Response

H. Sherry Zhang

22/08/2021

Many thanks for reviewing the manuscript and for your comments and suggestions. Please see below the answers to your review comments.

Reviewer 1

The Review

This work presents visual diagnostic tools for different methods of projection pursuit optimization. They use line plots, PCA and the tour method to visually compare the behavior of different optimizers. Finally, they showcased the analysis of three optimizers using their tool.

The tool is likely to be of interest to the R community. Most of my comments are about the technical aspect of the content. The writing is overall good, but please see my comment under the "writing" regarding to the clarity of the text.

Major comment

- "Sometimes, two optimisers can start with the same basis but finish with bases of opposite sign…However, this creates difficulties for comparing the optimisers since the end bases will be symmetric to the origin. A sign flipping step is conducted… " and the "Reconciling the orientation" section
  - For 1D projections, flipping the sign is sufficient. For p-D projections of $\mathbb{R}^n$, one may need a little more. To my knowledge most index functions are invariant to any rotations that happens within the p-subspace. This makes me wonder that in general, one may want to align two bases so that their difference does not contain any "within-subspace" components. In other words, one may want to compare representations in the Grassmannian manifold (space of k-subspaces) instead of in the Stiefel manifold (space of orthonormal k-frames).

We agree that most index functions should be rotation invariant (this is in fact a desirable property, but there are exceptions where indexes can still be very useful but not rotation invariant, see discussion in Laa, Cook (2020)). Thus when comparing frames defining projections to d>=2 dimensions, we may want to introduce a rotation that will align the bases as much as possible within the common subspace, while potentially quantifying e.g. the angle between two planes. This is an interesting problem, however the implementation of this general approach is beyond the scope of the current paper.

Minor comment

- The software seems to be specific to projection pursuit. I wish the authors could comment more on its potential generality to other optimizations. For example, can one generalizes the tour method to visualize any parameter trace of an optimizer?

It is feasible to generalise the tour method for diagnosing other optimisers, given that the parameter space can be simulated. In our examples, the orthonormality constraint regulates the parameter space of the projection basis ($p \times d$) to be an p-sphere, which can be simulated by the geozoo package. For other optimisers, reserachers need first to create a data collection object in a tibble that tracks the parameter during the optimisation, and then to use ferrn to produce the diagnostic plots as shown in the paper. Some examples of the data collection objects that we created are included in the ferrn package and should be referred to as the format for inputs into the expore_ functions.

- Figure 4 "... All the bases in PD have been flipped for easier comparison of the final bases..."
  - This is related to the major comment. In the figure, PD flipped in sign but stopped at a basis representation similar to the destination of the CRS. Flipping the sign back, does it mean that PD and CRS diverged but stopped at the similar subspace with opposite basis representations? Also, given that PD and CRS have the same starting bases and one is flipped, why wouldn't the dashed line pass the origin (gray dot) in the figure?

Thanks for sharing your interpretation on the effects of flipping, that's right! The reviewer is giving a decent explanation in describing the effect of flipping. The gray dot should land on the dashed line and this is a mistake due to the data not being updated after flipping in the code. This has now been corrected.

- "Recall that the columns in a 2D basis are orthogonal to each other, so the space of $p \times 2$ bases is a p-D torus".
  - It is not clear to me why "so". For p=2, the space of 2x2 bases is a pair of 1-D tori instead of just one (because you have two orientations for a pair of orthogonal vectors), right? For p=3, the space of 3x2 bases is SO(3), which is not a 3-D torus? Ref: https://en.wikipedia.org/wiki/Stiefel_manifold

We changed the wording of the sentence to: "a torus in the p-D space" and provide a reference for this statement: (Buja and Asimov, 1986).

Writing

- "Along with the bases recorded during the optimisation and a zero basis, the parameters (centre and radius) of the 2D space can be estimated by PCA. The centre of the circle is the first two PC coordinates of the zero matrix, and the radius is estimated as the largest distance from the centre to the basis."
  - It took me multiple passes to get what this means. Perhaps instead starting with something like: "We use PCA to project and visualize the parameters/bases in 2D. The centre of the 2D view is..."?

Changed to "We use PCA to estimate the 2D space that is best suited to visualise the bases. The centre of the 2D view is the first two PC coordinates of the zero matrix, and the radius is estimated as the largest distance from the centre to the basis."

- "Figure 6 shows some frames from the tour plot of the same two optimisations in its original 5D space."
  - I think this is the first mention of "5D space"/5D data in the main text (apart from the caption in Figure 4), so it may be worth clarifying what the original 5D space is.

The data set is first introduced in the subsection *Simulation setup* under *Diagnosing an optimiser*. However, to avoid confusion, we change the "original 5D space" to "original space" and give more details about the original space in the figure caption: "The basis space in this example is a 5D unit sphere ...".

- "While the previous few examples have looked at the space of 1D bases ... 2D bases"
  - I don't think the terms "1D bases" and "2D bases" are standard. Perhaps (orthonormal) "1-frames" and "2-frames"?

Changed to "While the previous few examples have looked at the space of 1-frames in a unit sphere, this section visualises the space of 2-frames."

Reviewer 2

These experiments are interesting and so is the methodology. The conclusions could do with beefing up What can we learn about how to optimise the holes index for 1 and 2d? Does this apply to other smooth indices? What about non-smooth indices? What about indices in other dimensions?

A paragraph further summarising the findings has been added in the second paragram in the conclusion

1. Should refer to earlier work on scagnostic indices, page 3

New references have been added as follows:

Added reference of (Wilkinson et al., 2005) and (Wilkinson and Wills, 2008):

L. Wilkinson and G. Wills. Scagnostics distributions. Journal of Computational and Graphical Statistics, 17 (2):473–491, 2008

L. Wilkinson, A. Anand, and R. Grossman. Graph-theoretic scagnostics. In IEEE Symposium on Information Visualization, 2005. INFOVIS 2005., pages 157–164, Oct 2005.

2. figure 2 caption "using the holes index" "and a boxplot". What is the colour scale for the points? The text related to figure 2 in the bottom of page 5 should mention the dataset and the index, giving references if needed.

The caption has now been re-written and color scale explanation added: "The color scale is from the customised botanical palette in the ferrn package."

3. "with higher index value" -> "with a higher index value". The word "the" is missing in a few other places in the paper.

That has now been corrected.

4. In Figure 3 would be better if the y-axes scales were the same for the two graphs.

It is now fixed

5. What is the star in figures 4 and 5? This is not mentioned in the text or caption

The caption has now been ammended to: "The black star represents the theoretical best basis the optimisers aiming to find." in the caption of figure 4.

6. In the text you say things like "details=TRUE" and "animate=TRUE" but what functions are you referring to?

Mention of "details=TRUE" and "animate=TRUE" is removed from subsection diagnostic 3a and 3b to the implementation section: ‘explore_space_pca()‘ produces the PCA plot of projection bases on the reduced space. Figure 4 includes the additional details of anchor and search bases, which can be turned on by the argument ‘details = TRUE‘. The animated version in Figure 5 is produced with argument ‘animate = TRUE‘.

7. In Figure 9, it seems CRS is doing better here than in Figure 8– maybe this could be mentioned.

Figure 8 and 9 focuses on different aspects: Figure 8 compares the effect of implementing an interruption while Figure 9 compares the effect of polishing. In the experiment of Figure 9, the parameter ‘max.tries‘ is set to 400 to allow guided tour to do its best instead of a default 25 in Figure 8. Cross comparing the two experiments is less meaningful.

8. "not the case for all the indexes" missing a "." page 11

It has now been added

9. For the Kolmogorov index, why are you using randomly generated data instead of normal quantiles?

The Kolmogorov Smirnoff index here looks at the difference between a theoretical pdf and an empirical one based on the data. In our case we are not assuming normality on the theoretical pdf and that is the reason why we are not using normal quantiles. To make it clearer, we change the wording from "Normal Kolmogorov Index" to "Kolmogorove Index".

10. Fix references with “et al”

Fixed!