THE UNIVERSITY *of* EDINBURGH

Optimal all-sky power spectrum estimation techniques for cosmic shear

A comparison of estimators for upcoming Stage-IV weak lensing experiments

Alex Maraio

Institute for Astronomy, University of Edinburgh

maraio@roe.ac.uk

Cosmology from Home, July 2022

Outline for my talk

1. What is cosmic shear?

2. All-sky power spectrum estimators
   - 2.1 The QML estimator
   - 2.2 Most efficient QML implementation

3. Results
   - 3.1 Numerical performance
   - 3.2 Power spectrum errors
   - 3.3 Fisher forecasts

4. Conclusions

What is cosmic shear?

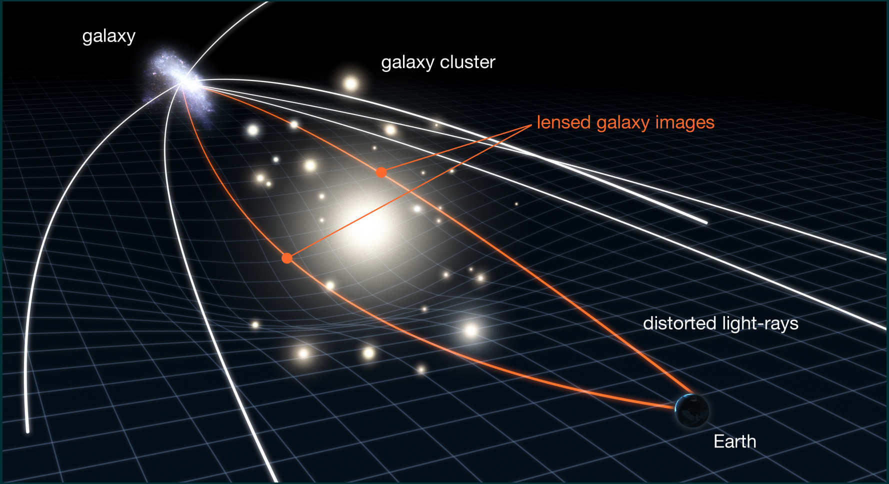

Figure 1: Cartoon showing light rays emitted from a source galaxy deflected by the intermediate lensing galaxy cluster, producing a lensed image. Source: ESA/Hubble

How do we observe weak lensing?

- We want to measure the distortion in the shapes of galaxies to infer properties about the intervening matter distribution
- Since galaxies have random intrinsic shapes, need to average over many galaxies to get the distortion from the matter distribution
  - The more galaxies we can image, the more accurate our measurements will be
- Hence, need high-precision telescopes - such as the upcoming *Euclid* space telescope (launch autumn 2024, hopefully!)
  - Will image around 15 000 deg^2 ($\sim$35%) of the sky

The effects of masking

- Since we cannot observe galaxy shapes over the full-sky, we need to mask our maps
- These masked regions arise for a number of reasons:
  - The galactic plane: Simply too many stars in our Milky Way
  - The ecliptic plane: Lots of dust in our Solar System
  - Extended objects: The Large Magellanic cloud, for example
  - Bright stars
  - The telescope cannot observe there: Ground-based observatories have fixed observing regions

The effects of masking

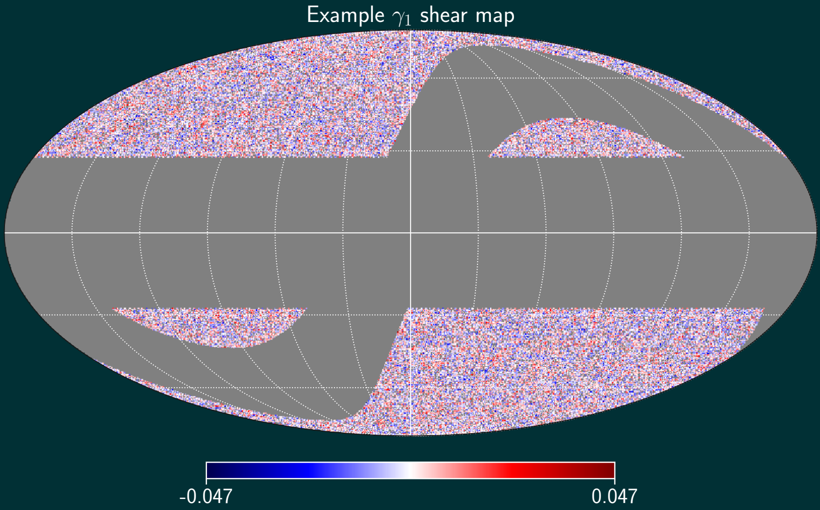

Figure 2: Example γ_1 shear map with *Euclid*-like mask applied

All-sky power spectrum estimators

Data compression

- A set of cosmic shear maps may contain many tens to hundreds of millions of pixels which makes comparing theory to observations at the map level quite difficult
  - But not impossible!
- Hence, we often want to use some form of summary statistic that characterises some of the physical properties of our maps, but using vastly fewer numbers
- Many methods have been developed and applied for cosmic shear:
  - Two-point correlation functions $\xi_{\pm}(\vartheta)$
  - Complete Orthogonal Set of E-/B-mode Integrals (COSEBIS)
  - Power spectrum coefficients C_{ℓ}

All-sky power spectrum estimators

- Using our masked maps as inputs, we want to recover the all-sky (unmasked) power spectrum of these maps
- This allows us to easily compare our observed maps to the theoretical predictions from different cosmological parameters / models
  - Allows easy comparison for use in a likelihood analysis
- The Pseudo-C_ℓ (PCl) estimator
  - Works in harmonic-space. Very efficient and allows processing of very high-resolution maps.
  - Has been shown to be non optimal
- The Quadratic Maximum Likelihood (QML) estimator
  - Works in pixel-space. Very numerically challenging which limits use to low-resolution maps.
  - QML estimator *is* optimal

The QML estimator

- First presented in the late 1990's in a series of papers by Max Tegmark[1]
- Finds a set of recovered $\tilde{C}_\ell$ values that maximises the Gaussian likelihood for a map $\vec{x}$ with covariance matrix $\mathbf{C}(C_\ell)$

$$\mathcal{L}(\tilde{C}_\ell \,|\, \vec{x}) = \frac{\exp\left(-\frac{1}{2}\,\vec{x}^\dagger\,\mathbf{C}^{-1}\,\vec{x}\right)}{(2\pi)^{N_{\rm pix}/2}\,|\mathbf{C}|^{1/2}}; \quad \mathbf{C} = \mathbf{S}(C_\ell) + \mathbf{N}. \tag{1}$$

- A minimum-variance quadratic estimator can be formed as

$$y_\ell = s_\ell - b_\ell; \quad s_\ell = \vec{x}^\dagger\,\mathbf{E}_\ell\,\vec{x}; \quad b_\ell = \mathrm{Tr}\,[\mathbf{N}\,\mathbf{E}_\ell]; \quad \mathbf{E}_\ell = \frac{1}{2}\,\mathbf{C}^{-1}\,\mathbf{P}_\ell\,\mathbf{C}^{-1}. \tag{2}$$

- This is related to the power spectrum through the Fisher matrix

$$\vec{C}_\ell = \mathbf{F}^{-1}_{\ell\ell'}\,\vec{y}_{\ell'}; \quad \mathbf{F}_{\ell\ell'} = \frac{1}{2}\,\mathrm{Tr}\left[\mathbf{C}^{-1}\,\frac{\partial\mathbf{C}}{\partial C_\ell}\,\mathbf{C}^{-1}\,\frac{\partial\mathbf{C}}{\partial C_{\ell'}}\right]. \tag{3}$$

[1] Tegmark 1997 `astro-ph/9611174`; Tegmark & de Oliveira-Costa 2001 `astro-ph/0012120`

The QML estimator

- All quantities ($\mathbf{C}$, $\mathbf{E}_\ell$, $\mathbf{F}_{\ell\ell'}$) are evaluated in pixel-space
  - This means that these matrices have dimensions that scale as $N_{\rm pix} \times N_{\rm pix}$
  - Hence, they become computationally intractable for even low resolution maps using existing methods ($N_{\rm side} \leq 64$; $\ell_{\rm max} \sim 128$)
- To analyse higher resolution maps, we need an alternative implementation that avoids direct computation of these quantities
- Introducing my new implementation of the QML estimator...

My efficient implementation of QML

- I have written a new implementation of the QML estimator that is the most efficient public implementation of QML yet!
- We avoid direct computation of the pixel covariance matrix $\mathbf{C}$ through the use of conjugate-gradient techniques
- We then use finite-differences differentiation to compute the C_ℓ-Fisher matrix
- TLDR: We can push our map resolutions up to $N_{\rm side} = 512$, where the previously computationally reasonable resolution using public codes was $N_{\rm side} = 64$

Conjugate gradient computation of $\mathbf{C}$

- We want to evaluate terms of the form $\mathbf{C}^{-1}\vec{x}$
- Instead of directly computing $\mathbf{C}$ and then inverting it, we can use a numerical iterative scheme that finds the solution vector $\vec{z}$ that satisfies

$$\mathbf{C}\,\vec{z} = \vec{x} \tag{4}$$

- We then split our covariance matrix $\mathbf{C}$ into the signal ($\mathbf{S}$) and noise ($\mathbf{N}$) contributions
  - $\mathbf{S}$ is best evaluated in harmonic-space (it's diagonal with C_ℓ values in appropriate places)
  - $\mathbf{N}$ is best evaluated in pixel-space (it's diagonal for uncorrelated noise)
- Hence, can use `map2alm` & `alm2map` from `HealPix` to rapidly transform our vector between pixel- and harmonic-space

Finite differnces to compute the Fisher matrix

- Under the Gaussian assumption, the Fisher matrix is given as

$$\mathbf{F}_{\ell_1\ell_2} = -\frac{\partial^2 \ln \mathcal{L}}{\partial\Theta_{\ell_1}\,\partial\Theta_{\ell_2}}, \tag{5}$$

  where Θ_ℓ are the power spectrum coefficients

- As a single derivative of the log-likelihood is given as

$$\frac{\partial \ln \mathcal{L}}{\partial\Theta_\ell} = s_\ell - b_\ell - \mathrm{Tr}[\mathbf{S}\,\mathbf{E}_\ell] \tag{6}$$

  we can use finite-differences to find the Fisher matrix as

$$F_{\ell\ell'}\,\Delta\Theta_{\ell'} = -\frac{1}{2}\left[\langle s_\ell(\Theta_{\text{fid}} + \Delta\Theta_{\ell'})\rangle - \langle s_\ell(\Theta_{\text{fid}})\rangle\right] \tag{7}$$

Results using my new estimator

Benchmarking my new estimator

We have benchmarked my new implementation against other public QML codes

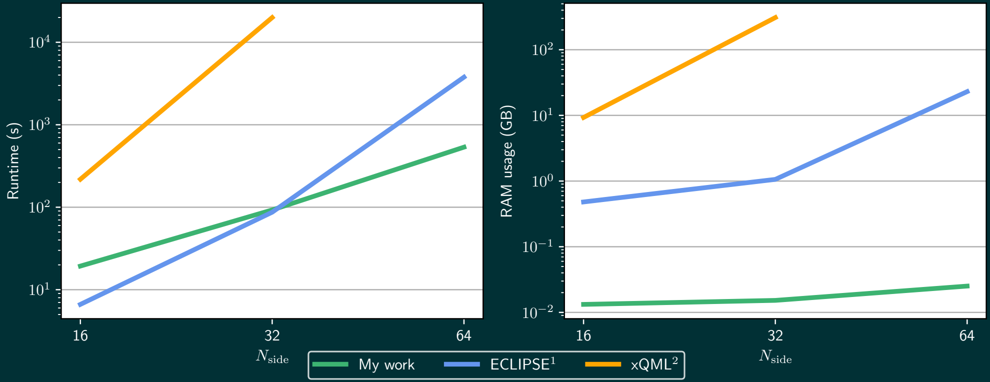

Figure 3: Timings and RAM usage for different QML implementations

1: `github.com/CosmoTool/ECLIPSE`

2: `gitlab.in2p3.fr/xQML/xQML`

Errors on the power spectrum

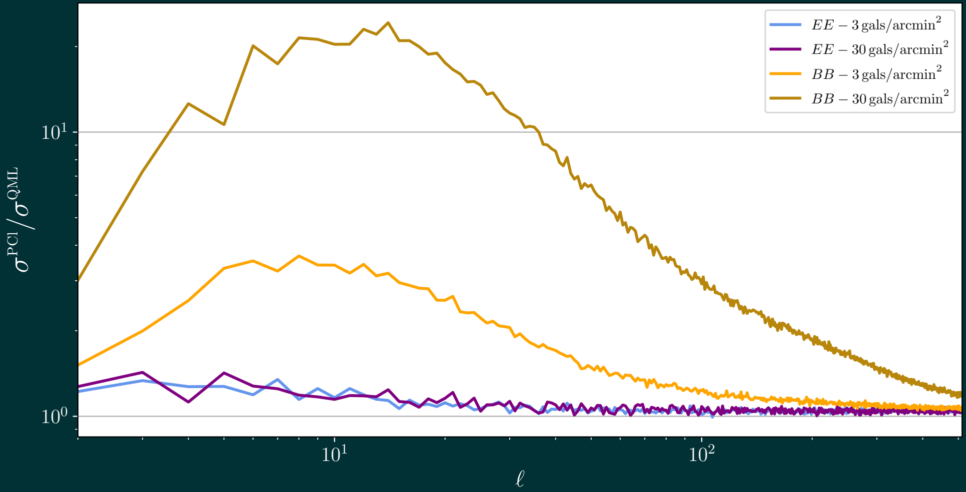

Figure 4: Ratio of Pseudo-C_ℓ to QML power spectrum errors for two different noise levels

Fisher forecasts

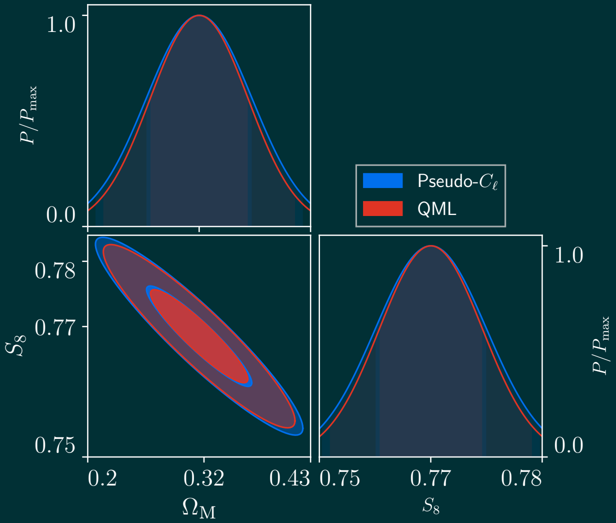

Figure 5: Parameter constraint comparisons between our two estimators

Conclusions

- We wanted implement and benchmark two power spectrum estimation techniques applied to a forthcoming Stage-IV weak lensing survey
- We found that the EE signal errors were reduced when using QML over Pseudo-C_ℓ
  - Also saw a large improvement in the errors for the low-ℓ B-modes
    - Provides promising constraints on B-mode physics for forthcoming surveys
- This had a small impact on cosmological parameter constraints.
  - Future analyses using Pseudo-C_ℓs should not induce any additional significant errors

Conclusions

- We wanted implement and benchmark two power spectrum estimation techniques applied to a forthcoming Stage-IV weak lensing survey
- We found that the EE signal errors were reduced when using QML over Pseudo-C_ℓ
  - Also saw a large improvement in the errors for the low-ℓ B-modes
    - Provides promising constraints on B-mode physics for forthcoming surveys
- This had a small impact on cosmological parameter constraints.
  - Future analyses using Pseudo-C_ℓs should not induce any additional significant errors

Thank you very much for attending my talk!