

**Alan Heavens, March 22**

As I see it, the paper has two broad aims: to explore a number of ways to compress weak lensing data; and to look at the sensitivity of results to covariance matrix errors, including seeing which elements of the covariance matrix are most important for parameter inference, and which should therefore be computed especially accurately if possible.

The paper makes a number of good points, but it has some problems, and this report is rather critical, but I hope constructively so. It might be worth stepping back and thinking what the main messages of the paper should be, what results can be presented that are genuinely new, and if some of the studies can be more systematic than now.

#### Summary

The executive summary is that the compression side tries a few things, and concludes that the best way is to use MOPED. Since MOPED is known to be optimal and can't be beaten in terms of the Fisher matrix (subject to some conditions), this shouldn't be a surprise. For suggestions of how to present something new here, see details below.

For the covariance matrix studies, this is an interesting question to tackle. The concern I have is that it's not clear how to do this in a systematic way, and I think that a referee will consider it to be too ad hoc to draw general conclusions from. This needs to be strengthened. I suggest a small technical improvement to avoid non-positive definite matrices, but I don't have any good ideas for a more systematic general approach.

#### Detailed comments

The first compression method involves modifying the highest or lowest eigenvalues of the covariance matrix, effectively discarding the eigenmodes with the highest or lowest eigenvalues. It's perhaps not surprising that this doesn't work very well, since the procedure knows nothing about the sensitivity of the covariance matrix to the parameters, and there is no formal way of ordering the modes by contribution to the information on each parameter. The second compression method is better, based around S/N eigenmodes (Vogeley and Szalay 1996 should be referenced). Here linear combinations of the field values are considered (field here corresponds to shear measurements on a number of tomographic slices). Compression to the highest S/N modes is optimal in one sense, if (and only if) it is the amplitude of the (various) power spectra that is the (single) target. This is a more principled approach, and it is not surprising that it does better. It is still not very good, as Fig. 4 shows, because in general, we want to infer multiple parameters, whose influence on the power spectra is more complicated. This problem was tackled by Tegmark, Taylor, Heavens (1997; TTH), and we found the modes that would be optimal for any single parameter, by including the sensitivity of the power spectra to the parameters. However, the challenge is to find a compression that works for all parameters simultaneously. This is not at all trivial. In TTH we proposed combining the

optimal modes determined for each parameter, and it turns out to offer something, but in the end it does not work very well, for reasons that I can explain if needed. The S/N eigenmodes do seem to be quite effective, in a restricted setting when applied in the radial direction only, as Davide Alonso found. Here we see that they are limited in this context, perhaps because once again they have no knowledge of the sensitivity of the signal to the parameters, so are therefore almost inevitably sub-optimal.

The final approach is to work with linear combinations of the two-point functions (rather than of the field). This uses a principled approach, and is the MOPED algorithm from Heavens, Jimenez, Lahav (2000), just without the (optional) Gram-Schmidt orthogonalization. A number of authors have rediscovered it since then, such as Gualdi et al., but the radical compression to the number of parameters goes back to HJL. TTH found the solution for a single parameter, but (as seen above in the compression of field values) the extension to multiple parameters is not trivial at all; guaranteeing that the Fisher matrix is unchanged on compression is not even obviously possible. What HJL showed mathematically is that in fact it is, and the credible regions can be unchanged (at least at the level of the Fisher approximation), even with such a radical compression, and this is what Fig.11 basically shows. This of course is wonderful, but I wonder if it tells us anything new – we already know that MOPED works (and that there is nothing that can do better for gaussian data in terms of Fisher matrix).

To present something new here, it might be best to angle it as the first analysis that applies MOPED to a weak lensing situation with  $> 15$  parameters. In the interest of full disclosure, we have a paper in late draft stage that applies MOPED to weak lensing data. Given the LSST internal review process, our paper will likely appear on arxiv before this one. However, it has fewer parameters, so you might want to stress the  $> 15$  parameters aspect. **[We thank you for this information. We have included this point in the paper; we also make it clear that we’re using MOPED (minus the normalising factor).]**

A suggestion for the covariance matrix study, to ensure that the perturbed matrix is positive definite, is to take the matrix log of the fiducial covariance matrix, and then to modify the elements of the log matrix in some way (how to do this systematically, I’m not sure). Taking the matrix exponential will then give a positive-definite matrix for sure. Its still a bit ad hoc, but avoids that particular problem. **[One of the issues here is how to introduce an error to the log matrix that would be similar to what we expect to see in the original matrix. A naive assumption would be that**

$$C_{\alpha\beta}^{new} = (1 + \delta)C_{\alpha\beta}^{old}$$

$$\rightarrow \log(C_{\alpha\beta}^{new}) = \log(1 + \delta) + \log(C_{\alpha\beta}^{old}) .$$

**We know this to be wrong, since obtaining the matrix logarithm is not as simple as that, so this approach would be incorrect. Also, applying error to the elements of the log matrix (like a 10% error) produces a**

covariance matrix with elements several orders of magnitude larger than the original one. This would work if we diagonalised the matrix and applied the transformation to the *log* of its eigenvalues, but since we already have a section for modifying the eigenvalues, we fear this would be repetitive.]

Smaller points are:

The comparison of the elements in Fig. 2 is a bit misleading (or rather, the conclusions are). Assuming that the small elements (which show the biggest fractional difference) are off-diagonals, then it probably doesn't much matter if they are  $10^{-4}$  or  $10^{-8}$  of the diagonal elements they are effectively zero. If they are on the diagonal, then it's a different matter, so it is worth checking. [We added a more thorough investigation into these differences, pointing out that the biggest difference are in the smallest elements and correspond to off-diagonal terms.]

The compression in Fig. 8 shows that for the compressed data, the two covariance matrices lead to compressed covariance matrices which agree much better. This is encouraging, and expected since the linear combinations involved will be dominated by the larger elements, which agree closely, and differences in the small elements are largely irrelevant. [These are the results we were hoping for.]

For the MOPED compression, it is suggested that the full covariance matrix is used to determine the fiducial model where the derivatives are computed, but this is slow. It is much faster to choose a rough fiducial model and find the maximum likelihood from a preliminary MOPED analysis and then repeat with this as the fiducial model. The process could be repeated as many times as necessary, but it's never been needed in my experience. [The fiducial values used were the bestfit values obtained in the previous analysis with the full covariance matrices. We agree with your comments, but we already had those results at hand.]

I'm sorry not to be able to suggest a route that makes the covariance matrix study likely to be strong enough, but I hope that these comments are constructive. [These points have helped immensely, and we are very thankful. We hope that we have now been able to provide a draft with clearer and more conclusive results.]

David Alonso, April 3

My overall comments would go along the same lines as Alan's, so I won't repeat them here. My main worry is the fact that there doesn't seem to be a clear lesson learnt or take-home message with regards to the problem of identifying the most relevant components of the covariance matrix. This is a hard problem! One possible angle that I didn't see explored here is whether MOPED would be able to circumvent errors in the covariance matrix. I.e. how much do the contours change if you use the two different covariances? How about their associated

minimum  $\chi^2$ ? [We explored this in more detail with the addition of the constraints on the intrinsic alignment parameter  $A_{IA}$ . We generated 9 new covariance matrices by introducing error to the compressed Full Covariance Matrix (FCM, which we previously referred to as CL), and saw that, while the constraints on  $\Omega_m$  and  $S_8$  did not alter by much, there was an alarming loss of constraining power for  $A_{IA}$ ]

Beyond this, I have a number of specific comments. Here they are in order:

- **L32** Individual  $\rightarrow$  independent? [Done]
- **L43** I'd remove this paragraph from the introduction. [Done]
- **L47** You seem to use the present continuous a lot we are trying. I think I'd go for the present simple. [Done]
- **L113** Could you explain in a bit more detail what the differences are between the two covariances? I.e. is it that one is missing the non-Gaussian parts? [We explain this better in the text now. The Gaussian Covariance Matrix (GCM) only consists of the Gaussian parts.]
- **Eq. 1** Does  $T$  here contain the noise bias contribution? I imagine not, but I wanted to check. [This is just the formal definition.]
- In both sections IV and V you seem to arbitrarily choose 200 as the number of modes to throw away. Can you justify this? [We modified the text to make this clearer. This is done to achieve a reduction of about 90% of the covariance matrix, so we have a fairer comparison between these methods and the other, more complicated compression schemes.]
- **Fig. 5** This would benefit from larger labels and legend. You could even consider making it a double-column figure. [Done]
- **L249** compression at the ell space - I think I'd say map-level compression. [Done]
- **Eq. 6** Use  $\langle$  and  $\rangle$  instead of  $<$  and  $>$ . [Done]
- **L314** It is worth justifying that ell-independent  $U^{ij}$  is a good approximation (which it usually is for cosmic shear). [Low ell spread is because of the cosmic variance, and is also lower weighted by the weighted average, high ell shows great consistency in Figure 5.]
- **Eq. 10** It'd be good to justify the 2ell+1 weighting. [Justified by the number of modes for each ell.]
- **Fig. 7** I think  $W_{ij}$  should be  $U^{ij}$  in the caption. [Done]
- **L354** is  $\rightarrow$  are [Done]

- **L354** The result also makes sense in terms of the cumulative nature of lensing (i.e. the further away the source, the more lensed it is), although the thinning out of the galaxy density at higher redshifts makes the 4th bin have a lower weight than the 3rd one. [\[Done\]](#)
- **Fig. 8** This plot is tiny! Do you need the right panel? [\[Done\]](#)
- **Fig. 10** Seems to appear in the text before Fig. 9. Actually, I'm not sure you discuss figure 9. [\[This was removed.\]](#)
- **Eq. 12** The transformation is parameter-dependent. I assume you perform it at a fiducial cosmological model? [\[We do indeed. We used the best-fit results from the full cosmological analysis and take those to be our fiducial cosmology everywhere.\]](#)
- **Eq. 13** Probably worth emphasizing that this is the Fisher matrix of your likelihood. [\[Done\]](#)
- **L415** Where does eighty thousand come from? [\[We were referring to the elements of the covariance matrix. We fixed this in the text.\]](#)
- The shape and variance of the posterior is not dependent on the derivative, but the best-fit value shifts according to the point where the derivative is taken. I'm a bit confused by this. Can you justify this statement? [\[We've removed this paragraph from the paper. The point of it was to explain that we needed a cosmological analysis had to be done first, in order to get the bestfit values.\]](#)
- You should try to reorganize the figures so they appear a bit closer to where they are discussed. [\[Done\]](#)
- **L449** We find that the mean values of the parameter constraints for the two methods agree to  $1\sigma$ , which shows that they are equivalent to each other. I don't understand this statement in the context in which it appears (the comparison between covariance matrices). [\[Since we are comparing the two compressed covariance matrices, there are two ways of doing this: we can either compress them using the same weight \(remember that the inverse of the covariance matrix is used for this result\), or with a different weight for each \(the weight would be obtained with the corresponding covariance matrix\). We wanted to check how different weights impact the parameter constraints and found the differences to be small \(as compared to the difference between the constraints for the two matrices\).\]](#)
- **L496** The qualitative jump between 25% and 30% errors is hard to digest. What is the cause of this? Also, did you do this for a single realization of the noisy covariance or for several of them? [\[We redid the analysis with 50 realizations of the covariance matrix. The jump was](#)

most likely due to a large fluctuation, which was corrected with more realizations. The new analysis brought new aspects to light, which are pointed out in the new draft.]

- **L510** Do you mean  $\delta > -1$ ? [We meant  $|\delta| < 1$ .]

**Jonathan Blazek, April 6**

I have followed with interest the progress of this important work, and I congratulate the authors on completing the paper. Please let me know if you have questions about the below comments/suggestions. Note that I deliberately wrote my comments before reading the other reviews, so there is some overlap.

One general comment before specifics. The paper has two main goals: 1) comparing different covariance codes/methods and 2) exploring data compression. I find that these two threads are not always clearly distinguished. For instance, much of the data compression is done only on the CL covariance matrix, and the conclusions about the comparison part are not given much attention. It could help to slightly restructure where the results are presented, or to add a few words making clear when CL-only is being used and when an important comparison is being done.

- For the reason mentioned above, the abstract should be more clear. What are the main goals and conclusions? Also start along the path in the first sentence sounds tentative for the start of an abstract. [Updated the abstract to include more information about our goals and conclusions.]
- Reference numbers are out of order. [The references are in alphabetical order, that's why they're not in increasing order in the text.]
- **L34:** As data sets increase  $\rightarrow$  you mean the number of elements in the data vector, not necessarily the overall amount of observational data. [We replaced *data sets* with *data vectors*.]
- **L68** Missing section symbol. [Done]
- **L83** Emphasize here (or elsewhere) that this analysis is done in configuration rather than Fourier space. Is there any reason we would expect the qualitative results to be different in a Fourier space analysis? [Map level compression actually works better in Fourier space, but we don't think MOPED is going to differ much in two setups. The reason we use real space is because it is in line with the DES Y1 analysis.]
- **Table 1** Here, and in the relevant parts of the text, it may improve clarity to add the  $_{IA}$  subscript to the A and eta variables. Also, you can cite the Troxel DES cosmic shear paper when mentioning the IA params. [Done]

- **L84** DES Y1 is the state-of-the-art for cosmic shear data sets, at least right now. You could mention this.
- **L97** Are all of these citations for CosmoSIS needed? [These are the references that CosmoSIS spits out after each run. We thought it would be polite to include all of them.]
- **L140** The difference of 4 orders of magnitude is casually discussed without further explanation. Although your later results show that this is mostly due to non-important elements, more explanation at this stage would be helpful. Why is this happening? How can we plausibly think that these two cov matrices are roughly the same when there are such large differences. [We try to address this by taking into consideration Alan's comments as well. The elements that show greatest discrepancies are the smallest ones (of order  $10^{-18}$ , this is probably why the constraints are not all that different. As can be seen, the diagonal elements (and the larger ones), have smaller differences.]
- **Paragraph beginning at L163** I was a bit lost by the logic of this paragraph. If possible, provide more explanation of what you are doing and why. [We removed the figure and the associated text.]
- **L179 (and elsewhere)** Why 200 eigenmodes? This number is not explained, and the results depend strongly on how many eigenmodes are kept. Similar question about the S/N analysis. [We try to explain this better in the text. This is done to achieve a reduction of about 90% of the covariance matrix, so we have a fairer comparison between these methods and the other, more complicated compression schemes.]
- **L183** Throwing away 90% of the information... By what metric? Where does this number come from? [Change to reducing the number of elements in the covariance matrix by 90%.]
- **Fig. 4 (and elsewhere)** The blue/purple/magenta color scheme is tough to parse. They look similar, especially with overlapping contours. [Changed one of the blueish color to green. The goal is trying to be colorblind-friendly as well as monochrome copy friendly.]
- **Fig. 4 caption** You probably dont need all of the detail on how you remove these modes (which is included in the text anyway). You can just say that you remove those modes. [Done]
- **L210** very small. How is that quantified? [We meant the lowest 200.]
- **L226** signal OF these modes... [Done]
- **L234** vanilla What does this mean? [Changed to 'simple'.]

- **L237** Explain that the shaded region shows the modes excluded in the previous analysis. **[This is in the text.]**
- **L249** Compression at the ell-space... This sentence is confusing. **[Changed to map-level compression.]**
- **L267-269** Eliminates even more modes the numbers are the same. These ideas seem to be in contradiction. Which point do you want to emphasize? **[This paragraph was removed.]**
- **Eq. 5** I dont understand the indices. What is the summation convention? **[They are really just matrix multiplication, simplified to matrix representation.]**
- **L298** double summation What is meant? **[It really just meant  $\sum_i \sum_j$ .]**
- **Eq. 8** This is the standard  $\xi_{+/-}$  expression, which is not how it is described in the text above. **[Changed so that it is only meant to be used to derive the  $\xi_{+/-}$  for the compressed data.]**
- **L323-324** Wrong symbol. Should be arcmin, not deg. **[Done]**
- **L342** increasing shades is confusing. I also found the plot itself hard to parse, since the contrast between different ell values isnt very much. **[Made it so it is a little easier to see the change in color.]**
- **L345** What is  $W_\ell$ ? **[It should be  $U_\ell$ .]**
- **L347** You say earlier that the transformation has to be ell independent, but then discuss here that this is some ell dependence. Is this a problem? If not, why not? **[For low ells, there is the cosmic variance that is inevitable, but they are lower weighted so it is fine. For higher ells, they converge to an acceptable level.]**
- **L353-354** I found this sentence confusing. **[Added more specifications of what the results reveal.]**
- **L369** First two KL modes and their cross correlation? **[Yes, and changed.]**
- **L415** Do you mean elements or linear combinations of the elements? **[We meant the elements.]**
- **L428** This discussion of parameter space volume is tough to understand without discussing the number of parameters. Also, do you want to claim that the other parameters agree well when they are only consistent at the 2sigma level? **[We wanted to be very conservative here. The cosmological and IA parameters show an agreement of > 99% (with the exception of  $\alpha_{IA}$ ). The real problem are the systematic parameters.]**



- **Fig. 7** As I mentioned before, the different colors are pretty subtle and hard to parse. Also, you should mention the ell range spanned by the different color shades. [Changed the color to blue-green-orange, should be better. Also specify the meaning of different shade level in the caption.]
- **Fig. 8** Larger text, like in Fig 12. Also, why are some of the diag elements below 1, when all of the others are above? I didnt see a discussion of this anywhere. [Specified in Section III A in v2.]
- **L443** Extra space in i.e. [Done]
- **L451** Agreeing to 1sigma does not imply equivalence. [Changed this.]
- **A general comment about figures** There are several similar figures, and they dont always appear right next to where they are described in the text. So it would help if each caption included info on which method of compression is being used. [Done]
- **Fig. 9** Where in the text is this figure mentioned? [We removed this figure.]
- **Section VII** Is this always with the CL covariance? Is there any tolerance testing for the other one? [The tolerance testing is done only on CL/FCM. We believe the results will be similar, if done with GCM.]
- **L475** Clarify that the eigenvalues here are not the original eigenvalues used in the compression method. [Done]
- **L479** What is this extra step which is mentioned a couple of places? Is it simply a check on the final cov matrix? If so, does this introduce a selection bias on the otherwise random noise in the cov? [We made this clearer in the text. For the elements, a constant is added to the diagonal and in the case of the eigenvalues,  $|\delta| > 1$ .]
- **L517** What is the takeaway of this result and this final sentence? The question of noise in the eigenvalues is just dropped. What does this result mean? [We have given more emphasis in this new draft.]
- **Conclusion** Say more about the results of the comparison between the two methods. What have you learned? [We included a discussion about this in the new version.]
- **L538** Is it actually untrue that these modes carry less information, or is it simply that they do carry some information? [We see that they do carry relevant information, since we lost constraining power for all the parameters by applying this cut (this wasn't the case for SNR, for instance).]

- **L549** The phrasing of IA being sensitive to low S/N may be misleading. You mean low S/N in cosmic shear – the cov doesn't include IA at all. **[Done]**
- **L559** consecutive → resulting? **[Done]**
- **Fig. 14** Its hard to see whats going on since we dont know how the x axis is ordered. **[We added labels to the x-axis, showing the parts corresponding to  $\xi_+$  and to  $\xi_-$ .]**
- **Fig. 15** Hard to see black vs blue lines. **[We replace the figure, and changed the colours.]**
- **Fig 15** englobes → covers. **[Done]**
- **L577** This is very interesting - is it truly lossless? **[We included more information regarding this: it is lossless only if the noise in the data does not depend on the model parameters.]**
- **L592** The sudden degradation in constraining power when going from 25% to 30% errors is surprising (to me, at least) and an interesting result. Are you able to explain what is going on? **[The results shown in the paper had been for only one realization of the perturbed covariance matrix. We redid the analysis for 50 of those, and the results are now different – the previous results were likely due to statistical fluctuations.]**
- **L595:** What is happening with the eigenvalue noise? Are there interesting lessons from this result? **[Introducing noise to the eigenvalues was done in an attempt to ensure positive definiteness of the covariance matrix. The constraints obtained are similar to those obtained when the perturbation is applied to the elements themselves. The interesting lesson is what happens to the  $A_{IA}$ : even though the error in the final elements is small, the constraining power is largely lost for this parameter.]**