

QML activities

Alessandro Gruppuso
(INAF-OAS BO)

with M.Migliaccio (Roma ToV) and P.Natoli (UniFE)

Euclid CMBXC meeting
Bologna - 4 November 2019

Outline

- Status of the XCMB QML implementations.
- Related Fisher matrix activity
- Summary and next steps

QML method

$$\hat{C}_\ell = F_{\ell\ell'}^{-1} [x^t E_{\ell'} x - \text{Tr}[NE_{\ell'}]]$$

Sum on ℓ' is understood

total covariance, in pixel space

$$E_\ell = \frac{1}{2} C^{-1} P_\ell C^{-1}$$

matrix E, for each multipole pixel space

$$C = S(\{C_\ell^{fid}\}) + N$$

$$P_\ell = \frac{\partial C}{\partial C_\ell^{fid}}$$

matrix P in pixel space,

$$F_{\ell\ell'} = \frac{1}{2} \text{Tr} [C^{-1} P_\ell C^{-1} P_{\ell'}]$$

related to the Legendre polynomials

Fisher matrix, in harmonic space

x map, in pixel space

QML method

$y_{\ell'}$

vector, in harmonic space

$$\hat{C}_{\ell} = F_{\ell\ell'}^{-1} [x^t E_{\ell'} x - \text{Tr}[N E_{\ell'}]]$$

Sum on ℓ' is understood

total covariance, in pixel space

$$E_{\ell} = \frac{1}{2} C^{-1} P_{\ell} C^{-1}$$

matrix E, for each multipole pixel space

$$C = S(\{C_{\ell}^{fid}\}) + N$$

$$P_{\ell} = \frac{\partial C}{\partial C_{\ell}^{fid}}$$

matrix P in pixel space,

$$F_{\ell\ell'} = \frac{1}{2} \text{Tr} [C^{-1} P_{\ell} C^{-1} P_{\ell'}]$$

related to the Legendre polynomials

Fisher matrix, in harmonic space

x map, in pixel space

QML method

$y_{\ell'}$

vector, in harmonic space

$$\hat{C}_{\ell} = F_{\ell\ell'}^{-1} [x^t E_{\ell'} x - \text{Tr}[N E_{\ell'}]]$$

Sum on ℓ' is understood

This estimator is **optimal** since it is “unbiased and minimum variance”. They are minimum variance because they saturate the Fisher-Cramer-Rao inequality (under very general assumptions it is possible to show that the variance of a given estimator has a lower bound).

$$\langle \hat{C}_{\ell} \rangle = C_{\ell}$$

unbiased

$$\langle (\hat{C}_{\ell} - C_{\ell})(\hat{C}_{\ell'} - C_{\ell'}) \rangle = F_{\ell\ell'}^{-1}$$

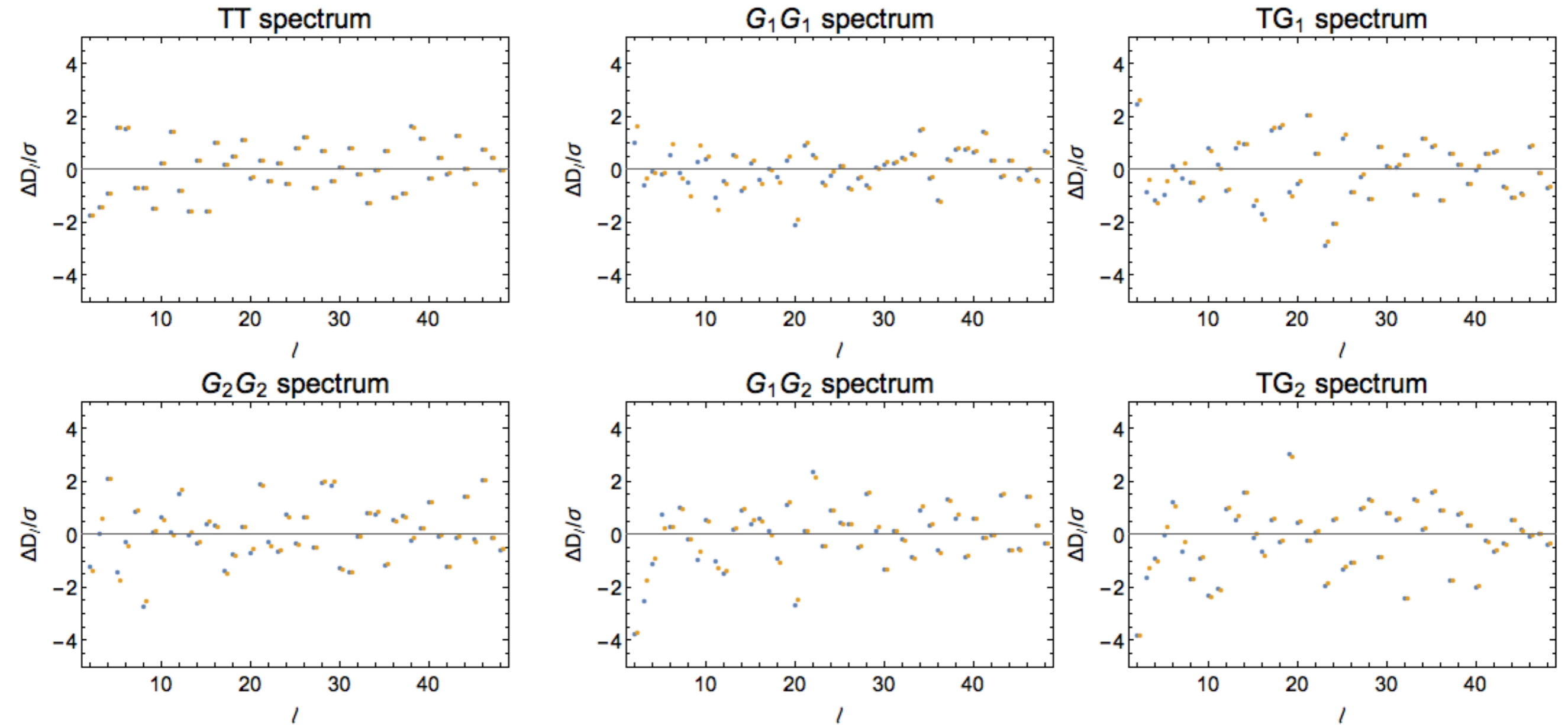
minimum variance

Status of the QML implementations

- Two CMBX QML implementations:
 - The first one is able to deal with one CMB map and two galaxy surveys. It provides six APS. Validated. Binning option. Computationally expensive and slow.
 - The second implementation is aimed at TG-only. Under development. It is lighter and faster.

First (slower) QML implementation

Validation (nside=16)



Distance of MC average (1000 sims) from the fiducial spectra
in units of standard deviation of the mean

Blue estimates = estimator built with the same
fiducial used for the MC

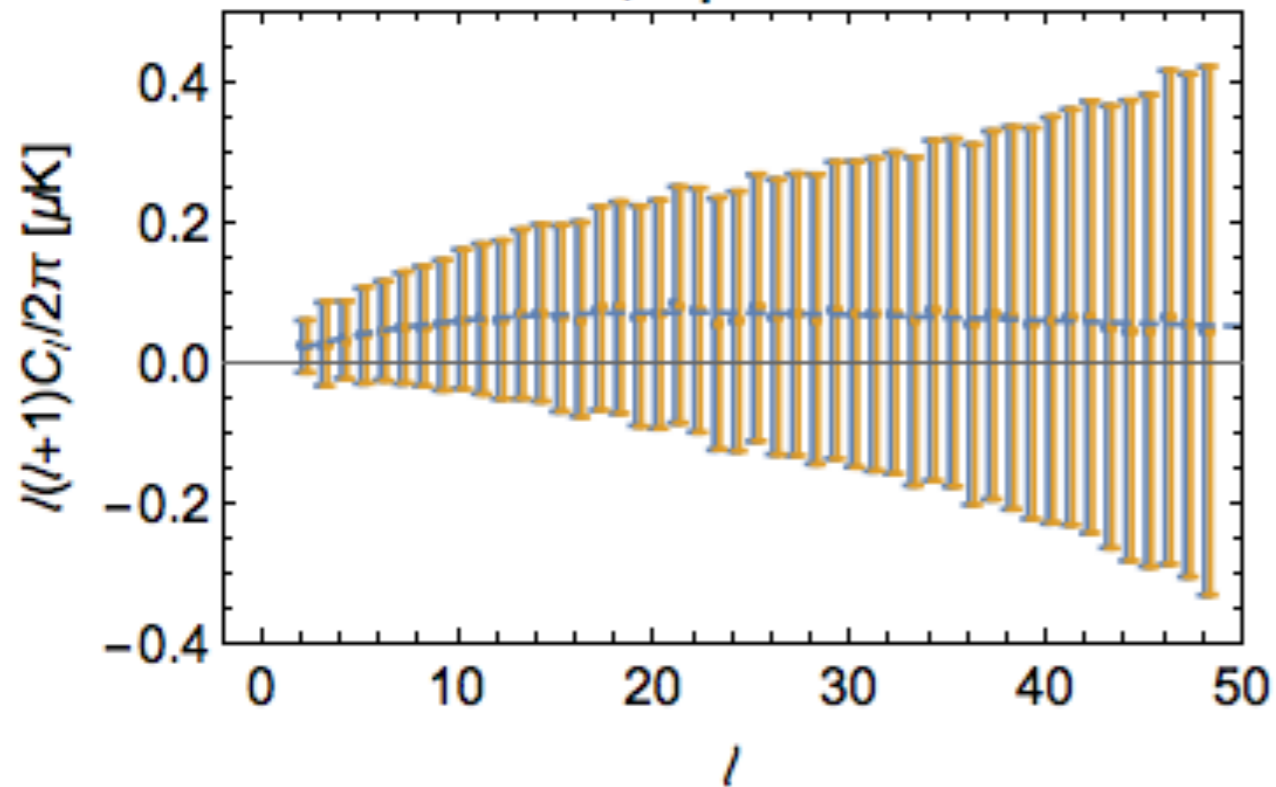
Orange estimates = estimator built with only
block diagonal covariances
(i.e. cross fiducial spectra set to zero)

fluctuations of the mean/st dev of mean

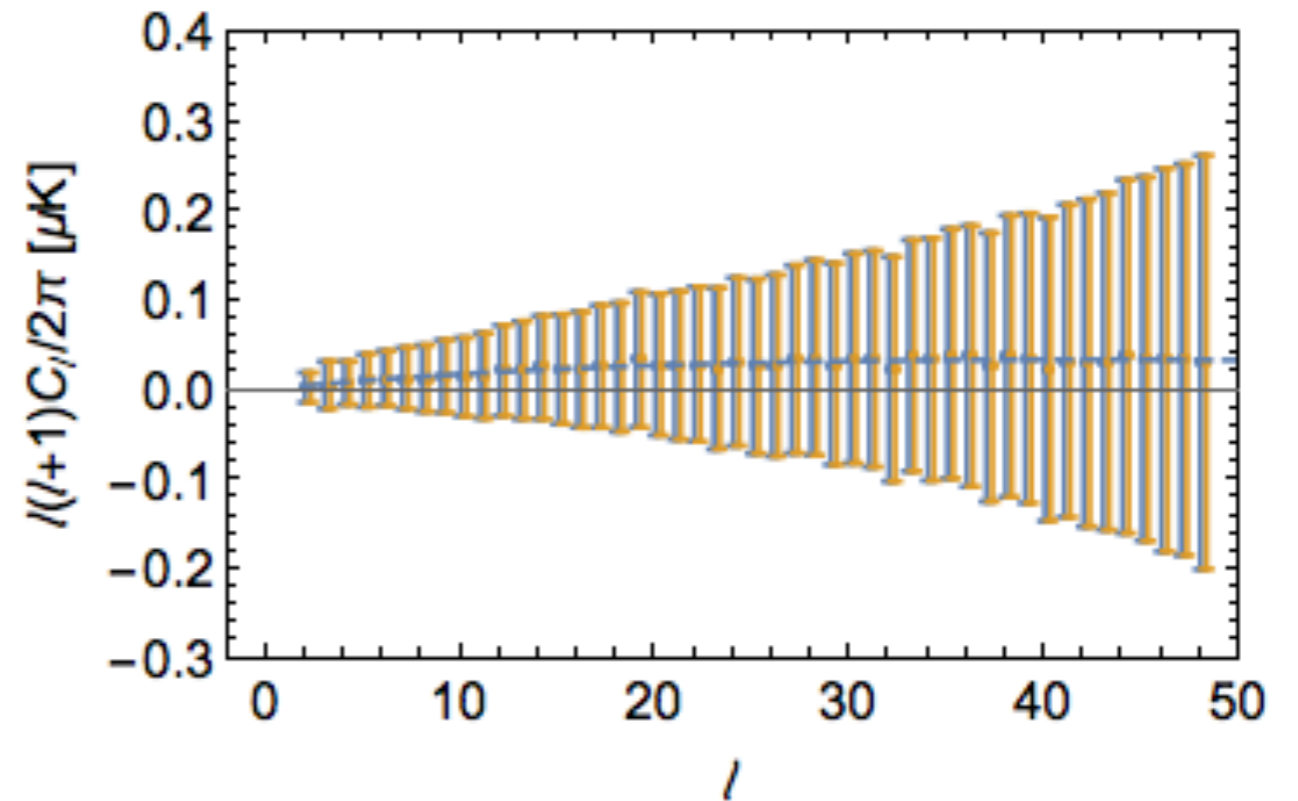
feature used to build the other implementation

Spectra (nside=16)/binning

TG₁ spectrum



TG₂ spectrum



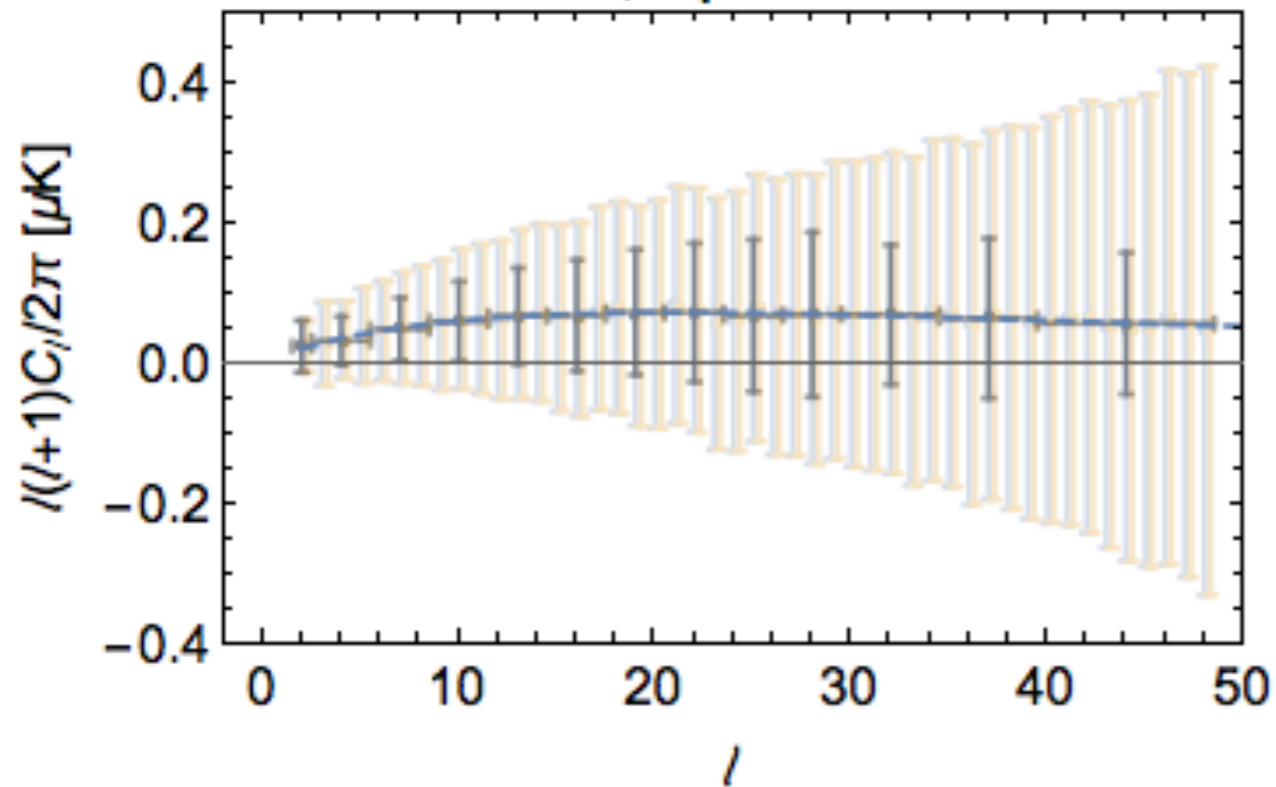
Blue estimates = estimator built with the same fiducial used for the MC

Error = Fisher matrix
(single realisation)

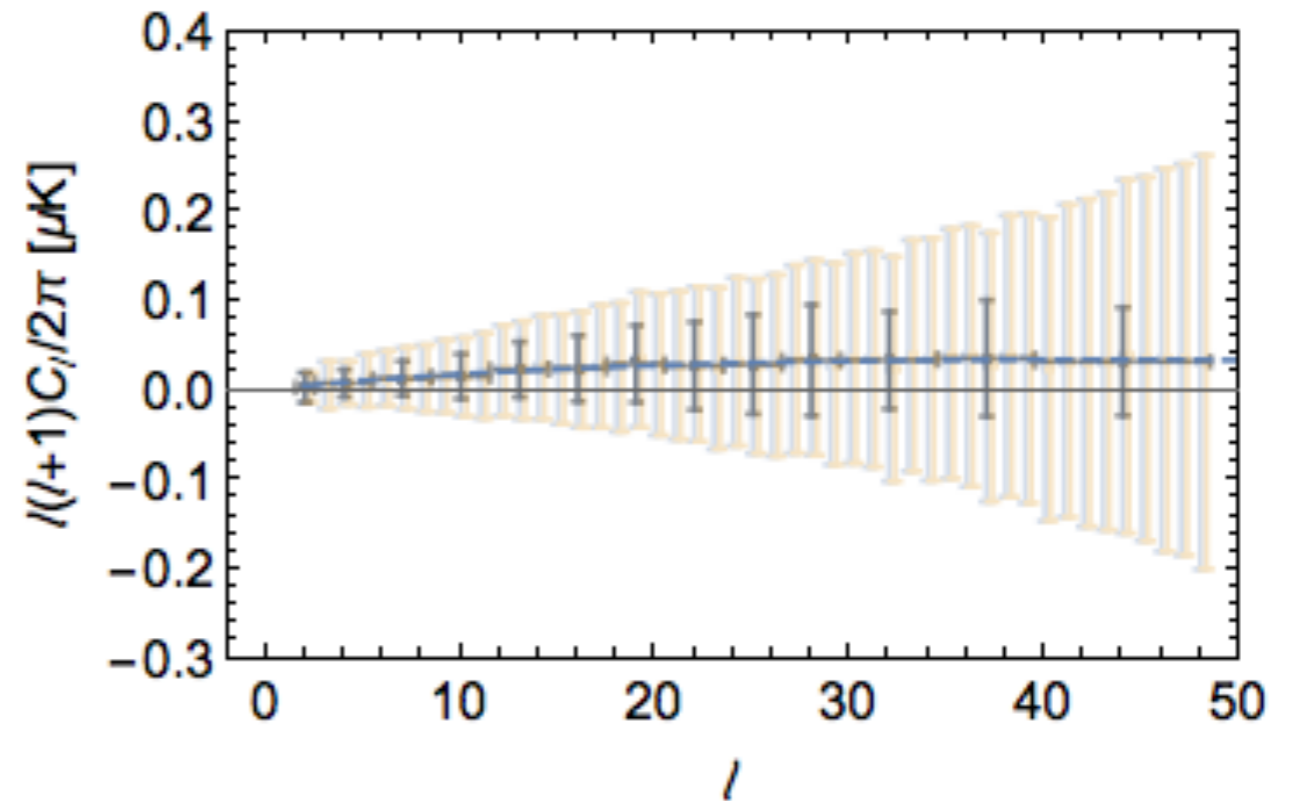
Orange estimates = estimator built with only block diagonal covariances (i.e. cross fiducial spectra set to zero)

Spectra (nside=16)/binning

TG₁ spectrum



TG₂ spectrum



Blue estimates = estimator built with the same fiducial used for the MC

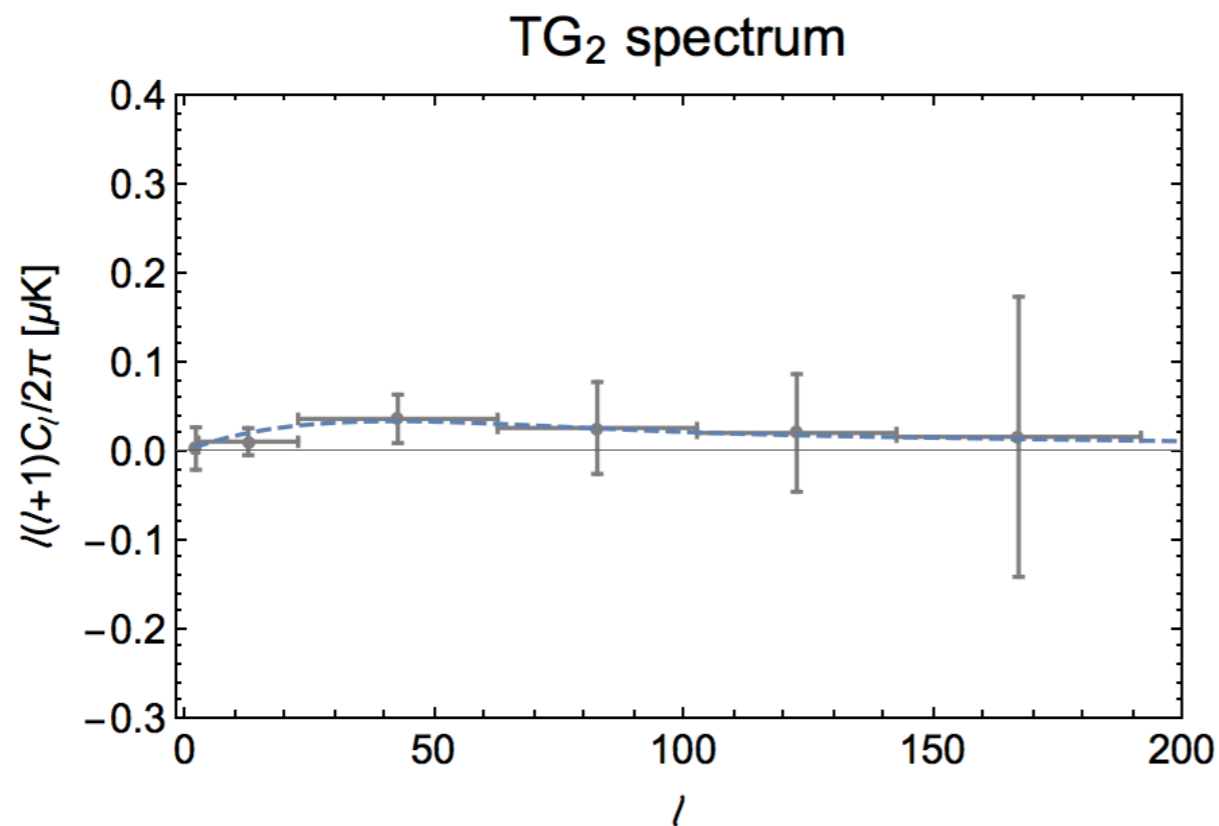
Error = Fisher matrix
(single realisation)

Orange estimates = estimator built with only block diagonal covariances
(i.e. cross fiducial spectra set to zero)

nside=64

Resolution we need for scientific exploitation

- Example at higher resolution, i.e. nside=64. This needs binning option active.



Aggressive binning

MC average of 100 sims
at nside 64. About 2h
running time with 900
processors (Edison,
NERSC)

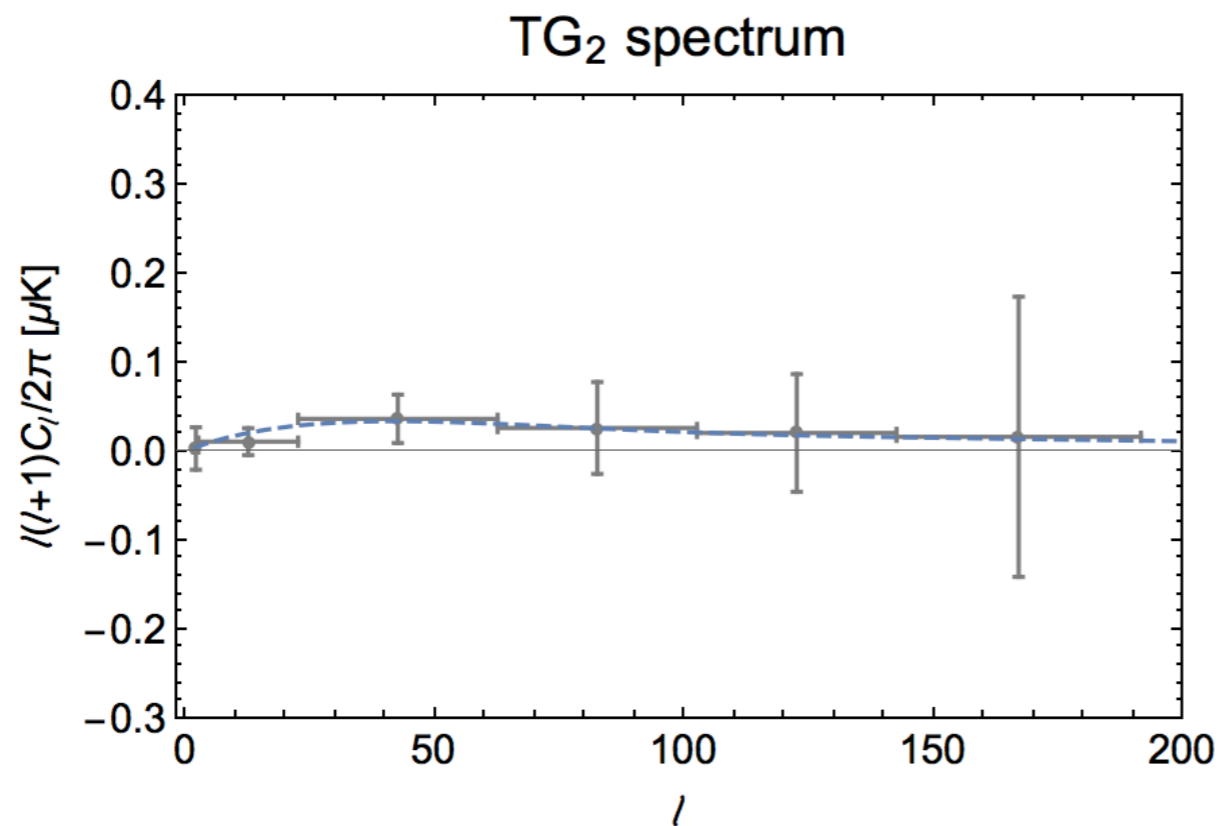
We need more bins

Not tried yet on Cori. Performances might be better.

nside=64

Resolution we need for scientific exploitation

- Example at higher resolution, i.e. nside=64. This needs binning option active.



MC average of 100 sims
at nside 64. About 2h
running time with 900
processors (Edison,
NERSC)

Aggressive binning

We need more bins

Having a less aggressive binning is too computationally expensive. For scientific exploitation we need a finer resolution in harmonic space. And also the comparison with other methods, e.g. pseudoCell, is not meaningful with such an aggressive binning.

=> we need a lighter implementation

A closer look at the heaviness of this implementation

CMB map and two galaxy surveys $x = (x_{CMB}, x_{G1}, x_{G2})$

QML algebra

$$\begin{pmatrix} \hat{C}_\ell^{TT} \\ \hat{C}_\ell^{TG_1} \\ \hat{C}_\ell^{G_1G_1} \\ \hat{C}_\ell^{G_2G_2} \\ \hat{C}_\ell^{TG_2} \\ \hat{C}_\ell^{G_1G_2} \end{pmatrix} = F_{\ell\ell'}^{-1} \begin{pmatrix} y_{\ell'}^{TT} \\ y_{\ell'}^{TG_1} \\ y_{\ell'}^{G_1G_1} \\ y_{\ell'}^{G_2G_2} \\ y_{\ell'}^{TG_2} \\ y_{\ell'}^{G_1G_2} \end{pmatrix}$$

explicitly written the elements of the vectors to make clear the dimensionality of the problem

192 elements for each multipole at nside=64

A closer look at the heaviness of this implementation

CMB map and two galaxy surveys $x = (x_{CMB}, x_{G1}, x_{G2})$
and the Fisher matrix is

$$F_{\ell\ell'} = \begin{pmatrix} TT, TT & TT, TG_1 & TT, G_1G_1 & TT, G_2G_2 & TT, TG_2 & TT, G_1G_2 \\ & TG_1, TG_1 & TG_1, G_1G_1 & TG_1, G_2G_2 & TG_1, TG_2 & TG_1, G_1G_2 \\ & & G_1G_1, G_1G_1 & G_1G_1, G_2G_2 & G_1G_1, TG_2 & G_1G_1, G_1G_2 \\ & & & G_2G_2, G_2G_2 & G_2G_2, TG_2 & G_2G_2, G_1G_2 \\ & & & & TG_2, TG_2 & TG_2, G_1G_2 \\ & & & & & G_1G_2, G_1G_2 \end{pmatrix}$$

symmetric

where each entry of the matrix stands for

$$F_{\ell\ell'}^{XY} = \frac{1}{2} \text{Tr} [C^{-1} P_{\ell}^X C^{-1} P_{\ell'}^Y]$$

$$X, Y = TT, TG_1, G_1G_1, G_2G_2, TG_2, G_1G_2$$

A closer look at the heaviness of this implementation

CMB map and two galaxy surveys $x = (x_{CMB}, x_{G1}, x_{G2})$
and the Fisher matrix is

$$F_{\ell\ell'} = \begin{pmatrix} TT, TT & TT, TG_1 & TT, G_1G_1 & TT, G_2G_2 & TT, TG_2 & TT, G_1G_2 \\ & TG_1, TG_1 & TG_1, G_1G_1 & TG_1, G_2G_2 & TG_1, TG_2 & TG_1, G_1G_2 \\ & & G_1G_1, G_1G_1 & G_1G_1, G_2G_2 & G_1G_1, TG_2 & G_1G_1, G_1G_2 \\ & & & G_2G_2, G_2G_2 & G_2G_2, TG_2 & G_2G_2, G_1G_2 \\ & & & & TG_2, TG_2 & TG_2, G_1G_2 \\ & & & & & G_1G_2, G_1G_2 \end{pmatrix}$$

0.2 GB at nside 16 (our masked case)
symmetric
54 GB at nside 64 (our masked case)

where each entry of the matrix
stands for

$$F_{\ell\ell'}^{XY} = \frac{1}{2} \text{Tr} [C^{-1} P_{\ell}^X C^{-1} P_{\ell'}^Y]$$

$X, Y = TT, TG_1, G_1G_1, G_2G_2, TG_2, G_1G_2$

Second (faster) QML implementation

Second (faster) QML implementation

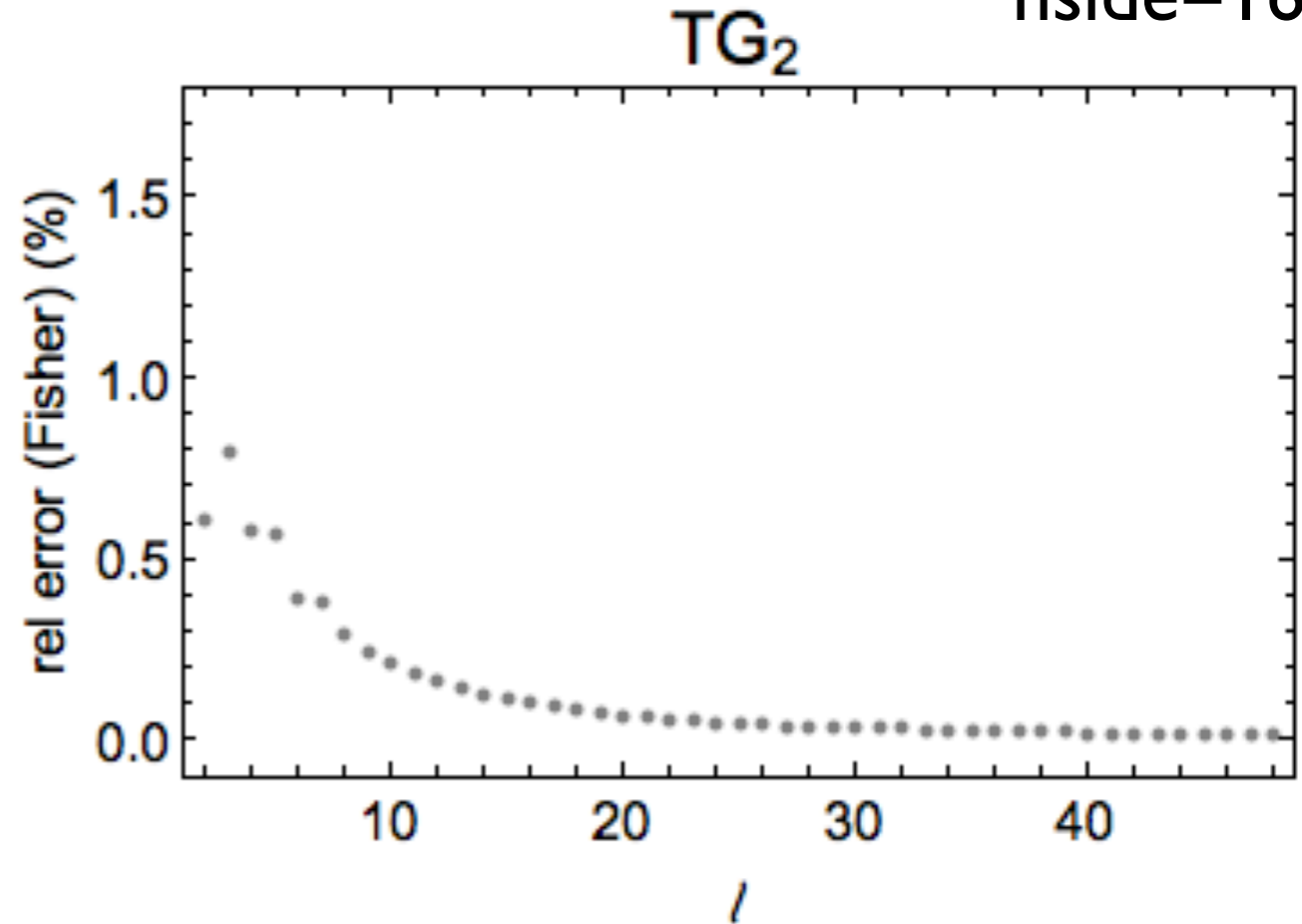
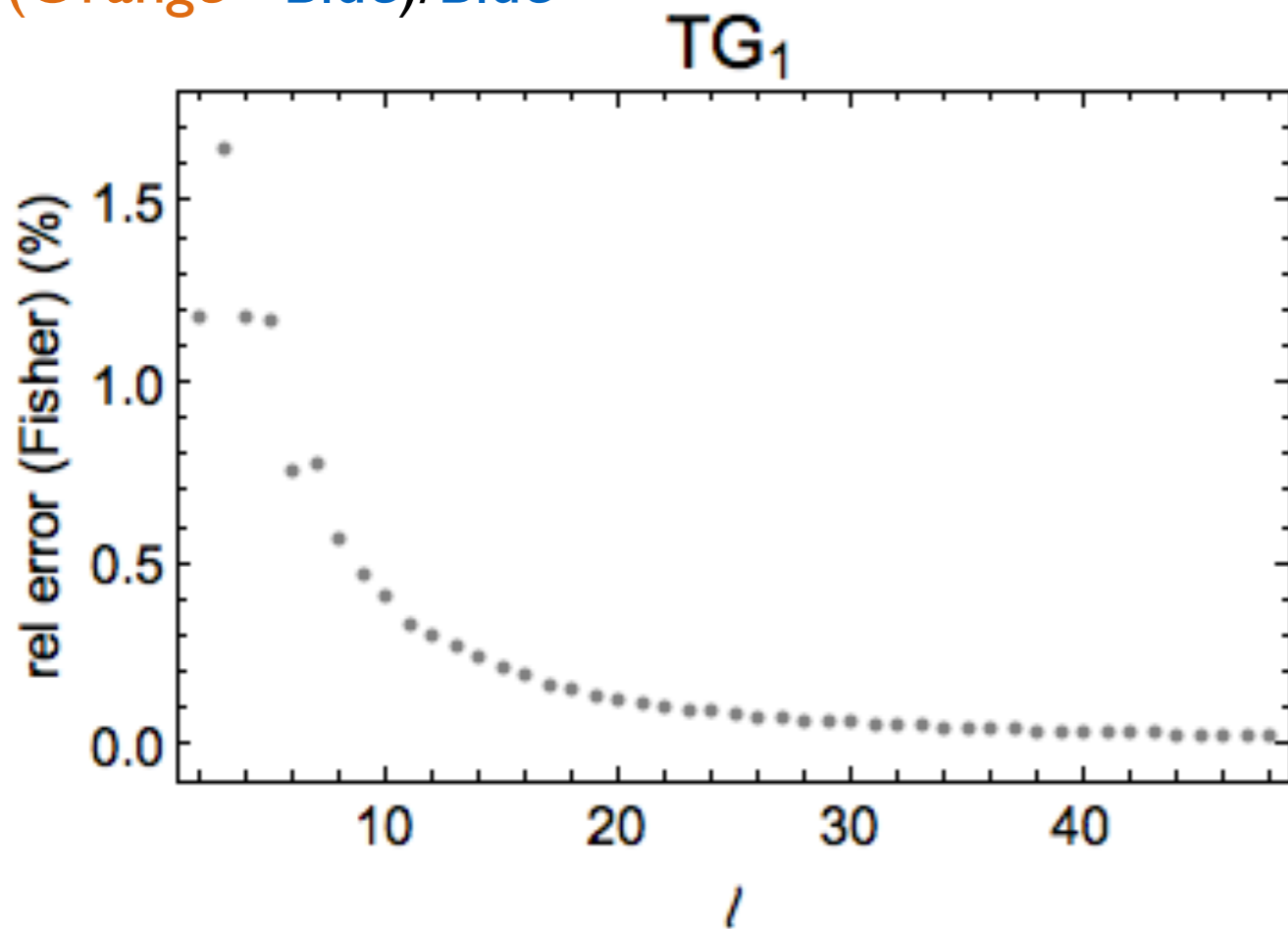
- This is a CMBX QML estimator which is aimed at providing the estimates of only TG.
- Disentangling TG from the other spectra is algebraically possible after setting cross-fiducial to zero (and selecting the spectrum of interest).
- The method with this choice, is still unbiased (shown before) and the increase of uncertainty is expected to be small.

Impact on Fisher uncertainties

when we consider or not fiducial cross spectra

“(Orange - Blue)/Blue“

inside=16



When the fiducial is not exactly the one used generate the MC, the QML is not exactly optimal anymore. This quantifies the increase of the error in percentage for each multipole when we use this simplification. Note that if the peak of the cross correlation is not below ~ 10 the impact might be mild. However this has to be quantified at the level of cosmological parameters.

Disentangling TG

Example: case of a CMB map and one Galaxy survey

QML algebra

$$x = (x_{CMB}, x_G)$$

and the covariance matrix is

$$\begin{pmatrix} \hat{C}_\ell^{TT} \\ \hat{C}_\ell^{TG} \\ \hat{C}_\ell^{GG} \end{pmatrix} = F_{\ell\ell'}^{-1} \begin{pmatrix} y_{\ell'}^{TT} \\ y_{\ell'}^{TG} \\ y_{\ell'}^{GG} \end{pmatrix}$$

$$C = \langle xx^t \rangle = \begin{pmatrix} C_{CMB} & C_{TG} \\ C_{TG}^t & C_G \end{pmatrix}$$

and the Fisher matrix is

$$F_{\ell\ell'} = \begin{pmatrix} TTTT & TTTG & TT\ GG \\ TGT T & TGTG & TG\ GG \\ GGTT & GGTG & GG\ GG \end{pmatrix}$$

where each entry of the matrix stands for

$$F_{\ell\ell'}^{XY} = \frac{1}{2} \text{Tr} [C^{-1} P_\ell^X C^{-1} P_{\ell'}^Y]$$

$$X, Y = TT, TG, GG$$

Disentangling TG

Example: case of a CMB map and one Galaxy survey

QML algebra

$$x = (x_{CMB}, x_G)$$

and the covariance matrix is

$$\begin{pmatrix} \hat{C}_\ell^{TT} \\ \hat{C}_\ell^{TG} \\ \hat{C}_\ell^{GG} \end{pmatrix} = F_{\ell\ell'}^{-1} \begin{pmatrix} y_{\ell'}^{TT} \\ y_{\ell'}^{TG} \\ y_{\ell'}^{GG} \end{pmatrix}$$

$$C = \langle xx^t \rangle = \begin{pmatrix} C_{CMB} & \text{X} \\ \text{X} & C_G \end{pmatrix}$$

$C_{\ell}^{fid, TG} = 0$

and the Fisher matrix is

where each entry of the matrix stands for

$$F_{\ell\ell'} = \begin{pmatrix} TTTT & \text{X} & \text{X} \\ \text{X} & TGTG & \text{X} \\ \text{X} & \text{X} & GG GG \end{pmatrix}$$

$$F_{\ell\ell'}^{XY} = \frac{1}{2} \text{Tr} [C^{-1} P_\ell^X C^{-1} P_{\ell'}^Y]$$

$$X, Y = TT, TG, GG$$

Disentangling TG

Example: case of a CMB map and one Galaxy survey

THREE INDEPENDENT QMLS

$$\hat{C}_\ell^{TT} = \left(F_{\ell\ell'}^{TT,TT} \right)^{-1} y_{\ell'}^{TT}$$

$$\hat{C}_\ell^{TG} = \left(F_{\ell\ell'}^{TG,TG} \right)^{-1} y_{\ell'}^{TG}$$

$$\hat{C}_\ell^{GG} = \left(F_{\ell\ell'}^{GG,GG} \right)^{-1} y_{\ell'}^{GG}$$

and the Fisher matrix is

$$F_{\ell\ell'} = \begin{pmatrix} TTTT & \text{~~TTGG~~} & \text{~~TTGG~~} \\ \text{~~TTTT~~} & TGTG & \text{~~TGGG~~} \\ \text{~~GGTT~~} & \text{~~GGTG~~} & GG GG \end{pmatrix}$$

$$x = (x_{CMB}, x_G)$$

and the covariance matrix is

$$C = \langle xx^t \rangle = \begin{pmatrix} C_{CMB} & \text{~~C}_{TG} \\ \text{~~C}_{TG} & C_G \end{pmatrix}~~~~$$

$C_{\ell}^{fid,TG} = 0$

where each entry of the matrix stands for

$$F_{\ell\ell'}^{XY} = \frac{1}{2} Tr [C^{-1} P_{\ell}^X C^{-1} P_{\ell'}^Y]$$

$$X, Y = TT, TG, GG$$

Second (faster) QML implementation

$$\hat{C}_\ell^{TG} = F_{\ell\ell'}^{-1} y_{\ell'}$$

$$F_{\ell\ell'} = \text{Tr} \left[C_{TT}^{-1} P_\ell C_{GG}^{-1} (P_{\ell'})^t \right] \quad y_\ell = m_T^t C_{TT}^{-1} P_\ell C_{GG}^{-1} m_G$$

Second (faster) QML implementation

$$\hat{C}_\ell^{TG} = F_{\ell\ell'}^{-1} y_{\ell'}$$

8GB at nside 64
(our masked case)

$$F_{\ell\ell'} = \text{Tr} \left[C_{TT}^{-1} P_\ell C_{GG}^{-1} (P_{\ell'})^t \right] \quad y_\ell = m_T^t C_{TT}^{-1} P_\ell C_{GG}^{-1} m_G$$

3GB at nside 64 (our masked case)

14GB at nside 64 (our masked case)

Fewer elements to be computed
and each of them are lighter (wrt 54GB estimated before)

Second (faster) QML implementation

Some details

- It computes the Fisher matrix separately. It might be useful because such an object represents the bottleneck of the whole computation. It can be read if already computed.
- Also the parallelisation strategy is new: different processors compute different multipoles (or groups of multipoles)

Second (faster) QML implementation

Preliminary results

- The code is not validated yet, but we have some preliminary (and interesting) results:
 - `nside=16`: the Fisher matrix can be computed with 5 nodes (160 processors on CORI) in 30”.
 - `nside=32`: the Fisher matrix can be computed with 30 nodes (640 active processors on CORI, more blocked for memory, still in debug queue) 23 minutes.
 - `nside=64`: still missing it but ~2h.

without any binning option active

Fisher matrix

new activity we started recently

- Fisher matrix is needed in the QML algebra and plays the role of a kernel. It is also the bottleneck of the computation.
- It represents the (analytic) covariance of the estimates.
- Therefore the QML method beyond providing the estimates with related uncertainties, offers the possibility to provide their covariance which might be also relevant for a Likelihood code for TG.

At least it is a natural candidate to be considered.

Fisher matrix

- In fact, in the TG-only version method the kernel is a very good approximation (but still an approximation) of the covariance of the estimates.
It impacts the lowest multipoles
- Therefore we started a new activity aiming at computing the corrections which switch on when the fiducial of TG is “small” but not exactly zero.
- The same can be done for the case of two galaxy surveys in order to evaluate the level of correlation we expect between TG1 and TG2, for example.
- Currently we are adopting an analytical approach. We wish to provide these analytical corrections.

Summary and next steps

- A CMBX QML code has been validated and put to work (it deals with one CMB map and two galaxy counts maps). Slow at $n_{\text{side}}=64$ if we wish to have more than ~ten bins in multipoles space. **Still working on optimisations**
- A QML-TGonly version has been implemented with a new parallelisation strategy. Validation and further optimisation to be done, but preliminary performances are very good. This will be used to make comparisons with pseudoCell estimators.
- Analytical computation of the correction terms in the Fisher matrix beyond the choice of fiducial $TG = 0$. This might be relevant for accurate likelihood modeling.
We need to evaluate whether these terms are useful or sufficient when we build a likelihood