The conditioning number of a manifold

Jānis Lazovskis
December 7, 2016

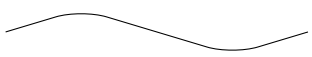
Abstract: We describe the conditioning number τ of a manifold (or an approximation of one) in several different ways, (CN1) through (CN4). The helix is investigated as a motivating example, and attempts are made in Section 2.1 to generalize the methods to higher dimensions. In Section 2.2 we use τ to reconstruct a manifold from a finite point sample on it, and we conclude with applications and connections to real-world scenarios in Section 3.

Note: This survey was prepared as an overview of an independent study with Professor Laura Schaposnik in the Fall 2016 semester, at the University of Illinois at Chicago.

Acknowledgments: Thank you to Professors Laura Schaposnik and Ben Antieau for their advice and guidance on this project, and to Professor Lou Kauffman for helpful suggestions.

Introduction

The *conditioning number* τ of an embedded manifold describes how "smooth" the manifold is. "Smooth" is not in the C^∞-sense, but rather in the sense of looking at the manifold from a distance we don't miss details from up close.



a well-conditioned manifold

not a well-conditioned manifold

Our d-dimensional manifold, for $d \geqslant 1$, will be called M, and will be assumed to be embedded in $\mathbf{R}^n$, for $n \geqslant d$. The particular embedding of M is important, though we will conflate notation for the manifold and

its image under the embedding. Some natural questions to ask regarding τ could be the following:

- Given a manifold, how do we find the conditioning number?
- Given points sampled on the manifold, how do we approximate the conditioning number?
- How does the conditioning number affect the homology of the manifold?

Section 2 will attempt to answer these questions, with varying degress of success. Different descriptions of τ will lend themselves to different ways of generalizing from the $d = 1$ case. A key property of τ we try to attain is *ease of computation*. For instance, the usual description (CN1) is simple, yet difficult to compute.

1 Setting

The manifold M may be "known," in the sense that we are given equations that define it, or we are given a description of points that lie on it. In a more realistic situation, we may only be given a finite number of points sampled on (or even worse, near) the manifold. We always assume the underlying shape is indeed a manifold. Situations, from least to most favorable are:

- Manifold unknown, finitely many points on it given
- Manifold unknown, can sample as many points on it as we want
- Manifold known, can sample as many points on it as we want

The approaches we describe will fall in all of these cases.

1.1 Definitions

normal bundle

Definition 1. The *normal space* of M at $p \in M$ is

$$N_pM := \{v \in \mathbf{R}^n \ : \ v \cdot x = 0 \ \forall \ x \in T_pM\},$$

for $\cdot$ the usual Euclidean inner product. The *normal bundle* of M is $NM = \bigsqcup_{p \in M} N_pM$. The *$\epsilon$-normal bundle* of M is $N^\epsilon M$, where $(N^\epsilon M)_p = N_pM \cap B(p, \epsilon)$.

Given some $\epsilon > 0$, if $N^\epsilon M$ can be embedded in $\mathbf{R}^n$ by extending canonically the embedding of M, then its embedding is called an *ϵ-tubular neighborhood* Tub_ϵ of M. It is natural to ask what is the largest ϵ for which the embedding extends from M.

cond. number

Definition 2. The *conditioning number* (or *reach*, or *injectivity radius*) of M is

$$\tau := \sup_{\substack{\text{embeddings} \\ N^\epsilon M}} \epsilon. \tag{CN1}$$

The tubular neighborhood theorem asserts that for any smooth manifold there always exists an $\epsilon > 0$ for which $N^\epsilon M$ embeds along with M. Unless the manifold is flat (in which case $\tau = \infty$), by starting with a very small ϵ-tubular neighborhood and slowly increasing the radius, we will reach a point where the embedding intersects itself. The points in $\mathbf{R}^n$ where this happens are important to the study of τ.

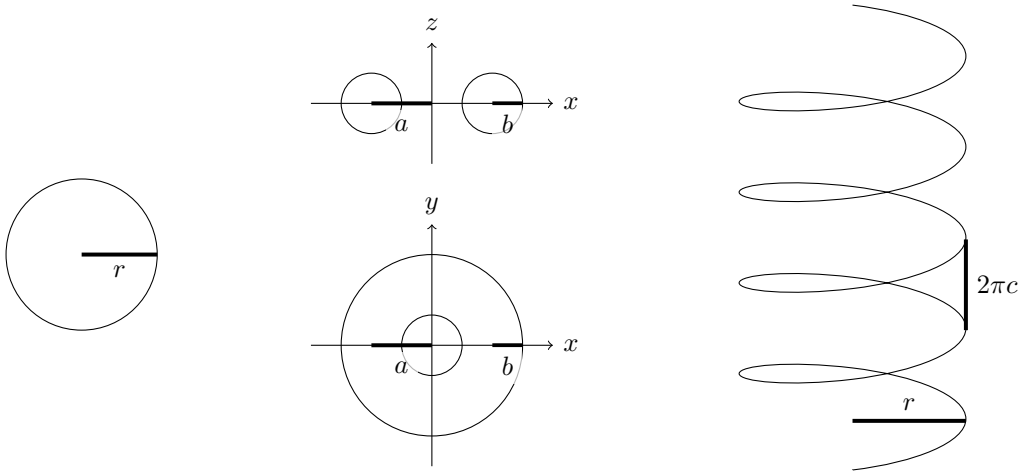
medial axis

Definition 3. The *medial axis* of M is the closure of the set

$$\left\{ x \in \mathbf{R}^n \ : \ \exists \ p \neq q \in M \text{ such that } \inf_{y \in M} \{d(x, y)\} = d(x, p) = d(x, q) \right\}.$$

The medial axis is empty if and only if the manifold is flat (the charts are the identity maps).

1.2 Examples



A circle A torus A helix

The circle: Let $M \subset \mathbf{R}^2$ be a circle of radius r (the zero locus of $x^2 + y^2 - r^2$) centered at the origin of $\mathbf{R}^2$ (so $d = 1$, $n = 2$). It is clear that $\tau = r$. Its medial axis is the single point $(0, 0)$.

The torus: Let $M \subset \mathbf{R}^3$ be the torus of major radius a and minor radius b, that is, the zero locus of $(\sqrt{x^2 + y^2} - a)^2 + z^2 - b^2$ (so $d = 2$, $n = 3$). In this case $\tau = \min\{b, a - b\}$. Its medial axis is the disjoint union of the circle $x^2 + y^2 = a^2$ in the plane $z = 0$ and the z-axis.

The helix: Let $M \subset \mathbf{R}^3$ be the helix of radius r and vertical period $2\pi c$, that is, the zero locus of $x - r\cos(z/c)$ and $y - r\sin(z/c)$ (so $d = 1$, $n = 3$). To find τ, first consider local restrictions from how curved the helix is at a point p. This may be found by taking the limit of the distance to the intersection of the the normal planes at p and q as $q \to p$, which is $\frac{r^2+c^2}{r}$. We must also consider how close the curve comes to itself, for instance the curve may have a small stretch $c \ll r$. By unfolding the helix we find the shortest distance between the helix and itself on the normal planes may be one of two numbers, so the conditioning number is the smallest of these three different values (see Examples 4, 7, 9 for details).

2 Reconstruction

Here we show how to find and use the conditioning number, uncovering new descriptions of it along the way.

2.1 Computing the conditioning number

We begin by assuming we know the manifold under consideration.

2.1.1 Intersections of normal planes

Consider a curve C embedded in $\mathbf{R}^n$ via $n - 1$ independent equations $f_1, \ldots, f_{n-1}$. For an arbitrary curve, we may not have such equations, or we may have more equations, so this is a special case. One approach to finding the conditioning number is to take two points $p, q \in C$, whose embedded normal planes N_pC and N_qC, having codimension 1, must intersect in $\mathbf{R}^n$. Then

$$\tau = \min_{p,q \in C} \left\{ \min_{x \in N_pC \cap N_qC} \left\{ \frac{d(p,x) + d(q,x)}{2} \right\} \right\}. \tag{CN2}$$

The normal planes are described as

$$N_pC = \left\{ x \in \mathbf{R}^n \ : \ \det \begin{bmatrix} x-p \\ df_1(p) \\ \vdots \\ df_{n-1}(p) \end{bmatrix} = 0 \right\}, \qquad N_qC = \left\{ x \in \mathbf{R}^n \ : \ \det \begin{bmatrix} x-q \\ df_1(q) \\ \vdots \\ df_{n-1}(q) \end{bmatrix} = 0 \right\}.$$

Letting $g_p(x), g_q(x) \in \mathbf{R}[x]$ be the linear polynomials describing these planes, their intersection is

$$N_pC \cap N_qC = \{x \in \mathbf{R}^n \ : \ g_p(x) = 0, g_q(x) = 0\} = \{(h_1, h_2, t_1, \ldots, t_{n-2}) \ : \ t_i \in \mathbf{R}\},$$

for $h_1, h_2 \in \mathbf{R}[t_1, \ldots, t_{n-2}]$ linear polynomials. Hence for $x \in N_pC \cap N_qC$, the distance to p and q is given by

$$d(p,x) = \sqrt{(p_1 - h_1(t_1, \ldots, t_{n-2}))^2 + (p_2 - h_2(t_1, \ldots, t_{n-2}))^2 + (p_3 - t_1)^2 + \cdots + (p_n - t_{n-2})^2},$$
$$d(q,x) = \sqrt{(q_1 - h_1(t_1, \ldots, t_{n-2}))^2 + (q_2 - h_2(t_1, \ldots, t_{n-2}))^2 + (q_3 - t_1)^2 + \cdots + (q_n - t_{n-2})^2}.$$

local cond. number

Example 4. Recall the helix example from Section 1.2, for which we compute a special case of (CN2). Instead of taking the minimum over all $p, q \in C$, we take the limit as $q \to p$. This is called the *local conditioning number* τ_p^ℓ of C at p. Let $p_1, p_2 \in C$ and z_1, z_2 the z-values that define p_1 and p_2, respectively. The normal planes at p_i satisfy

$$\det \begin{bmatrix} x - r\cos(z_i/c) & y - r\sin(z_i/c) & z - z_i \\ 1 & 0 & r\sin(z_i/c)/c \\ 0 & 1 & -r\cos(z_i/c)/c \end{bmatrix} = -xr\sin(z_i/c)/c + yr\cos(z_i/c)/c + z - z_i = 0.$$

Points on the line L where the two planes intersect then have coordinates

$$x = \frac{c((z-z_1)\cos(z_2/c) - (z-z_2)\cos(z_1/c))}{r\sin((z_1-z_2)/c)}, \qquad y = \frac{c((z-z_1)\sin(z_2/c) - (z-z_2)\sin(z_1/c))}{r\sin((z_1-z_2)/c)},$$

and z is a free variable. Summing the distances from p_1 and p_2 to a point $(x, y, z) = L_z$ on this line gives us the function we need to minimize. Since the helix is symmetric around its axis, without loss of generality $p_1 = (r, 0, 0)$, so $z_1 = 0$, simplifying the distances to

$$d(p_1, L_z) = \sqrt{z^2\left(1 + \frac{c^2}{r^2}\right) + \left(r + \frac{cz(\cos(z_2/c) - 1) + cz_2}{r\sin(z_2/c)}\right)^2},$$
$$d(p_2, L_z) = \sqrt{(z-z_2)^2 + \left(\frac{cz}{r} + r\sin(z_2/c)\right)^2 + \left(r\cos(z_2/c) + \frac{cz(\cos(z_2/c) - 1) + cz_2}{r\sin(z_2/c)}\right)^2}.$$

Taking the limit of the sum of these distances as $p_2 \to p_1$ or, equivalently, as $z_2 \to 0$, is not too difficult, and we find

$$\frac{1}{2}\lim_{z_2 \to 0}[d(L_z, p_1) + d(L_z, p_2)] = \frac{\sqrt{(c^2 + r^2)(c^2 + r^2 + z^2)}}{r},$$

which attains its minimum at $z = 0$, giving a value $\tau_{p_1}^\ell = (c^2 + r^2)/r$ of the local conditioning number. Note that it does not depend on p_1.

Remark 5. For an arbitrary d-manifold embedded in $\mathbf{R}^n$, the normal planes at any two points, having codimension $n - d$, only intersect when $d = 1$. Hence this approach does not generalize from curves.

Remark 6. The helix example was simple, but in general taking the minimium over an infinite number of points is quite difficult, and the square root function does not make it easier. There are two ways out: one is to use a finite point sample, which we will explore further below; another one is to use a completely different approach, adapted from [GM99], which we consider in the next section.

The definition (CN2) of τ may be simplified a bit. Restrictions on the conditioning number may also come from "local" properties as above, but also from the curve "being close to itself." Increasing the size of the tubular neighborhood T of C, we hit non-local constraints when tangent spaces of ∂T coincide. When tangent spaces are the same, the normals defining them also are the same, meaning that if ∂T is tangent to itself, there exist $p, q \in C$ such that

$$p \in N_q C, \qquad q \in N_p C.$$

non-local cond. number

With this, set

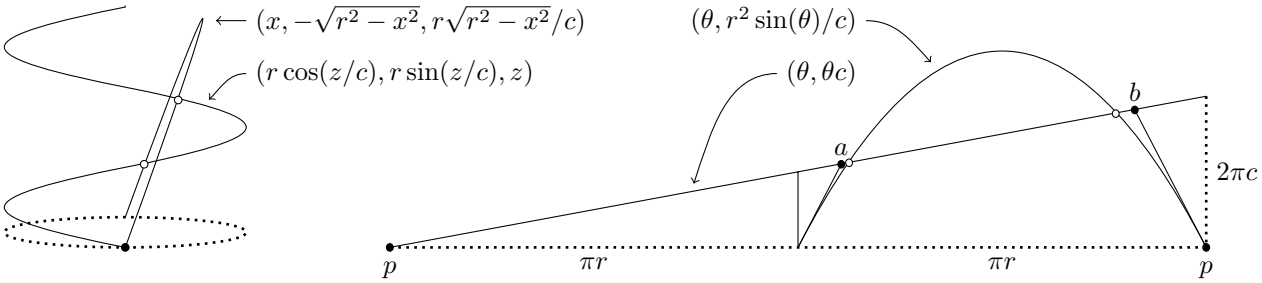
$$\tau_p^{n\ell} := \frac{1}{2} \min_{\substack{q \in N_p C \cap C \\ q \neq p}} d(p,q)$$

to be the *non-local conditioning number* (also *distance of closest approach*) of C at p. It follows immediately that the conditioning number of C may be defined in terms of the local and non-local ones. That is,

$$\tau = \min_{p \in C} \left\{ \tau_p^{\ell}, \tau_p^{n\ell} \right\}. \tag{CN3}$$

We avoid using the term "global," as that implies all restrictions, local and non-local, are satisfied.

Example 7. To find the non-local conditioning number of the helix, we need to find the distance along the normal plane from a point p to the curve. Unfortunately this is quite difficult to solve exactly, but we may make some approximations. Unravel a full period of the helix as below, with parametrizations in rectangular and cylindrical coordinates as given.



As the transformation into cylindrical coordinates shows, solving $\theta c = r^2 \sin(\theta)/c$ is quite difficult, so instead we find when θc intersects the tangent lines of the curve at $z = p_z$. Note this curve is the intersection of the normal plane at p intersecting the cylinder on which the helix lies. Knowing the angle at p is the same as the angles between the verticals and tangent lines, we find the distances from p to the approximations (black dots) a and b of the actual intersection points (white dots) to be

$$d(p,a) = \sqrt{2r^2\left(1 + \cos\left(\frac{\pi c^2}{r^2 - c^2}\right)\right) + \left(\frac{\pi c r^2}{r^2 - c^2}\right)^2}, \quad d(p,b) = \sqrt{2r^2\left(1 - \cos\left(\frac{2\pi c^2}{r^2 + c^2}\right)\right) + \left(\frac{2\pi c r^2}{r^2 + c^2}\right)^2}.$$

Note that $a = b$ when $r = c\sqrt{3}$, so certainly when $c > r/\sqrt{3}$ we do not need to worry about the non-local conditioning number.

2.1.2 Spheres on the manifold

Instead of working from the normal spaces, the conditioning number problem may be restated in terms of d-spheres lying on the embedded d-manifold M in $\mathbf{R}^n$. We again begin with the motivating $d = 1$ case.

curvature

Definition 8. Let $\gamma : \mathbf{R} \to \mathbf{R}^n$ be a smooth curve. The *curvature* of γ at $p = \gamma(t)$ is

$$\kappa(t) = \frac{\sqrt{\left|\frac{d\gamma}{dt}\right|^2 \left|\frac{d^2\gamma}{dt^2}\right|^2 - \left(\frac{d\gamma}{dt} \cdot \frac{d^2\gamma}{dt^2}\right)^3}}{\left|\frac{d\gamma}{dt}\right|^3}. \tag{1}$$

When γ is parametrized by arc length (that is, $\left|\frac{d\gamma}{dt}\right| = 1$), this simplifies to $\kappa(t) = \left|\frac{d^2\gamma}{dt^2}\right|$.

radius of curvature

The reciprocal $\rho(t)$ of the curvature is called is the *radius of curvature*. It may be viewed as the radius of a circle having the same curvature as γ at $\gamma(t)$. Since it is not immediate how to generalize κ into "d-curvature" on the manifold, we take another approach at defining ρ, first on curves. Take $p_1, p_2, p_3 \in C$ and set

$$r(p_1, p_2, p_3) = (\text{radius of circle through } p_1, p_2, p_3).$$

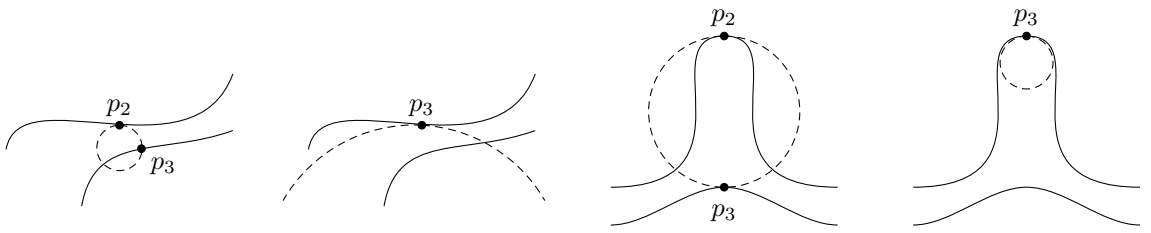
If p_1, p_2, p_3 are colinear, set the value to be ∞. If $\gamma \in \mathbf{R}^3$, then this circle is the great circle of a unique 2-sphere. In general, if $\gamma \in \mathbf{R}^n$, this circle lies on the equator of a unique $(n-1)$-sphere, defined by using the equation (5) to find the center and radius of the circle defined by the three points in $\mathbf{R}^n$ (and a center and radius in $\mathbf{R}^n$ is enough to get an $(n-1)$-sphere).

It is immediately clear that τ is bounded above by the infimum of $r(p_1, p_2, p_3)$, taken over all triples of non-colinear points on C. To see that equality holds, first consider the limits

$$\lim_{p_1 \to p_2} r(p_1, p_2, p_3), \tag{2}$$

$$\lim_{p_1, p_2 \to p_3} r(p_1, p_2, p_3). \tag{3}$$

The circle defined by the first limit lies tangent to M and passes through p_3, though is not necessarily tangent to it there, as it may contain parts of M in its interior. The circle defined by the second limit has radius the local radius of curvature of M at p_3. Below are some examples to illustrate the differences between the two limit types on a curve.



(4)

Radius (2) circle Radius (3) circle Radius (2) circle Radius (3) circle

Hence locally, as in [GM99], the limit of type (3) is $\rho(p_3) = \tau^{\ell}_{p_3}$, and the circle it defines is called the *osculating circle* at p_3. Non-locally, for appropriate p_3, the limit (2) becomes $\tau^{n\ell}_{p_2}$. Indeed, if the $(n-1)$-sphere defined by p_1, p_2, p_3 in the limit (2) contains a part of C inside it, switching p_3 to that part of C gives a strictly smaller circle. Hence this limit gives us the distance of closest approach of M at p_2, which is the non-local conditioning number.

Example 9. Recall again the helix example, with parametrization (and derivatives)

$$\gamma(t) = (r\cos(t/c), r\sin(t/c), t)\,, \quad \frac{d\gamma}{dt} = \left(\frac{-r\sin(t/c)}{c}, \frac{r\cos(t/c)}{c}, 1\right), \quad \frac{d^2\gamma}{dt^2} = \left(\frac{-r\cos(t/c)}{c^2}, \frac{-r\sin(t/c)}{c^2}, 0\right).$$

Applying equation (1) and simplifying, we confirm $\rho = (r^2 + c^2)/r$, irrespective of t (that is, z).

Let us now try to generalize this notion for $d > 1$. Take $d+2$ points $p_1, \ldots, p_{d+2}$ on our manifold M in general position (naturally defining a copy of $\mathbf{R}^{d+1}$ in $\mathbf{R}^n$), for which there is a unique d-sphere through these points. Indeed, if we view the points $p_i = (p_{i,1}, \ldots, p_{i,d+1})$ as lying in the copy of $\mathbf{R}^{d+1}$ with coordinates

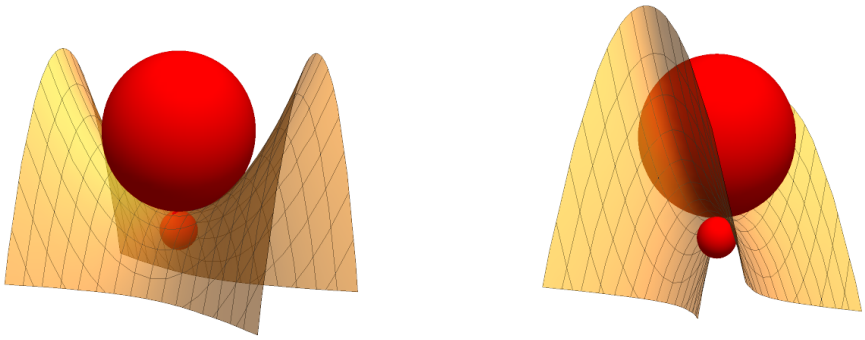
$x_1, \ldots, x_{d+1}$, then the d-sphere is defined (generalizing from [Zwi12], Section 4.8.1) by

$$\det \begin{bmatrix} \sum_{i=1}^{d+1} x_i^2 & x_1 & x_2 & \cdots & x_{d+1} & 1 \\ \sum_{i=1}^{d+1} p_{1,i}^2 & p_{1,1} & p_{1,2} & \cdots & p_{1,d+1} & 1 \\ \sum_{i=1}^{d+1} p_{2,i}^2 & p_{2,1} & p_{2,2} & \cdots & p_{2,d+1} & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \sum_{i=1}^{d+1} p_{d+2,i}^2 & p_{d+2,1} & p_{d+2,2} & \cdots & p_{d+2,d+1} & 1 \end{bmatrix} = 0. \tag{5}$$

As before, let $r(p_1, \ldots, p_{d+2})$ be the radius of this sphere if the points are in general position, and ∞ otherwise. We still have the property if some part of M lies in the interior of this d-sphere, then it may be shrunk. More concretely, if a neighborhood of p_i on M lies in the interior of the sphere, there is some p_i' in this neighborhood such that $r(p_1, \ldots, p_i', \ldots, p_{d+2}) < r(p_1, \ldots, p_i, \ldots, p_{d+2})$. This gives us a third definition

$$\tau = \inf_{p_i \in M} r(p_1, \ldots, p_{d+2}) \tag{CN4}$$

of the conditioning number of M. However, if we try to generalize a limit of the type (2) or (3), we run into some problems. Indeed, consider a 2-manifold with a saddle point, which has two distinct "osculating 2-spheres" at that point, which may have different radii, as in the surface below.



On the other hand, extending the notion of taking limits along paths on M also fails - we can easily construct paths with very sharp turns, yielding a small radius of curvature (a lower bound for τ_p^ℓ). As shown by the saddle point example, we can get also paths with radius of curvature larger than the local conditioning number (that is, an upper bound for τ_p^ℓ). Hence we have no coherent approach.

2.1.3 Finite point sampling

To alleviate the problems in higher dimensions, consider the case where we are given a finite number of points $X = \{p_1, \ldots, p_N\}$ from the manifold M. This also generalizes the previous discussion in the sense that we do not know the equations defining the manifold.

Vietoris Rips complex

Definition 10. Let $\epsilon > 0$. The *Vietoris–Rips complex* V of a finite set $X \subset \mathbf{R}^n$ is a simplicial complex for which a k-tuple of points $\{x_1, \ldots, x_k\}$ defines a $(k-1)$-simplex in V iff $d(x_i, x_j) < \epsilon$ for all $1 \leqslant i, j \leqslant k$.

It is slightly more natural to instead use the *Čech complex*, or *nerve* of radius $\epsilon/2$ of X, which has $\{x_1, \ldots, x_k\}$ defining a $(k-1)$-simplex iff $B(x_i, \epsilon/2) \cap B(x_j, \epsilon/2) \neq \emptyset$ for all i, j. This however is much more difficult to compute, and the two often agree. The conditioning numbers may now be defined as

$$\tau_p^\ell = \min_{(p, p_{i_j}) \in V} r(p, p_{i_1}, \ldots, p_{i_{d+1}}), \qquad \tau = \min_{p_{i_j} \in X} r(p_{i_1}, \ldots, p_{i_{d+3}}).$$

As before, we let $r(p_1, \ldots, p_{d+2}) = \infty$ whenever the points not all distinct or are not in general position. This definition agrees with the (CN4) description in the limit $X \to M$. Since we are taking a minimum over a finite number of points, and the computation is just a polynomial (the determinant), this is quite doable for a computer.

Example 11. Suppose we have $d+2$ points $p_1, \ldots, p_{d+2} \in \mathbf{R}^{d+1}$ in general position on a d-manifold M. Setting $A_{i,j}$ to be the (i,j)-minor of the matrix (determinant of the submatrix when the ith row and jth column are removed) on the left side of equation (5), the radius of this sphere through these $d+2$ points is

$$\sqrt{\frac{(-A_{1,2}/2A_{1,1})^2 + (-A_{1,3}/2A_{1,1})^2 + \cdots + (-A_{1,d+2}/2A_{1,1})^2 - (-1)^{d+4}A_{1,d+3}}{A_{1,1}}} = \sqrt{\sum_{j=2}^{d+2} \frac{A_{1,j}^2}{4A_{1,1}^3} - \frac{(-1)^d A_{1,d+3}}{A_{1,1}}}.$$

This comes from comparing equation (5) to the usual definition $(x_1 - a_1)^2 + \cdots + (x_{d+1} - a_{d+1})^2 = r^2$ of a d-sphere centered at $(a_1, \ldots, a_{d+1})$ with radius r.

2.2 Computing homology

Now suppose that we know τ of M, but not much else. It is possible to calculate the homology of M, within some certainty bound, by sampling points on M.

2.2.1 Procedure

Such an approach is described in [NSW08], which we discuss now. The authors assume we know the d-volume of the d-manifold (length of a 1-manifold, surface area of a 2-manifold, etc), denoted $\mathrm{vol}(M)$.

Algorithm 12.

1. Choose $0 < \epsilon < \tau$ and $\delta \in (0,1)$.
2. Cover M with n-balls of radius $\epsilon/4$.
3. Sample N points (a function of ϵ, τ, δ) on M so every $\epsilon/4$-ball has at least one point.
4. Compute the homology of the Vietois–Rips complex of radius ϵ of the sampled points.

The authors of [NSW08] show there are, with probability $1-\delta$, homotopy equivalences $\mathrm{Tub}_\tau \simeq U \simeq M$, for U the cover of M with n-balls of radius epsilon around the sampled points. Hence the homology of the complex is the homology of M. We also assume the probability distribution on M by which points are sampled is uniform, though adjustments can be made if it is not uniform. The minimums number of points to be sampled is

$$N = \frac{4^d \mathrm{vol}(M)\Gamma(\frac{d+1}{2})}{\epsilon^d(1-(\frac{\epsilon}{8\tau}))^{d/2}2\pi^{(d+1)/2}} \log\left(\frac{8^d \mathrm{vol}(M)\Gamma(\frac{d+1}{2})}{\delta\epsilon^d(1-(\frac{\epsilon}{16\tau}))^{d/2}2\pi^{(d+1)/2}}\right)$$

Giving this bound, the authors claim that $\epsilon < \tau\sqrt{3/5}$, but they incorporate the possibility of points being sampled near the manifold, not necessarily on it, which we do not consider. Now we compute N for some examples, using $\epsilon = \tau$ for simplicity.

2.2.2 Visual and computational examples

Given a finite set X of N points in $\mathbf{R}^n$ and a conditioning number τ, the exact steps to code and actually compute the homology as described above are not immediate. We take the following route:

1. Sample N points randomly on M with Sage, using the `random` package;
2. Construct the Vietoris–Rips complex of X of radius τ with Sage, using the fast algorithm [Zom10];
3. Compute the Betti numbers $h_i = \dim(H_i)$ of the $(d+1)$-skeleton of X with `CHomP` software [Har16].

Step 1 may be skipped if the points are given (not the case in the examples below). If we know that M is a d-manifold, $H_i(M) = H_i(X^{d+1})$ with probability $1-\delta$ for all $0 \leqslant i \leqslant d$, where X^{d+1} is the $(d+1)$-skeleton of the Vietoris–Rips complex of radius τ of X. When running the computations, we only do one run of the procedure described above, due to time and computing power limitations.

Example 13. Let M be the unit sphere centered at the origin in $\mathbf{R}^3$, for which $\tau = 1$ and surface area is $\mathrm{vol}(M) = 4\pi$. Note in the table below that increasing τ has more of an impact than decreasing δ. On the right we have Betti number calculations of X^3 for the $\tau = 1$ case.

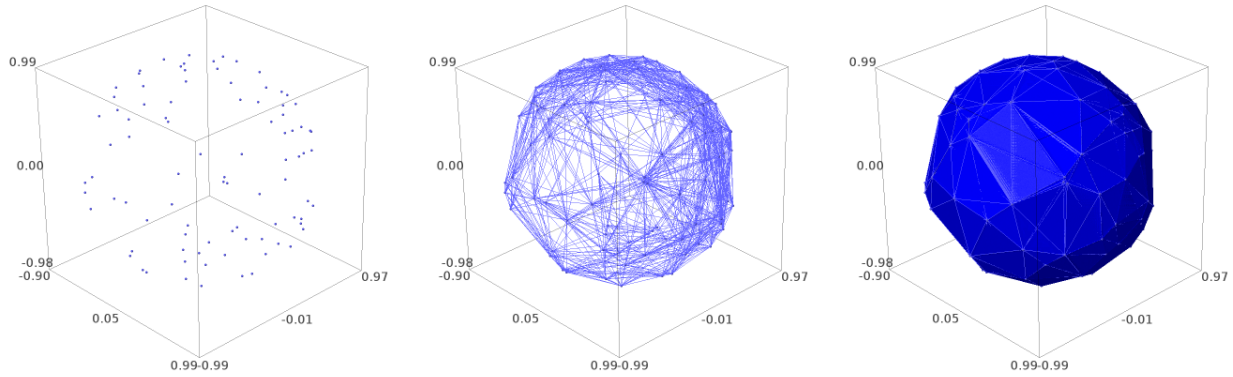
N	$\tau = 1/4$	$\tau = 1/2$	$\tau = 1$
$\delta = 1/2$	2250	462	90
$\delta = 1/10$	2721	579	120
$\delta = 1/100$	3395	748	162
$\delta = 1/1000$	4068	916	204

h_0	h_1	h_2	h_3
1	0	0	13958
1	0	0	38922
1	0	1	107245
1	0	1	305580

To actually sample and plot k points on the sphere, begin with the constant probability distribution on the square $[0,1] \times [0,1] \ni (a,b)$ to get one on the sphere $[0,2\pi] \times [0,\pi] \ni (\theta,\phi)$, via the transformation

$$\theta = 2\pi a, \qquad \phi = \arccos(2b-1).$$

To give an impression of the Vietoris–Rips complex contructed, we include diagrams of the 0-, 1- and 2-skeleta of X with 90 points, the same used to calculate h_i for $\delta = 1/2$ above.



Example 14. Let M be the torus of major radius 2 and minor radius 1 centered at the origin in $\mathbf{R}^3$, for which we already know the conditioning number is $\min\{1, 2-1\} = 1$. The surface area is $\mathrm{vol}(M) = 8\pi^2$, so we get the following table of values N.

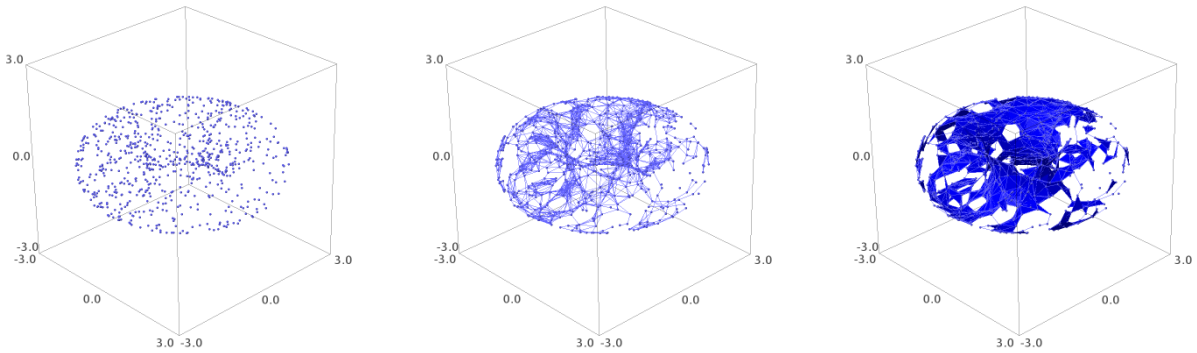
N	$\tau = 1/4$	$\tau = 1/2$	$\tau = 1$
$\delta = 1/2$	17514	3742	777
$\delta = 1/10$	20472	4481	961
$\delta = 1/100$	24705	5540	1226
$\delta = 1/1000$	28938	6598	1491

h_0	h_1	h_2	h_3
1	6	0	799723

Due to the high number of points, (naive) implementation of sampling, and physical constraints, we only manage to calculate the Betti numbers for the $\delta = 1/2$ case. We again use the constant probability distribution on $[0,1]^2$ to get a distribution on the torus via the transformation

$$\theta = 2\pi a, \qquad \phi = 2\pi(\cos(2\pi b)/2 + 1),$$

as described in [DHS13]. The diagrams below are the 777 point case at $\delta = 1/2$.



The diagrams, implementations, and experimental results indicate the sampling on the surfaces may not have been as uniform or efficient as possible. Fine-tuning this process is something we hope to explore in the future.

3 Applications

We conclude with some applications of the conditioning number. These only deal with the 1-dimensional case, but the ideas may be extended to higher dimensions.

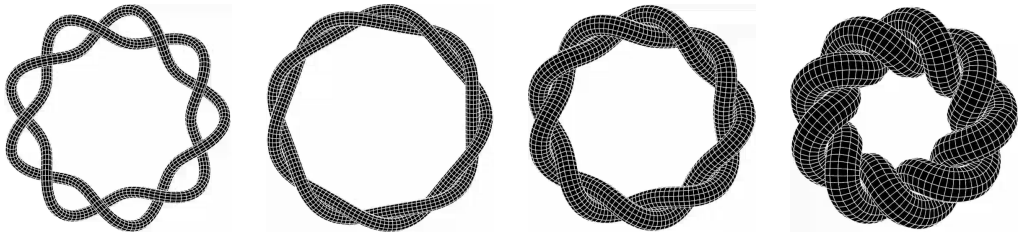
3.1 Knot theory

Recall a *knot* K is an embedding of S^1 into $\mathbf{R}^3$, and we view knots with equivalent (up to ambient isotopy) embeddings to be the same. We may ask which embedding will give the largest τ, for a fixed length ℓ of the embedding. An embedding for which the ratio ℓ/τ, called the *ropelength*, is smallest is called the *ideal shape* of K. This question may also be applied to links, which are embeddings of several circles into $\mathbf{R}^3$.

rope length

Example 15. The unknot has ideal shape the embedding given by $x^2 + y^2 = r^2$ in the plane $z = 0$ in $\mathbf{R}^3$. We already know $\tau = r$ for the circle, so the smallest ropelength of the unknot is $2\pi r/r = 2\pi$.

As Theorem 7 in [CKS02] shows, every (finite length) link has a C^1 embedding that minimizes ropelength. However, often this minimization is done by trial and error - by slowly increasing the thickness of an embedding and "wiggling" the knot around (this changes the emebdding, but keeps isotopy class the same) until the it can be made no thicker.

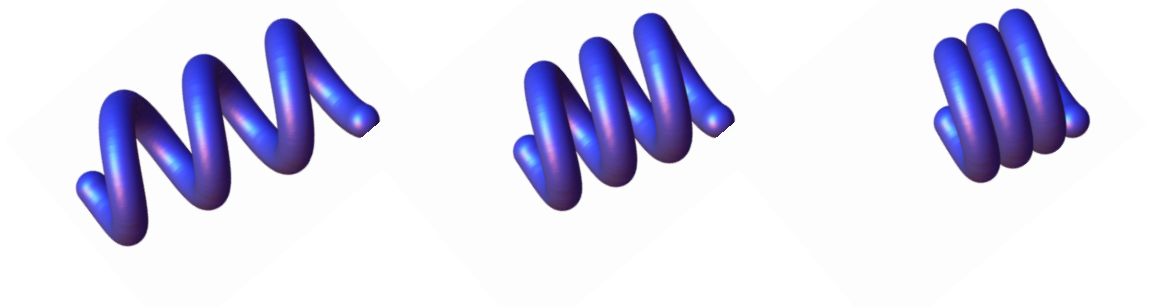


These images, from `jasoncantarella.com/movs/`, show the process of attaining the ideal shape for a knot that lies on a torus.

3.2 Mechanics

The physical distance from a manifold to itself affects properties we may be interested in.

Example 16. Suppose we have a wire of radius τ and length ℓ, and we would like to coil it into a spring. If the spring is to have radius r, the height of the spring, when compressed, will be $\pi r\ell/\tau$.

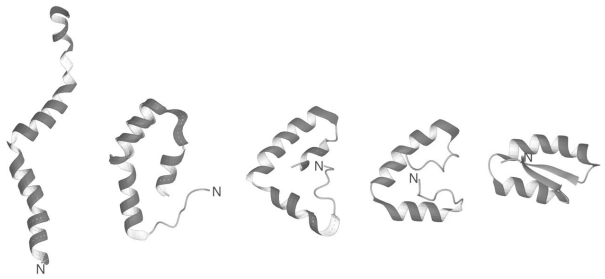


Example 17. Suppose electricity with current I is running through a wire, creating a magnetic field of strength $\mu I/2\pi r$ at a perpendicular distance r from it (μ is a magnetic permeability constant). Then the magnetic field at a point within distance τ from the wire is affected only by the field from the closest point on the wire, and from points whose normal planes contain the point (the magnetic field is a vector quantity).

3.3 Biology

protein folding

A protein is a chain of amino acids and *folding* describes how its physical states change. Some states, "tightest" in a sense, make the protein function properly, while others cause problems. That is, if τ is small, the parts of the protein defining the medial axis at a distance τ will interact more with each other than with any surrounding objects. A diagram of a protein folding is given below.



Given a potential embedding of a protein, knowing τ also indicates if it is not a valid embedding.

References

[CKS02] Jason Cantarella, Robert B. Kusner, and John M. Sullivan. "On the minimum ropelength of knots and links". In: *Invent. Math.* 150.2 (2002), pp. 257–286. ISSN: 0020-9910.

[DHS13] Persi Diaconis, Susan Holmes, and Mehrdad Shahshahani. "Sampling from a Manifold". In: *Advances in Modern Statistical Theory and Applications: A Festschrift in honor of Morris L. Eaton.* Vol. Volume 10. Collections. Beachwood, Ohio, USA: Institute of Mathematical Statistics, 2013, pp. 102–125.

[GM99] Oscar Gonzalez and John H. Maddocks. "Global curvature, thickness, and the ideal shapes of knots". In: *Proc. Natl. Acad. Sci. USA* 96.9 (1999), 4769–4773 (electronic). ISSN: 1091-6490.

[Har16] Shaun Harker. *CHomP – Computational Homology Project software.* `https://github.com/shaunharker/CHomP`. 2016.

[NSW08] Partha Niyogi, Stephen Smale, and Shmuel Weinberger. "Finding the homology of submanifolds with high confidence from random samples". In: *Discrete Comput. Geom.* 39.1-3 (2008), pp. 419–441. ISSN: 0179-5376.

[Zom10] Afra Zomorodian. "Fast construction of the Vietoris-Rips complex". In: *Computers & Graphics* 34.3 (2010). Shape Modelling International (SMI) Conference 2010, pp. 263 –271. ISSN: 0097-8493.

[Zwi12] Daniel Zwillinger. *CRC standard mathematical tables and formulae*. second. CRC Press, Boca Raton, FL, 2012, pp. xiv+819. ISBN: 978-1-4398-3548-7.

Department of Mathematics, Computer Science, and Statistics, University of Illinois at Chicago, Chicago, Illinois, USA

E-mail address: jlazov2@uic.edu