a $p$-simplex if $\left\{\ell_{i}, \ell_{j}\right\}$ span an edge for all $i \neq j$.

Examples of witness complexes are shown in Figure 2b alongside the associated Delaunay triangulation. Four landmark points were chosen using the maxmin procedure described below. The complex on the left has $R=0.0329$, and the complex on the right has $R=0.1317$. Note that the larger value of $R$ yields a complex with more simplices. In addition, note that the witness complex is a coarse approximation of the Delaunay triangulation.

We make some observations about this definition. Let $D$ be the $n \times N$ matrix of distances from points in $L$ to points in $X$ :

- If $R=0$, then $\ell, \ell^{\prime} \in L$ form an edge if there is an $x_{i} \in X$ such that $d\left(x_{i}, \ell\right)$ and $d\left(x_{i}, \ell^{\prime}\right)$ are the two smallest entries in the $i$-th column of $D$. This is analogous to the existence of an edge in the Delaunay triangulation $\operatorname{Del}(L)$.
- For $R>0$, one may think of relaxing the boundaries of the Voronoi diagram of $L$ and taking the nerve of the resulting covering of $X$.
- If $0 \leq R<R^{\prime}$, then there is an inclusion of simplicial complexes $W(X, L, R) \subseteq W\left(X, L, R^{\prime}\right)$.

By a theorem of de Silva and Carlsson [11], this complex is a natural analogue of the Delaunay triangulation for a space represented by point cloud data.

Suppose that $X$ is a sample of points from some object $M \subset \mathbb{R}^{m}$. There is no guarantee that $W(X, L, R)$ recovers the topology of $M$, but experiments on familiar geometric objects [11] (spheres, for example) suggest that, for a suitable range of values of $R$ and good choices of landmarks $L$, the topology of $W(X, L, R)$ is the same as that of $M$. This begs the question:

1. How should the landmark set $L$ be chosen?
2. What is the correct value of $R$ ?

The second question is best handled via the use of persistent homology, which we discussed in Section 2.3 above. As for the choice of landmarks, there are three standard options:

1. Select landmarks at random.
2. Use the maxmin procedure: Choose a seed $\ell_{1}$ at random. Then, if $\ell_{1}, \ldots, \ell_{n}$ have been chosen, let $\ell_{n+1} \in X-\left\{\ell_{1}, \ldots, \ell_{n}\right\}$ be the point which maximizes the function

$$
z \mapsto \min \left\{d\left(z, \ell_{1}\right), d\left(z, \ell_{2}\right), \ldots, d\left(z, \ell_{n}\right)\right\}
$$

3. Use a density-based strategy.

The maxmin procedure yields more evenly-spaced landmarks, but tends to emphasize extremal points. It is generally more reliable than a random selection [11]. Another useful resource is [12]. In our experiments in Section 3.5 below, we use the maxmin process to generate landmarks.

### 2.5. Sampling Procedures

To build a Vietoris-Rips or witness complex on points in $G_{k}\left(\mathbb{R}^{n}\right)$, we need to develop a sampling procedure. The first question to be asked is in which Euclidean space do we embed $G_{k}\left(\mathbb{R}^{n}\right)$ ? This is highly nontrivial. Even in the case of projective spaces $(k=1)$, it is not so obvious how to proceed. A whole industry has been devoted to the question of the minimal embedding dimension of $\mathbb{R} P^{n}$ [13], but the proof of the minimality of any particular embedding rarely comes with an explicit formula for the map. An exception is if one insists on an isometric embedding [14], but the minimal dimension of such an embedding for $\mathbb{R} P^{n}$ is $n(n+3) / 2$, which grows rather quickly.

For arbitrary Grassmannians, one could try to use the Plücker embedding $G_{k}\left(\mathbb{R}^{n}\right) \rightarrow$ $P\left(\bigwedge^{k}\left(\mathbb{R}^{n}\right)\right)=\mathbb{R} P^{\binom{n}{k}-1}$ defined by

$$
\left(x_{1}, \ldots, x_{k}\right) \mapsto\left[x_{1} \wedge \cdots \wedge x_{k}\right]
$$

(where $[v]$ denotes the line spanned by the vector $v$ ) and then embed the target projective space into Euclidean space. Of course, this explodes the dimension further, making this an impractical solution. Aside from some low dimensional projective spaces, we will instead approach this problem via the following result.

Proposition 2. The manifold $G_{k}\left(\mathbb{R}^{n}\right)$ is diffeomorphic to the smooth manifold consisting of all $n \times n$ symmetric, idempotent matrices of trace $k$. The map $\varphi$ realizing this takes a $k$-plane $X$ to the operator defined by orthogonal projection onto $X$.

Proof. If $X$ is a $k$-plane with orthonormal basis $x_{1}, \ldots, x_{k}$, denote by $A$ the $n \times k$ matrix having the $x_{i}$ as columns. Define a map $\varphi: G_{k}\left(\mathbb{R}^{n}\right) \rightarrow M_{n}(\mathbb{R})$ by $X \mapsto A A^{T}$. This map is clearly smooth since it consists of polynomials in the entries of the various $x_{i}$. Moreover, it is well-defined since, if $y_{1}, \ldots, y_{k}$ is another orthonormal basis of $X$ with associated matrix $B$, then there is an orthogonal matrix $O$ such that $B=A O$. Then, $B B^{T}=(A O)(A O)^{T}=A O O^{T} A^{T}=A A^{T}$. The matrix $A A^{T}$ is symmetric: $\left(A A^{T}\right)^{T}=\left(A^{T}\right)^{T} A^{T}=A A^{T}$. It is idempotent: $\left(A A^{T}\right)^{2}=A A^{T} A A^{T}=A I_{k} A^{T}=A A^{T}$ (note that $A^{T} A=I_{k}$, the $k \times k$ identity matrix, since the columns of $A$ are orthonormal). Finally, the trace of $A A^{T}$ is $k$ since its rank is $k$ and its only eigenvalues are 0 and 1 . Thus, the image of $\varphi$ lies in the set of symmetric, idempotent matrices of trace $k$. To see that $\varphi$ surjects onto this set, note that such a matrix $B$ is projection onto a $k$-dimensional subspace $X$ and there exists a basis $x_{1}, \ldots, x_{k}$ with $\varphi(X)=B$. Injectivity of $\varphi$ follows since the subspace determined by a projection is unique.

Now, to generate a sample of points on which to build a Vietoris-Rips or witness complex, we will use the embedding $\varphi$. A crude sampling is then obtained by the following procedure:

- $\quad$ Select $k$ random vectors in $\mathbb{R}^{n}$.
- Perform the Gram-Schmidt orthogonalization algorithm to yield an orthornomal set $x_{1}, \ldots, x_{k}$. Let $A$ be the matrix with $x_{i}$ as columns.
- Compute $A A^{T}$.

One immediate problem with this process is that the $k$-plane it constructs lives in the top-dimensional Schubert cell with probability 1. However, since we know the space we are interested in, and we know its homology, we can bias our sample to ensure we include points from each Schubert cell. The following procedure implements this idea:

- Determine the percentage of sample points desired from each Schubert cell. For example, one might choose $5 \%$ from a 1 -cell, $10 \%$ from a 2 -cell, and so on.
- Elements of a given Schubert cell correspond to the column space of a particular matrix form. Generate such a matrix $B$ using random vectors of the required form.
- Generate a random $n \times n$ orthogonal matrix $X$.
- Add the matrix $A=X\left(B B^{T}\right) X^{T}$ to the point cloud.

Note the final step above. If we merely took the matrix $B$, we would not end up with a well-distributed sample. For example, in the case of $G_{2}\left(\mathbb{R}^{4}\right)$, such a matrix lying in the 1-cell of the Schubert decomposition has the following form:

$$
B=\left[\begin{array}{ll}
1 & * \\
0 & 1 \\
0 & 0 \\
0 & 0
\end{array}\right]
$$

The corresponding point in $\mathbb{R}^{16}$ would have most coordinates equal to 0 , which is clearly not what we want. Conjugating the various $B B^{T}$ by a random orthogonal matrix $X$ (a different $X$ for each $B$ ) yields a wider distribution of points in $G_{k}\left(\mathbb{R}^{n}\right)$.

The MATLAB files we used to generate samples in various projective spaces and Grassmannians are available at https:/ / github.com/niveknosdunk/grassmann.

### 2.6. Approximate Triangulations

We are now ready to search for simplicial complexes modeling the spaces $G_{k}\left(\mathbb{R}^{n}\right)$. The procedure we employ is as follows:

- Construct a sample of points on $G_{k}\left(\mathbb{R}^{n}\right)$.
- Construct a collection of Vietoris-Rips or witness complexes on the point cloud.
- Compute the persistent homology of this filtration.
- Determine a range of parameters where the homology of the complexes agrees with that of $G_{k}\left(\mathbb{R}^{n}\right)$.

Definition 3. Let $K_{r}$ denote either $V R(X, r)$ or $W(X, L, r)$. If there exists a parameter $r>0$ for which the $\mathbb{Z} / 2$ homology of $K_{r}$ agrees with that of $G_{k}\left(\mathbb{R}^{n}\right)$, then we call $K_{r}$ an approximate triangulation of $G_{k}\left(\mathbb{R}^{n}\right)$.

Note that $K_{r}$ is a subcomplex of the Euclidean space in which we have embedded $G_{k}\left(\mathbb{R}^{n}\right)$. However, it does not necessarily lie inside the embedded $G_{k}\left(\mathbb{R}^{n}\right)$. Still, its vertices do lie on $G_{k}\left(\mathbb{R}^{n}\right)$ and so we can think of this as being close to a triangulation of this manifold. For further verification, we can also compute the $\mathbb{Z} / 3$ homology of $K_{r}$ and check it against that of $G_{k}\left(\mathbb{R}^{n}\right)$.

## 3. Results

## 3.1. $\mathbb{R} P^{2}$, Part I

Let us begin by embedding $\mathbb{R} P^{2}$ into $\mathbb{R}^{4}$ using the map $\psi: S^{2} \rightarrow \mathbb{R}^{4}$ defined by

$$
\psi:(x, y, z) \mapsto\left(x y, x z, y^{2}-z^{2}, 2 y z\right)
$$

Note that $\psi(-x,-y,-z)=\psi(x, y, z)$ and so it descends to a map $\mathbb{R} P^{2} \rightarrow \mathbb{R}^{4}$. Generate a sample of 100 points on $S^{2}$ and then use this map to get the points in $\mathbb{R}^{4}$. The persistence diagrams are shown in Figure 3. There is a tiny window, around $r=0.87$, where we get the correct homology.


Figure 3. Vietoris-Rips persistence diagrams for 100 points on $\mathbb{R} P^{2}$ (a) $H_{1}$ persistence and (b) $\mathrm{H}_{2}$ persistence.

Now generate a sample of 200 points. As expected, the Vietoris-Rips complex has the correct homology for a longer range of parameters, as indicated in Figure 4. Here, we see a long interval $0.69<r<0.87$ where we get the correct homology. Thus, the Vietoris-Rips complex built on these 200 points in $\mathbb{R}^{4}$ is a good approximation to $\mathbb{R} P^{2}$.


Figure 4. Vietoris-Rips persistence diagrams for 200 points on $\mathbb{R} P^{2}$ (a) $H_{1}$ persistence and (b) $\mathrm{H}_{2}$ persistence.
3.2. $\mathbb{R} P^{2}$, Part II

The embedding of $\mathbb{R} P^{2}$ into $\mathbb{R}^{4}$ is not an isometric embedding, though. For that, we need $\mathbb{R}^{5}$ :

$$
(x, y, z) \mapsto\left(y z, x z, x y, \frac{1}{2}\left(x^{2}-y^{2}\right), \frac{1}{2 \sqrt{3}}\left(x^{2}+y^{2}-2 z^{2}\right)\right)
$$

If we then generate 100 random points on this surface, we obtain the Vietoris-Rips barcodes in Figure 5. This works better than the embedding into $\mathbb{R}^{4}$; we get the correct answer for $0.625<r<0.871$. The result for 200 points is even better, and is shown in Figure 6.


Figure 5. Vietoris-Rips persistence diagrams for 100 points on $\mathbb{R} P^{2}$, using the isometric embedding into $\mathbb{R}^{5}$ (a) $H_{1}$ persistence and (b) $H_{2}$ persistence.


Figure 6. Vietoris-Rips persistence diagrams for 200 points on $\mathbb{R} P^{2}$, using the isometric embedding into $\mathbb{R}^{5}$ (a) $H_{1}$ persistence and (b) $H_{2}$ persistence.

## 3.3. $\mathbb{R} P^{3}$

We use the fact that $\mathbb{R} P^{3}$ is diffeomorphic to $S O(3)$, the space of $3 \times 3$ orthogonal matrices of determinant 1 . If we select 100 random points on this space in $\mathbb{R}^{9}$, we find that there is only a tiny window where $\beta_{2}=1$, so 100 points probably is not enough to yield a good approximate triangulation. The barcodes are shown in Figures 7 and 8.


Figure 7. Vietoris-Rips persistence diagrams for 100 points on $\mathbb{R} P^{3}$, realizing it as the Lie group $S O(3) \subset \mathbb{R}^{9}(\mathbf{a}) H_{1}$ persistence and (b) $H_{2}$ persistence.


Figure 8. The $H_{3}$ barcode for 100 points on $\mathbb{R} P^{3}$.
If we now sample 200 points at random on $\mathbb{R} P^{3}$ (computation time 6:54), we obtain the barcodes in Figures 9 and 10. Note that we get the correct homology for $2.1<r<2.4$.


Figure 9. Vietoris-Rips persistence diagrams for 200 points on $\mathbb{R} P^{3}$ (a) $H_{1}$ persistence and (b) $\mathrm{H}_{2}$ persistence.


Figure 10. The $H_{3}$ barcode for 200 points on $\mathbb{R} P^{3}$.

## 3.4. $G_{2}\left(\mathbb{R}^{4}\right)$, Part I

We now consider the first Grassmannian that is not a projective space. Embed the 4-manifold $G_{2}\left(\mathbb{R}^{4}\right)$ as the space of symmetric idempotent $4 \times 4$ matrices of trace 2 . As a first attempt, we take the naïve sampling approach of generating random pairs of orthonormal vectors to build a point cloud of such matrices. However, persistence calculations now become rather cumbersome. Table 1 shows some statistics on computation times for point clouds of various sizes on a MacBook Pro, 16 GB RAM.

Table 1. Computation times on a MacBook Pro, 16 GB RAM, of Vietoris-Rips persistence for various spaces. An $X$ indicates that the software could not complete the calculation; a indicates the software was not used for the run. The ? indicates the number of simplices is unknown. Note the rapid explosion in the number of simplices.

| Space | \# Points | Top Dim | \# Simplices | Eirene | Ripser |
| :---: | :---: | :---: | :---: | :---: | :---: |
| $\mathbb{R} P^{2} \subset \mathbb{R}^{4}$ | 100 | 2 | 206 K | $0: 00.53$ | - |
|  | 200 | 2 | 2.1 M | $0: 01$ | - |
| $\mathbb{R} P^{2} \subset \mathbb{R}^{5}$ | 100 | 2 | 436 K | $0: 02$ | - |
|  | 200 | 2 | 4.2 M | $0: 07$ | - |
| $\mathbb{R} P^{3} \subset \mathbb{R}^{9}$ | 100 | 3 | 7.6 M | $0: 09$ | - |
|  | 200 | 3 | 146 M | $6: 54$ | - |
| $G_{2}\left(\mathbb{R}^{4}\right) \subset \mathbb{R}^{16}$ | 100 | 4 | 107 M | $1: 51$ | $1: 15$ |
|  | 150 | 4 | 792 M | $1: 04: 45$ | X |
|  | 200 | 3 | 112 M | $3: 01$ | $3: 07$ |
|  | 200 | 4 | $?$ | $X$ | $X$ |

Eirene could compute homology for 200 points up to dimension 3 in about 3 min , producing a parameter value of $r=0.95$ where the homology is correct in these dimensions. It seems that $H_{4}$ is the sticking point. The barcodes for 150 points are shown in Figures 11 and 12. At $r=0.96$, the homology is correct up to dimension 3 , but $H_{4}=0$ there.


Figure 11. Vietoris-Rips persistence diagrams for 150 points on $G_{2}\left(\mathbb{R}^{4}\right)$ (a) $H_{1}$ persistence and (b) $\mathrm{H}_{2}$ persistence.


Figure 12. Vietoris-Rips persistence diagrams for 150 points on $G_{2}\left(\mathbb{R}^{4}\right)$ (a) $H_{3}$ persistence and (b) $H_{4}$ persistence.

In a quest for more memory, we received an offer from Mikael Vejdemo-Johannson to use his machine. It has 256 GB RAM. We began the 200 point Vietoris-Rips calculation in Eirene in the background and logged out. After 10 hours, it was still processing and was using $97 \%$ of the system memory. The next morning the process was complete; the output file (in JLD2 format) was 74 GB (!). Since Eirene uses PlotlyJS to render barcodes, they cannot be viewed remotely. Even if the file could be retrieved, it is unclear that our laptop could even open it, nor is there any guarantee that the barcodes are correct.

## 3.5. $G_{2}\left(\mathbb{R}^{4}\right)$, Part II

We then took a different approach. The Vietoris-Rips complex is nice because it is easy to compute, but it suffers from combinatorial explosion. We turned to witness complexes and made the associated computations using the Javaplex package [15] in MATLAB.

The initial attempt simply generated elements of $G_{2}\left(\mathbb{R}^{4}\right)$ by taking a pair of orthonormal vectors in $\mathbb{R}^{4}$ and using them to build a certain $4 \times 4$ matrix. For this experiment, we biased the sample in the following way. For a given number $M$ of points on $G_{2}\left(\mathbb{R}^{4}\right)$, we took $5 \%$ from the 1-cell, $15 \%$ from each of the 2 -cells, $25 \%$ from the 3 -cell, and $40 \%$ from the 4 -cell. One could choose different proportions, of course.

This worked remarkably well. We generated 5000 points on $G_{2}\left(\mathbb{R}^{4}\right)$ and constructed the witness complex on 100 landmarks chosen using the max-min process. The barcodes for one such trial are shown in Figure 13. Note that we get the correct mod 2 and mod 3 homology for $r>0.125$. This witness complex, which has 118,220 simplices, is therefore a good approximate triangulation of $G_{2}\left(\mathbb{R}^{4}\right)$. The point cloud and witness points are available as text files at https:/ / github.com/niveknosdunk/ grassmann. Note that the number of simplices in this witness complex is three orders of magnitude smaller than a Vietoris-Rips complex on the same number of points (cf. Table 1), so this construction is more effective all the way around.

(a)

Figure 13. Cont.

(b)

Figure 13. Barcodes for a witness complex on 100 points in a 5000 -point sample on $G_{2}\left(\mathbb{R}^{4}\right)$ (a) $\mathbb{Z} / 2$ coefficients and (b) $\mathbb{Z} / 3$ coefficients.

## 4. Conclusions

In this paper, we demonstrated the utility of using Vietoris-Rips and witness complexes to obtain approximate triangulations of the Grassmann manifolds $G_{k}\left(\mathbb{R}^{n}\right)$. We were able to construct such spaces with relatively few vertices, but some questions remain for further study.

1. How small of a sample can we use to generate an approximate triangulation? For example, a result in [3] asserts that any triangulation of $G_{2}\left(\mathbb{R}^{4}\right)$ must have at least 14 vertices. We built an approximate triangulation using a witness complex on 100 landmarks. Surely, our algorithm will not work with only 14 points, but we plan to investigate how few we can get away with. A theorem of Niyogi-Smale-Weinberger [16] provides lower bounds on the number of points required to compute homology correctly with high probability, but these are certainly too high and can be improved in practice.
2. Can we push the computations further? The next Grassmannian to study is $G_{2}\left(\mathbb{R}^{5}\right)$. This is a nonorientable 6-manifold, and, using our procedure, we would embed it in $\mathbb{R}^{25}$. The machine used to compute the persistent homology of the witness complexes on $G_{2}\left(\mathbb{R}^{4}\right)$ in MATLAB ran out of memory on 100 landmarks in $G_{2}\left(\mathbb{R}^{5}\right)$. We therefore need either a bigger machine running MATLAB, or software that can handle witness complexes. The GUDHI package [17] is one option, but we have not attempted it yet.
3. The author expects to gain access to a new GPU based supercomputer at his institution in the next year. This may allow for similar computations on higher-dimensional $G_{k}\left(\mathbb{R}^{n}\right)$.

Funding: This research received no external funding.
Acknowledgments: This problem was suggested to me by Vidit Nanda; I thank him for the inspiration and helpful conversations. Henry Adams provided useful tips for Javaplex. I am also grateful to Mikael Vejdemo-Johansson for the use of his rather powerful computer. Marc Lange had a good suggestion about the proof of Proposition 2. Finally, thanks are due to a pair of anonymous referees for many useful comments.

Conflicts of Interest: The author declares no conflict of interest.

## References

1. Whitehead, J. On C ${ }^{1}$-complexes. Ann. Math. 1940, 41, 809-824. [CrossRef]
2. Duan, H.; Marzantowicz, W.; Zhao, X. Estimate of number of simplices of triangulations of Lie groups. arXiv 2020, arXiv:2003.13125.
3. Govc, D.; Marzantowicz, W.; Pavešić, P. How many simplices are needed to triangulate a Grassmannian? arXiv 2020, arXiv:2001.08292.
4. Milnor, J.; Stasheff, J. Characteristic Classes; Annals of Mathematics Studies 76; Princeton University Press: Princeton, NJ, USA, 1974.
5. Anderson, L.; Davis, J. There is no tame triangulation of the infinite real Grassmannian. Adv. Appl. Math. 2001, 26, 226-236. [CrossRef]
6. Aanjaneya, M.; Teillaud, M. Triangulating the real projective plane. MACIS Proc. 2007. preprint version. Available online: https:/ / arxiv.org/abs/0709.2831 (accessed on 16 July 2020).
7. Casian, L.; Kodama, Y. On the Cohomology of real Grassmann Manifolds. 2013, preprint. Available online: https: / /arxiv.org/abs / 1309.5520 (accessed on 22 June 2020).
8. Henselman, G. Eirene, a Software Package for Computing Persistent Homology. Available online: https: / / github.com/Eetion/Eirene.jl (accessed on 22 June 2020).
9. Bauer, U. Ripser, a Software Package for Computing Persistent Homology. Available online: https:/ / github. com/Ripser/ripser (accessed on 22 June 2020).
10. Nanda, V. Perseus, the Persistent Homology Software. Available online: http:/ /www.sas.upenn.edu/ $\sim$ vnanda/ perseus (accessed on 22 June 2020).
11. de Silva, V.; Carlsson, G. Topological estimation using witness complexes. In Eurographics Symposium on Point-Based Graphics; A.K. Peters: Wellesley, MA, USA, 2004.
12. Chazal, F.; Oudot, S. Towards persistence-based reconstruction in Euclidean spaces. In Proceedings of the 24th ACM Symposium on Computational Geometry, College Park, MD, USA, 9-11 June 2008; pp. 232-241.
13. Davis, D. Embeddings of real projective spaces. Bol. Soc. Mat. Mex. 1998, 4, 115-122.
14. Zhang, Y. Isometric embeddings of real projective spaces into Euclidean spaces. Differ. Geom. Appl. 2009, 27, 100-103. [CrossRef]
15. JavaPlex Persistent Homology and Topological Data Analysis Library. Available online: http:/ /appliedtopology. github.io/javaplex/ (accessed on 22 June 2020).
16. Niyogi, P.; Smale, S.; Weinberger, S. Finding the homology of submanifolds with high confidence from random samples. Disc. Comp. Geom. 2008, 39, 419-441. [CrossRef]
17. GUDHI: Geometry Understanding in Higher Dimensions. Available online: https:/ / gudhi.inria.fr / (accessed on 22 June 2020).
(C) 2020 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).
