Reviewer report on the paper ‚On the use of the Gram matrix for multivariate functional principal components analysis' by Golovkine et al

The paper investigates the problem of multivariate functional principal component analysis (MFPCA) and proposes an alternative estimation method that uses the Gram matrix, i.e. functional inner products, instead of being based on covariance smoothing. It thus combines ideas from existing MFPCA methods with existing methods from PCA in multivariate statistics and fully develops the resulting method for the multivariate functional case. Overall, I enjoyed reading the paper. The alternative estimation method seems to improve on existing competitors at least in some settings, and I also enjoyed seeing an honest and helpful discussion of the pros and cons of the different methods in different settings. I have some comments on the paper.

1. The paper is lacking any application besides the simulation study. I would find it illustrative to see a real data setting where the alternative estimation strategy improves on existing methods in terms of computation and/or results, as well as a discussion of any potential differences in results.
2. Your assumption of no error hardly seems realistic in practice. What are your recommendations in practice? Pre-smoothing can be an option, but also induces error that is later not taken into account, and can have problems in not so dense settings. It seems that in a setting with error, covariance smoothing based options may have an advantage and it would be good to discuss this more thoroughly and include a corresponding setting in the simulations.
3. As the scalar product was introduced without weights, the difference and correspondence between weights in the scalar product and rescaling of the data does not become very clear in Section 3.4. I would suggest to briefly introduce the weighted version of the scalar product and the correspondence of rescaling with the square-root of the weight (if not rescaling point-wise). E.g. your w_p(t_p) in the second formula really corresponds to the point-wise version of the square root of the proposal in the first formula, as the first formula uses weighting and the second formula rescaling. Not sure how your own proposal fits in there – with the square-root it should be a rescaling rather than a weighting as stated, surely? Why do you propose the integrated squared norm of C_p rather than the variance and how do they relate? You could also mention the option of weights already on p. 3, as it seems relevant in many applications.
4. In the simulation study, I have several questions regarding the choices made.
i) You do not seem to give sufficient detail to judge whether your comparison to Happ & Greven (2018) is fair. Not sure whether it is sufficient to use the FCP-TPA (Allen, 2013) option, as this seems to assume some separability (?) that may or may not be warranted, it would be better to also include another sensible basis choice such as tensor product B-splines. Right now it is hard to judge whether worse results in the second scenario are due to the image setting or this particular choice, given similarly good results in the first scenario. Also, how did you chose the smoothing parameter and how many univariate basis functions did you retain, as truncating too early will lead to a loss of information?
ii) The log-AE in (12) can become negative if the absolute difference is below 1. This does not seem like a sensible error measure and the relative versions below non-sensical. Please check.
iii) Which are the orthonormal basis functions chosen in Simulation scenario 1?

iv) Why do you only consider a univariate functional (image-valued) setting in Scenario 2 if your method targets multivariate functional data, not say a second one-dimensional function as in Happ & Greven?
v) Figures have small axis labels etc. and are hard to read. Figures 7 and 9 are particularly bad. Figure 9 seems to be empty besides results for p=2 – is this due to imcomlete results or choice of a bad depiction? You state that in the other settings besides p=2 the results of the two methods are 'identical' (which would indeed yield a boxplot not distinguishable from the 1-line), which is hardly exactly possible, please check. Sometimes the text does not seem to match the (later) figures, please check. There are likely too many figures and some parts may need to be delegated to an appendix.

<u>Smaller comments:</u>
1. Please introduce any notation and names you use, even if also used by one of the papers you cite. For example, 'feature' middle of p. 3. What is \mathcal{Z}_{n,.} in 4.1?
2. Your notation should distinguish between theoretical and empirical quantities, i.e. their estimators. For example \mu^{(p)} or C_{p,q} on p. 3 are used for both.
3. distance should be similarity, last line of p. 3. diag would be clearer as blockdiag, p. 4 ff.
4. The distinction between f and M_f could be clearer (p. 6 ff). Can it be that M_f = M_g for f \neq g?
5. The discussion of (9) and (10) could more clearly point out where quantities in (10) are the p-th element of the corresponding quantities in (9) (\mu and c_k) and where they aren't (\phi_k vs. \varphi_k) as the relationship can otherwise be confusing.
6. In 4.1, I believe it should be "For a feature X(p), the eigenfunctions and eigenvectors are computed **using** [not as] a matrix decomposition of the estimated covariance …" as you will need to do a rescaling of eigenvectors to ensure orthonormality with respect to the functional not vector inner product, correct?
7. On p. 9, v_k is used for the eigenvectors of both Z and M, please change one notation.
8. You seem to go a bit back and forth between the weighted and equally weighted observation case, could you make this a bit more consistent?
9. For the statements of computation complexity in Section 4.3, I could easily see some, while others seemed less obvious. It would be good to have brief derivations in an appendix. Could you also state the resulting complexities when using SVD in Remark 5?
10. In the discussion, you say "… whereas the decomposition of the covariance operator necessitates the specification of the percentage of variance accounted for by each individual univariate feature. Specifying the percentage of variance explained for each feature does not guarantee that we recover the nominal percentage of variance explained for the multivariate data." An option that should be mentioned here as it is usually advisable is to retain 100% of the possible univariate components and to only truncate the multivariate expansion, which does not cause this particular problem.
11. In the appendix, please make clearer where one proof ends and another starts and which formula in the paper you are showing at any given time.