Approximate triangulations of Grassmannians

Kevin Knudson

October 28, 2020
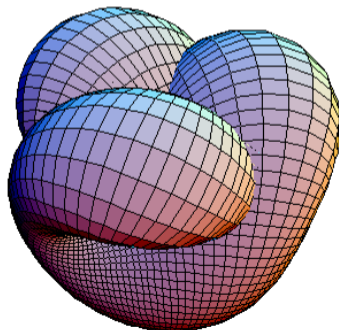
Department of Mathematics
University of Florida
kknudson@ufl.edu
http://kpknudson.com
@niveknosdunk

UF

Outline

Sampling, Triangulation, Persistence (Oh my!)

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

$\mathbb{R}P^2$, Part II

$\mathbb{R}P^3$

$G_2(\mathbb{R}^4)$, Part I

$G_2(\mathbb{R}^4)$, Part II

UF

Sampling, Triangulation, Persistence (Oh my!)

Can you give me an explicit triangulation of the Grassmann manifold $G_k(\mathbb{R}^n)$ of k-planes in $\mathbb{R}^n$?

The real projective plane is well-studied in this regard, but beyond that there doesn't seem to be much out there. Using characteristic classes and cohomological arguments there are estimates on the minimal number of simplices required for a triangulation of $G_k(\mathbb{R}^n)$, but even that is fairly recent (and doesn't produce explicit triangulations).

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 $(a_1, \ldots, a_i, 0, \ldots, 0)$. Any k-plane X gives rise to a sequence of integers

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

Consecutive integers differ by at most 1.

A *Schubert symbol* $\sigma = (\sigma_1, \ldots, \sigma_k)$ is a sequence of k integers satisfying

$$1 \leq \sigma_1 < \sigma_2 < \cdots < \sigma_k \leq n.$$

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

$$\dim(X \cap \mathbb{R}^{\sigma_i}) = i, \dim(X \cap \mathbb{R}^{\sigma_i - 1}) = i - 1.$$

Each $X \in G_k(\mathbb{R}^n)$ belongs to precisely one of the sets $e(\sigma)$.

Fact. $e(\sigma)$ is an open cell of dimension
$d(\sigma) = (\sigma_1 - 1) + (\sigma_2 - 2) + \cdots + (\sigma_k - k)$.

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

$$\begin{bmatrix} * & \cdots & * & 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ * & \cdots & * & * & * & \cdots & * & 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & & & & & & & & & & & & & & \vdots \\ * & \cdots & * & * & * & \cdots & * & * & * & \cdots & * & 1 & 0 & \cdots & 0 \end{bmatrix}$$

where the i-th row has σ_i-th entry positive (say equal to 1) and all subsequent entries zero.

UF

Theorem. The $\binom{n}{k}$ sets $e(\sigma)$ form the cells of a CW-decomposition of $G_k(\mathbb{R}^n)$.

Proposition. The number of r-cells in $G_k(\mathbb{R}^n)$ is equal to the number of partitions of r into at most k integers each of which is $\leq n - k$.

So, for example, the possible Schubert symbols and cells for $G_2(\mathbb{R}^4)$ are as follows. Such a symbol has the form $\sigma = (\sigma_1, \sigma_2)$ where $1 \leq \sigma_1 < \sigma_2 \leq 4$.

σ	$d(\sigma)$
$(1,2)$	0
$(1,3)$	1
$(1,4)$	2
$(2,3)$	2
$(2,4)$	3
$(3,4)$	4

Homology

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

For example,

$$H_i(G_2(\mathbb{R}^4), \mathbb{Z}/2) = \begin{cases} \mathbb{Z}/2 & i = 0 \\ \mathbb{Z}/2 & i = 1 \\ \mathbb{Z}/2 \oplus \mathbb{Z}/2 & i = 2 \\ \mathbb{Z}/2 & i = 3 \\ \mathbb{Z}/2 & i = 4 \end{cases}$$

UF

So, to get a triangulation of $G_k(\mathbb{R}^n)$ we can just "triangulate the Schubert cells," right?

Yeah, OK. That's a cheap answer.

Here's another idea:

- embed $G_k(\mathbb{R}^n)$ into a euclidean space
- sample some points
- build the Vietoris-Rips complex on that sample
- compute persistent homology
- find a parameter where the complex gives you the correct homology

The resulting complex is then what we'll call an *approximate triangulation* of $G_k(\mathbb{R}^n)$.

UF

Lots of questions:

- Which embedding of $G_k(\mathbb{R}^n)$ into which $\mathbb{R}^m$?
- How to sample?
- What's a "Vietoris-Rips" complex?
- Umm, $G_k(\mathbb{R}^n)$ is a manifold of dimension $k(n-k)$. That can get pretty big and most persistence software will run out of memory before computing all the homology, right?
- Also, you only mentioned $\mathbb{Z}/2$-coefficients for homology. How do we know this is good enough?

UF

Some answers:

In theory, real projective spaces can be embedded into fairly small euclidean spaces sometimes. Tracking down explicit formulas for these, however, can be a challenge. If one insists on an *isometric* embedding, there are explicit formulas in the literature, but the minimal dimension for $\mathbb{R}P^n$ is $\frac{n(n+3)}{2}$. That's pretty big.

The Plücker embedding puts $G_k(\mathbb{R}^n)$ into $P(\mathbb{R}^{\binom{n}{k}})$, but then that needs to be embedded into a euclidean space (see first answer).

The Gospel of Milnor and Stasheff tells us that $G_k(\mathbb{R}^n)$ is diffeomorphic to the space of symmetric idempotent $n \times n$ matrices of trace k. This gives an easy way to produce a sample of points on this Grassmannian, but the points then lie in a euclidean space of dimension n^2.

UF

Vietoris-Rips complexes

Given a (finite) metric space X with distance function d, the *Vietoris-Rips complex* with parameter $r \geq 0$ is the simplicial complex $VR(X, r)$ with vertex set X in which $\sigma = [x_0, x_1, \ldots, x_k]$ is a k-simplex if the diameter of the set $\{x_0, x_1, \ldots, x_k\}$ is less than r.

Note that if $r < r'$ we have an embedding $VR(X, r) \subseteq VR(X, r')$ and so we obtain a nested family of complexes $\{VR(X, r)\}_{r \geq 0}$.

UF

Fix a field k. Suppose we have a simplicial complex K and an increasing filtration of K by subcomplexes:

$$K_{R_1} \subset K_{R_2} \subset \cdots \subset K_{R_p} = K$$

where $R_1 < R_2 < \cdots < R_p$.

Then we obtain a sequence of homology groups and linear transformations:

$$H_i(K_{R_1}; k) \to H_i(K_{R_2}; k) \to \cdots H_i(K_{R_p}; k).$$

Say a homology class z is born in $H_i(K_{R_s}; k)$. One of two things happens:

- z maps to 0 in some $H_i(K_{R_t}; k)$, $R_s < R_t$ (that is, the cycle representing z gets filled in); or
- z maps to a nontrivial element in $H_i(K; k)$.

In the latter case, we see that z is a genuine topological feature of the complex K. In the first case, z may or may not be topologically significant, depending on how long the interval $[R_s, R_t]$ is.

In any case, we can summarize what happens for each Betti number by using a *barcode*. This is a collection of intervals defined by

- If z is born at level R_s and dies at level R_t, then we represent it by the interval $[R_s, R_t]$.
- If z is born at level R_s and lives all the way to K, then we represent it by the interval $[R_s, \infty)$.

The associated barcodes:
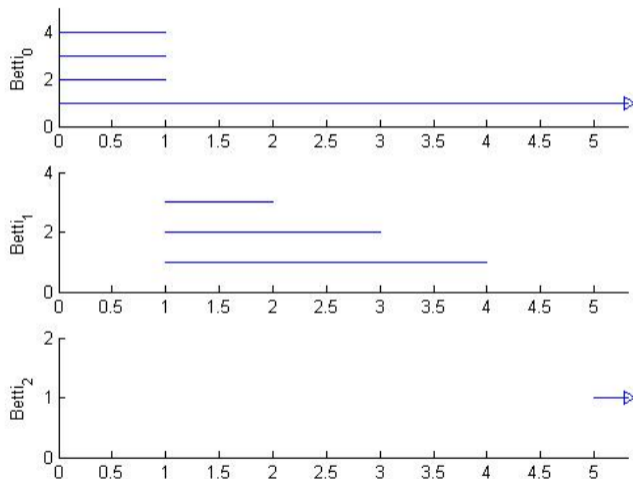
Then there's the homology calculation issue. But let's just forge ahead with some examples.

As for the $\mathbb{Z}/2$-coefficients question, we can also compute $\mathbb{Z}/3$-homology and check that. If it's correct then we can be confident about our approximate triangulation.

Ancillary question: Just how many points do you need?
Niyogi-Smale-Weinberger gives an estimate on the number of points needed to compute the correct homology with high probability. A back of the envelope calculation for $\mathbb{R}P^2$ suggests that one needs approximately 10^7 points to get the homology correct with probability > 0.99. Surely we don't need that many points.

UF

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

Let's embed $\mathbb{R}P^2$ into $\mathbb{R}^4$ using our favorite embedding

$$(x, y, z) \mapsto (xy, xz, y^2 - z^2, 2yz)$$

Generate a sample of 100 points on S^2 and then use this map to get the points in $\mathbb{R}^4$. Compute persistence:
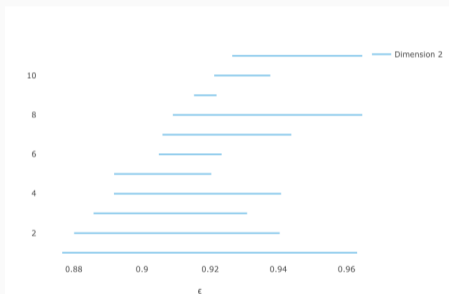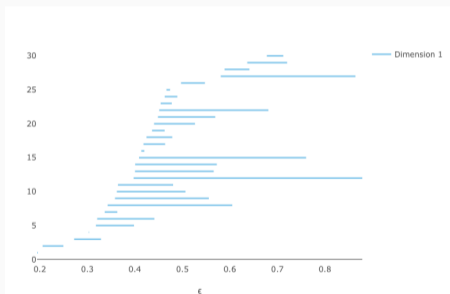
There's a tiny window, around $r = 0.87$ where we get the correct homology.
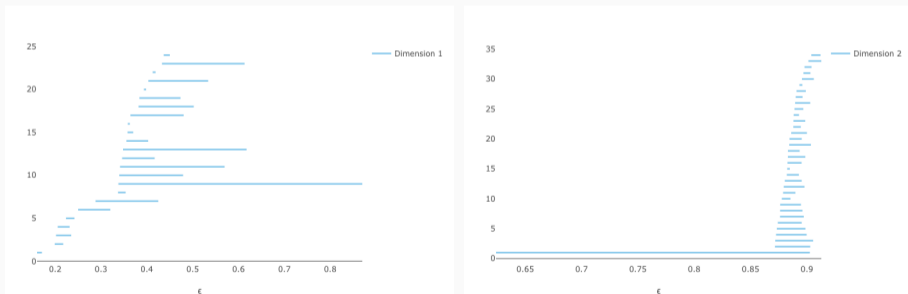
Let's take 200 points.
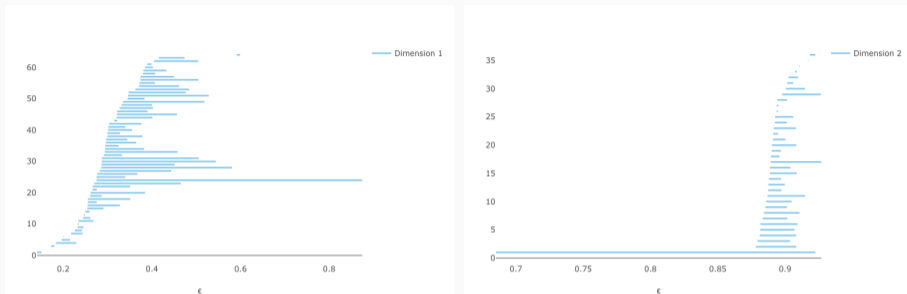
Here we see a long interval $0.69 < r < 0.87$ where we get the correct homology. So the VR complex built on these 200 points in $\mathbb{R}^4$ is a good approximation to $\mathbb{R}P^2$.

$\mathbb{R}P^2$, Part II

The embedding of $\mathbb{R}P^2$ into $\mathbb{R}^4$ is not an isometric embedding, though. For that we need $\mathbb{R}^5$:
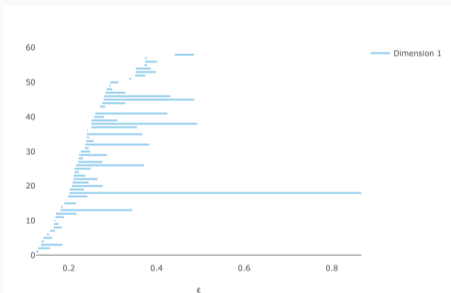
$$(x, y, z) \mapsto \left(yz, xz, xy, \frac{1}{2}(x^2 - y^2), \frac{1}{2\sqrt{3}}(x^2 + y^2 - 2z^2) \right)$$
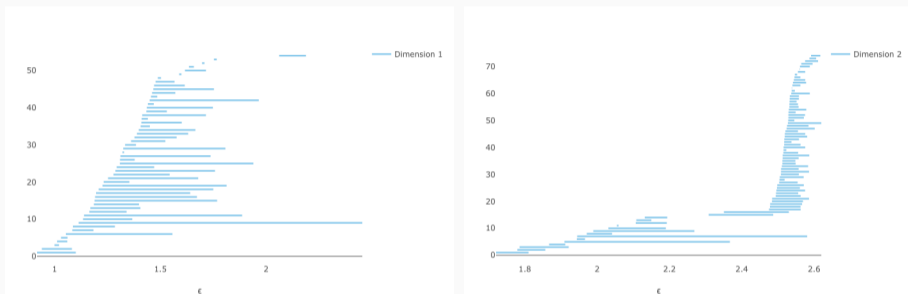
100 random points on this:

This works better. We get the correct answer for $0.625 < r < 0.871$.
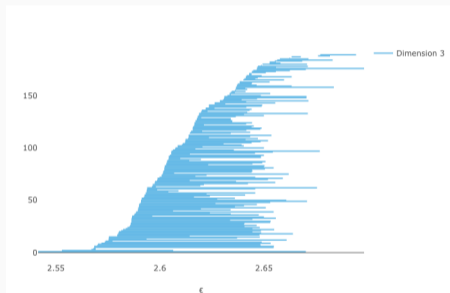
200 random points:
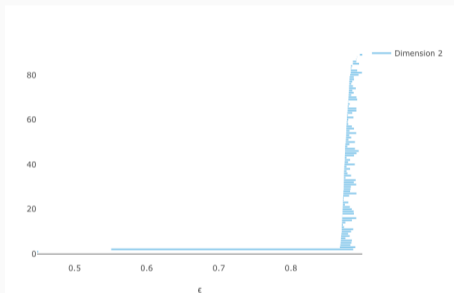
$\mathbb{R}P^3$

Let's use the fact that $\mathbb{R}P^3$ is diffeomorphic to $SO(3)$, the space of 3×3 orthogonal matrices of determinant 1. 100 random points on this space in $\mathbb{R}^9$:
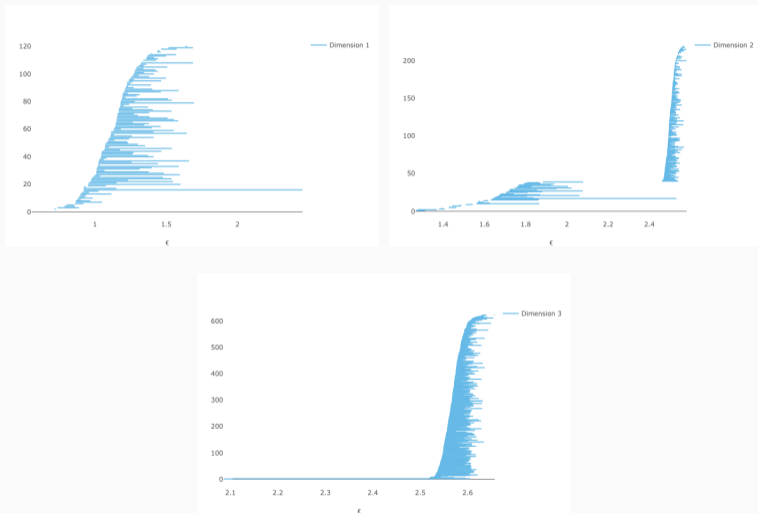

There is only a tiny window where $\beta_2 = 1$, so 100 points probably isn't enough.

Here's the H_3 barcode, just for fun.

200 points at random (computation time 6:54)

Homology correct for $2.1 < r < 2.4$.

$G_2(\mathbb{R}^4)$, **Part I**

Now the real fun begins. Embed $G_2(\mathbb{R}^4)$ as the space of symmetric idempotent 4×4 matrices of trace 2. The sampling procedure is to take a pair of random vectors in $\mathbb{R}^4$, apply Gram-Schmidt, make them the columns of a matrix A and then compute AA^T. This 4×4 matrix has the required properties (exercise).

Persistence calculations now become pretty cumbersome. Here are some statistics on how this goes, on my MacBook Pro, 16GB RAM, computing homology up to dimension 4.

# points	Eirene	Ripser
100	1:51	1:15
150	1:04:45	X
200	X	X

Eirene would compute homology for 200 points up to dimension 3 in about 3 minutes, producing a parameter value of $r = 0.95$ where the homology is correct up to dimension 3. So H_4 seems to be the sticking point.

I tried this both ways–point cloud input and distance matrix input.

NEED MOAR MEMORY!

UF

Here are the barcodes for 150 points:
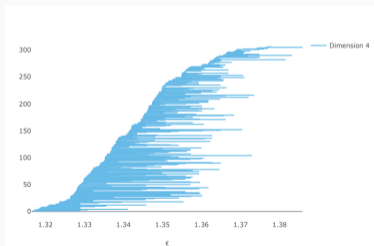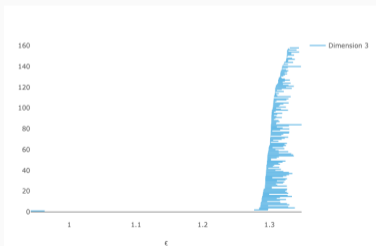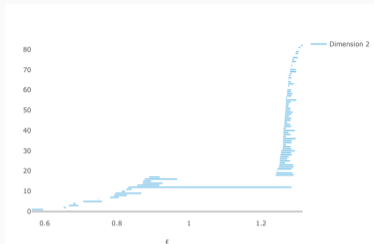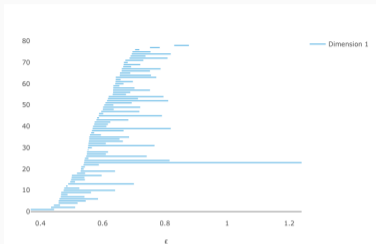
At $r = 0.96$, the homology is correct up to dimension 3. But $H_4 = 0$ there.

In my quest for more memory, I received an offer from Mikael Vejdemo-Johannson to use his machine. It has 256GB RAM. So, I tried the 200 point calculation.

I got it going in the background and logged out. When I logged back in about 10 hours later it was still going and was using 97% of the system memory. I checked back the next morning and the process was complete. The output file (in JLD2 format) was 74 GB (!). Since Eirene uses PlotlyJS to render barcodes, I was unable to view these remotely. I tried fetching the file back to my own machine but couldn't keep a connection long enough (and even if I had gotten it I doubt I could open it locally). And, I suspect the barcodes won't give the correct answer anyway.

UF

$G_2(\mathbb{R}^4)$, **Part II**

We need a different approach. The Vietoris-Rips complex is nice because it's easy to compute, but it suffers from combinatorial explosion. Let's try a different model: witness complexes. I'll skip the formal definition, but given a point cloud X, one chooses a subset $L \subset X$ called *landmarks* and builds a simplicial complex $W(X, L, r)$ ($r \geq 0$) with vertex set L according to a certain rule. Experiments show that the spaces $W(X, L, r)$ approximate the underlying topology quite well (and there are relationships between these spaces and various Rips complexes).

Sampling revisited

The initial attempt simply generated elements of $G_2(\mathbb{R}^4)$ by taking a pair of orthonormal vectors in $\mathbb{R}^4$ and using them to build a certain 4×4 matrix. Is this really the right thing to do? That is, which Schubert cells are likely to be represented in this procedure? (This would be an interesting experiment in its own right.)

Because we're trying to get a good approximation to $G_2(\mathbb{R}^4)$, we are free to bias our sample. So, let's make sure we get representatives from each Schubert cell. And let's sample proportionally according to dimension. That is, let's grab say 5% of the sample from the 1-cell, 10% from each of the 2-cells, and so on.

UF

Explicitly, we know the form of matrices that correspond to each Schubert cell. Generate a random such matrix; call it A. Then take AA^T. But just to make sure that we move things around enough, let's conjugate this by a random 4×4 orthogonal matrix. This should give a pretty good sample of points in $G_2(\mathbb{R}^4)$.

Here's the result. Generate 5000 points on $G_2(\mathbb{R}^4)$ and construct the witness complex on 100 landmarks chosen using the max-min process. Here are the barcodes for one such trial:

UF
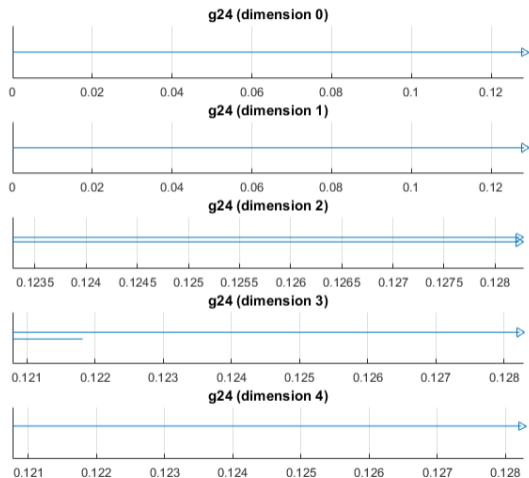
ℤ/2-coefficients

ℤ/3-coefficients
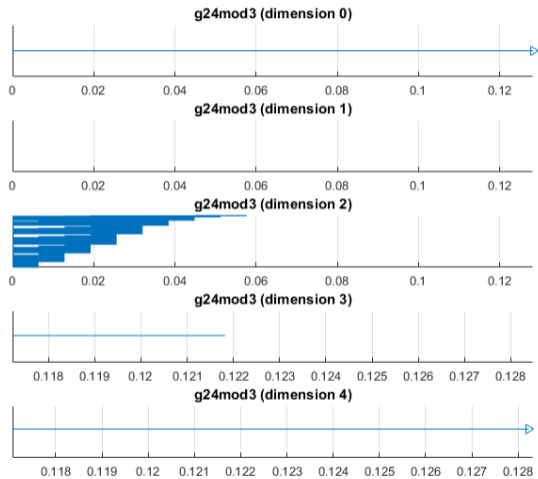
Still to do: more calculations, of course. $G_2(\mathbb{R}^5)$ is already too big for MATLAB on my desktop, so I need to try something else. Stay tuned.

Thanks to Vidit Nanda for suggesting this question to me.