October 22, 2024

Dear Editor,

We would be very grateful if you would consider this revision, entitled *'On the use of the Gram matrix for multivariate functional principal components analysis'* for publication in *JMVA* journal. This is a revised version of the manuscript JMVA-D-24-00251. We were able to address the comments made by the reviewers and to strengthen our results. A point-to-point reply is provided below.

We quite naturally hope that this version matches reviewers' expectations, and we thank you in advance for your consideration.

Sincerely,

Steven Golovkine
(on behalf of the co-authors Edward Gunning, Andrew J. Simpkin and Norma Bargary)

Comments received for *On the use of the Gram matrix for multivariate functional principal components analysis*

The changes are in blue in the main text.

Reviewer 1

Comment #1 :

It seems to me that the sentence "FPCA was introduced by Karhunen (1947) and Loève (1945) and developped by Dauxois et al. (1982)" is not entirely true. In fact, Karhunen and Loève independently proved the so-called Karhunen-Loève decomposition (written on p. 9, eq. (8)) but they were not aware of the possible applications of FPCA at the time. FPCA was later developed by Kleffe (1973); Deville (1974). Note also that at the same time as Kari Karhunen and Michel Loève, an orthogonal series decomposition was also proved by Kosambi (1943).

Reply:

We thank the Reviewer for these details on the history of the FPCA. We have modified the text accordingly.

Comment #2 :

The sentence "the principal components obtained from a PCA run on the rows data matrix are the same as the ones obtained from a PCA run on the columns of the matrix" is not precise: the eigenvalues are the same but the eigenvectors are different (even if there is a relationship between them).

Reply:

We thank the Reviewer for this comment. We have replaced "principal components" by "eigenvalues".

Comment #3 :

The estimation of $\widehat{C}_{p,p}(s_p, s_p)$ depends on the noise variance σ_p^2 which is unknown in practice. How do you estimate it?

Reply:

The last sentence of Section 2 refers to the estimation of the noise variances. We acknowledge that it was not clear that this sentence also referred to the noise variances in $\widehat{C}_{p,p}$. We have modified the sentence to make it clearer.

Comment #4 :

"the time complexity [...] of the univariate score is $\mathcal{O}(NM_pK_p)$". It is not written very clearly but I assume that it is the time complexity of computing the scores of all individuals?

Reply:

We thank the Reviewer for this comment. We now use "univariate scores".

Comment #5 :

It seems that the two sentences "we focus solely on players who have made more than 1000 shots" and "We remove [...] players that have made fewer than 100 shots" contradict each other. Or maybe I do not understand the overall meaning. Which players exactly are you removing?

Reply:

First, we exclude players who have made less than 1000 shots during the five-season period. We then exclude shots from a restricted area (close to the hoop). Finally, we exclude players

who have made less than 100 shots outside the restricted area. We have modified the main text to clarify the process.

Comment #6 :

I see no difference between the eigenfunctions of shots attempted and shots made, the eigenvalues are exactly the same. This seems very surprising to me. Do you have an explanation? Also, plotting the scores can be informative for the interpretation.

Reply:

We agree that the eigenfunctions of shots attempted and shots made are similar. A possible explanation is that the shots made are included in the shots attempted. We are not sure what the Reviewer means by "the eigenvalues are exactly the same". If the Reviewer refers to the percentage of variance explained in the Figure 10, we recall that the eigenfunctions are vector-valued $\phi_k = (\phi_k^{(1)}, \phi_k^{(2)})^\top$ and thus the eigenvalues are related to ϕ_k and not $\phi_k^{(1)}$ or $\phi_k^{(2)}$ individually. While we acknowledge that a plot of the scores may be informative for the interpretation, we choose to not include it as the main interest of the paper is not the analysis of this specific dataset, but the methodology to estimate the multivariate principal components.

Reviewer 2

Comment #1 :

The novelty of the research can be improved and highlighted by providing another data example to show the differences from previous works and the supplementary python codes need more documentation.

Reply:

We thank the Reviewer for the comment. We have added some comments to the Python code and another example on Canadian weather in the Supplementary material.