Principal Nested Torii

December 19, 2016

Abstract

Ideas from discussions wtih Andy Wood and Huiling Le, as to how to adapt a Principal Nested Sphere approach to torus spaces.

1 Background

For dimension d, consider data objects $X_1, ..., X_n$ lying in the d-dimensional torus space $\left(S^1\right)^d$, where S^1 is the standard unit circle. The goal is try to find an analog of the Principal Nested Spheres approach of Jung et al (2012) [2]. This may be a special case of backwards PCA, as discussed in Damon and Marron (2014) [1].

The main idea is to find a sequence of approximating torii, that are nested in dimension.

Using the full $\mathbb{R}^{2d}$ embedding of the d-dimensional torus, write

$$\left(S^1\right)^d = \left\{ \begin{pmatrix} \cos\theta_1 \\ \sin\theta_1 \\ \vdots \\ \cos\theta_d \\ \sin\theta_d \end{pmatrix} : \theta_1, \cdots \theta_d \in S^1 \right\}. \tag{1}$$

This is different from the common "donut embedding", as shown in Figure 1. An advantage of a donut embedding is it is easy to visualize in the case $d = 2$. A disadvatage is the different radii entail treating the two angles differently.

The full 2d embedding (1) is also different from the "flat tiled embedding" shown in Figures 2 and 3.

There are two useful properties of the full 2d embedding (1). The first, described in Subsection 1.1, gives a useful indication of where in $\mathbb{R}^{2d}$ the embedded data lie. The second, described in Subsection 1.2, shows that the operation of projection onto the full $\mathbb{R}^{2d}$ embedded d-dimensional torus, in $\mathbb{R}^{2d}$ can be computed as d separate projections onto circles S^1.

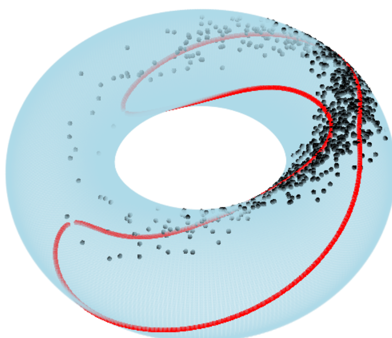

Figure 1: Example of Donut Embedding.

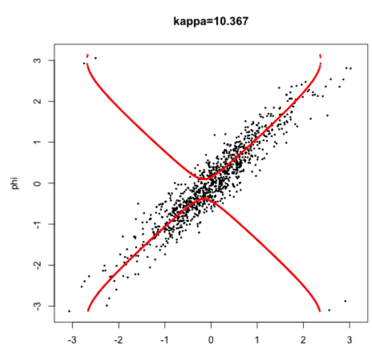

Figure 2: Same Example, shown as flat tile embedding.

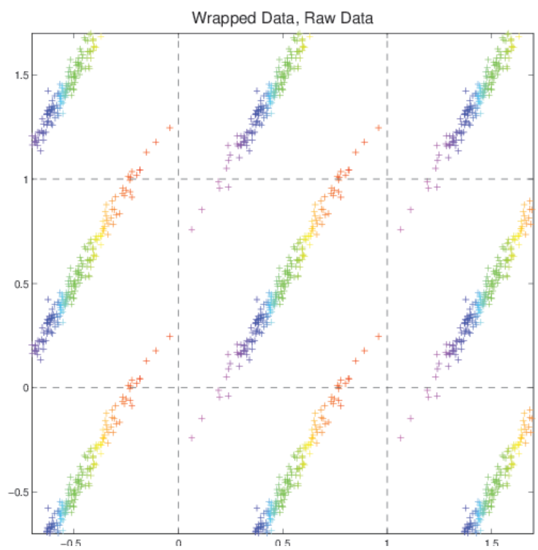

Figure 3: Another flat tile embedding.

1.1 Embedded torus lies in a sphere

Note that the torus is also a subset of (i.e. is embedded in) $\sqrt{d} * S^{2d-1}$ (i.e. the sphere in $\mathbb{R}^{2d}$ of radius $\sqrt{d}$), since

$$\left\| \begin{pmatrix} \cos\theta_1 \\ \sin\theta_1 \\ \vdots \\ \cos\theta_d \\ \sin\theta_d \end{pmatrix} \right\| = \sqrt{\sum_{j=1}^{d} \left((\cos\theta_j)^2 + (\sin\theta_j)^2 \right)} = \sqrt{d}.$$

Hence one approach to PCA on torii might be PNS on the data on this sphere, although that would ignore the toroidal structure of the data.

1.2 Projection in embedded space happens one circle at a time

First simplify notation by taking $d = 2$. Hence consider the projection of an arbitrary point $x \in \mathbb{R}^4$ onto $\left(S^1\right)^2$. This is

$$\arg\min_{t \in (S^1)^2} \|t - x\|^2 = \arg\min_{t \in (S^1)^2} \left[(t_1 - x_1)^2 + (t_2 - x_2)^2 + (t_3 - x_3)^2 + (t_4 - x_4)^2 \right].$$

But $t \in \left(S^1\right)^2$ can be written as $t = \begin{pmatrix} \cos\theta_1 \\ \sin\theta_1 \\ \cos\theta_2 \\ \sin\theta_2 \end{pmatrix}$, so the projection can be rewritten as

$$\arg\min_{t\in(S^1)^2} \|t - x\|^2 = \arg\min_{\theta_1,\theta_2} \left[(\cos\theta_1 - x_1)^2 + (\sin\theta_1 - x_2)^2 + (\cos\theta_2 - x_3)^2 + (\sin\theta_2 - x_4)^2\right] =$$

$$= \arg\min_{\theta_1} \left[(\cos\theta_1 - x_1)^2 + (\sin\theta_1 - x_2)^2\right] + \arg\min_{\theta_2} \left[(\cos\theta_2 - x_3)^2 + (\sin\theta_2 - x_4)^2\right].$$

This shows that projection can be computed essentially "one S^1 at a time". This easily generalizes to the case of $d > 2$, and motivates thinking carefully about decomposition of 2 dimensional subspaces of $\mathbb{R}^{2d}$, i.e. a type of "great torii" of $\left(S^1\right)^d$, composed of d "great circles".

2 Principal Nested Torii

These ideas motivate a new analog of PCA in torii spaces, which is loosely motivated by Principle Nested Spheres. This is essentially a backwards PCA approach, where at each step data on the torus S^d is decomposed into *main components on* S^{d-1} and *small residuals on* S^1. The residuals over all steps provide the resulting set of scores as angles on S^1, which then gives another S^d toroidal representation of the data. This torus representation attempts to "concentrate variation as well as possible in the first few components", just as PCA provides a representation of Euclidean data with the same goal. It can be Euclideanized by finding splits (cut loci) of each S^1 component separately (e. g. at maximal gaps), which are now hopefully oriented in a way that makes this process more sensible.

2.1 The Two Dimensional Case

Again simplify notation by taking $d = 2$. An extrinsic approach to finding the small S^1 residual component is based on finding a rank two projection matrix P, on $\mathbb{R}^4$, (thus $P^2 = P$, $tr(P) = 2$ and there is a two dimensional orthonormal basis $\{u, v\} \subseteq \mathbb{R}^4$ (i.e. $\|u\| = \|v\| = 1$ and $u^t v = 0$) so that $Px = (x^t u)\, u + (x^t v)\, v$. Note that in $\mathbb{R}^4$, $\|Px\| = \sqrt{(x^t u)^2 + (x^t v)^2}$. Given data $X_1, ..., X_n$ lying in $\left(S^1\right)^2$, we seek the projection matrix P which simultaneously makes the PX_i as close as possible to their unit circle projections, and also explain as little variation as possible (in the spirit of backwards PCA).

First, using subscripts in a different manner, note that for $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2$, the squared distance from x to its projection on the unit circle is

$$\left(x_1 - \frac{x_1}{\sqrt{x_1^2 + x_2^2}}\right)^2 + \left(x_2 - \frac{x_2}{\sqrt{x_1^2 + x_2^2}}\right)^2 = \left(x_1^2 + x_2^2\right)\left(1 - \frac{1}{\sqrt{x_1^2 + x_2^2}}\right)^2 =$$

$$= \left(\sqrt{x_1^2 + x_2^2} - 1 \right)^2 = x_1^2 + x_2^2 - 2\sqrt{x_1^2 + x_2^2} + 1 =$$

$$= \left(\sqrt{x_1^2 + x_2^2} - 1 \right)^2 = \left(\|x\| - 1 \right)^2 .$$

Thus, given data $X_1, ..., X_n \in \left(S^1\right)^2$, and a regularization parameter λ, we seek u, $v \in \mathbb{R}^4$ which give the projection matrix $P = uu^t + vv^t$ where

$$\arg\min_{u,v} \left[\sum_{i=1}^{n} \left(\frac{1}{\sqrt{2}} \|PX_i\| - 1 \right)^2 + \lambda \sum_{i=1}^{n} \|(I-P)\, X_i\|^2 \right] .$$

Note that this criterion seeks both to make the projected data close to the unit circle, and also seeks to minimize the variation in the direction of the subspace determined by P. Note that when $\sum_{i=1}^{n} \|(I-P)\, X_i\|^2 = 0$, it follows that $\sum_{i=1}^{n} \left(\frac{1}{\sqrt{2}} \|PX_i\| - 1 \right)^2 = 0$, i. e, that the PX_i all lie on a circle of radius $\sqrt{2}$ in $\mathbb{R}^4$. To find a scores representation of the data, essentially map the projections in $u - v$ space into angles, e.g. with the Matlab function $atan2$, to get the first scores. For the second scores, do the same, with the projections of the data with respect to $I - P$.

This $u - v$ parametrization can be used to generate interesting examples on $\left(S^1\right)^2$. In particular, a given u and v determine a great circle on S^3, as

$$\left\{ u \cos\alpha + v \sin\alpha : \alpha \in (0, 2\pi) \right\} .$$

Then the $\left(S^1\right)^2$ representation of this set of points is computed as:

$$\left\{ \begin{pmatrix} atan2\left(u_1 \cos\alpha + v_1 \sin\alpha, u_2 \cos\alpha + v_2 \sin\alpha\right) \\ atan2\left(u_3 \cos\alpha + v_3 \sin\alpha, u_4 \cos\alpha + v_4 \sin\alpha\right) \end{pmatrix} : \alpha \in (0, 2\pi) \right\} ,$$

where again $atan2$ is the e.g. Matlab four quadrant version of the arc tangent.

2.2 Higher Dimensions

To extend these ideas to the case of $d > 2$, the main idea is to similarly decompose the S^{2d-1} or $\mathbb{R}^{2d}$ embedding of $\left(S^1\right)^d$, in terms of two dimensional projection matrices, $P_1, \cdots, P_d$, using again a projection and optimization approach.

In the first step the goal is to find projection matrices with $P_1, \cdots, P_{d-1}$, with $P_j^2 = P_j$, $tr(P_j) = 0$ and $P_j P_k = 0$ for $j \neq k$, as

$$\hat{P}_1, \cdots \hat{P}_{d-1} = \arg\min_{P_1, \cdots, P_{d-1}} \left[\sum_{i=1}^{n} \sum_{k=1}^{d-1} \left(\frac{1}{\sqrt{d}} \|P_k X_i\| - 1 \right)^2 + \lambda \sum_{i=1}^{n} \left\| \left(I - \sum_{k=1}^{d-1} P_k \right) X_i \right\|^2 \right] , \tag{2}$$

where again λ is a regularization parameter. This results in representing each data point X_i by its lower dimensional projection $X_i^{(1)} = \sum_{k=1}^{d-1} \frac{\hat{P}_i X_i}{\left\| \hat{P}_i X_i \right\|}$. {need to think more carefully about degenrate cases where $\left\| \hat{P}_i X_i \right\| = 0$. This also gives the level d scores as points in S^1, again using an atan2 calculation on $\left(I - \sum_{k=1}^{d-1} P_k \right) X_i$ as above.

This process is repeated, so that at step l, we solve

$$\hat{P}_1^{(l)}, \cdots \hat{P}_{d-l}^{(l)} = \arg \min_{P_1, \cdots, P_{d-l}} \left[\sum_{i=1}^{n} \sum_{k=1}^{d-l} \left(\frac{1}{\sqrt{d-l+1}} \left\| P_k X_i^{(l-1)} \right\| - 1 \right)^2 + \lambda \sum_{i=1}^{n} \left\| \left(I - \sum_{k=1}^{d-l} P_k \right) X_i^{(l-1)} \right\|^2 \right],$$

where λ_l is the regularization parameter, and where $P_1, \cdots, P_{d-l}$ satisfy $P_j^2 = P_j$, $tr(P_j) = 0$ and $P_j P_k = 0$ for $j \neq k$. Next find the level $d-l+1$ scores as points in S^1, again using an atan2 calculation on $\left(I - \sum_{k=1}^{d-1} P_k \right) X_i$ as above.

Note: multiple steps are needed, as opposed to just stopping with the projection matrices found in (2), to give a complete ordering of the subspaces (i.e. the S^1 like components).

3 Variations and Extensions

There are many other things do on this, and also variations of the main idea to be explored. Some of these are:

- Change S^1 to S^k (eg. $k = 2$ for s-reps), to handle the polysphere case. At first glance, this looks rather straightforward.
- This approach is mostly extrinsic in nature. Is there a somewhat parallel fully intrinsic approach?
- Is there a way to replace the greedy search by some analog of Xavier Pennec's global flag manifold idea?
- Relationship to Stiefel and Grassmann manifolds?
- Instead of fitting to circles of radius 1, one could modify the algorithm to consider circles of smaller radii, and one could optimize over the radii. This would correspond to modifying Principal Nested Great Spheres, to allow for smaller radius spheres.
- Can all of this be notationally simplified (or even methodologically improved) by using complex number representations?

References

[1] James Damon and James Stephen Marron. Backwards principal component analysis and principal nested relations. *Journal of Mathematical Imaging and Vision*, 50(1-2):107–114, 2014.

[2] S. Jung, I.L. Dryden, and JS Marron. Analysis of principal nested spheres. *Biometrika*, 99(3):551–568, 2012.