MLPSE: documentation

April 25, 2022

1 Likelihood

$$L = (2\pi)^{N/2} \left[\det(C)\right]^{-1/2} \exp\left[-\frac{1}{2}x^{\dagger}C^{-1}x\right] \tag{1}$$

2 Log-likelihood

The log-likelihood is defined as

$$\mathcal{L} = -2\ln L \tag{2}$$

such that maximizing L is the same as minimizing $\mathcal{L}$. Ignoring the constant term in $\mathcal{L}$, it derives

$$\mathcal{L} = \ln\left[\det\left(C\right)\right] + \mathrm{Tr}(C^{-1}D) \tag{3}$$

where $D = xx^{\dagger}$.

3 Parameterisation

We use a set of band power $\{p_\alpha\}$ to parameterise the covariance matrix C such that

$$C = \sum_{\alpha} p_\alpha Q_\alpha + N, \tag{4}$$

where p_α are scalar parameters and Q_α are the response matrices which can be understood as the covariance matrices of unit cosmological band powers. N is the noise (instrumental noise + foregrounds) covariance matrix.

4 Perturbations

Perturb $\mathcal{L}$ by making variations on the parameters. Then the perturbation on the covariance matrix reads

$$\delta C = \sum_{\alpha} \delta p_{\alpha} Q_{\alpha} \tag{5}$$

The perturbation on the log-likelihood in terms of δC is

$$\delta \mathcal{L} = \mathrm{Tr}\left[C^{-1}\delta C(I - C^{-1}D)\right] + \mathrm{Tr}\left[C^{-1}\delta C C^{-1}\delta C(C^{-1}D - \frac{1}{2}I)\right] + O\left(\delta C^3\right) \tag{6}$$

Rewrite the perturbation in terms of δp_{α}:

$$\begin{aligned} \delta \mathcal{L} = &\sum_{\alpha} \delta p_{\alpha} \,\mathrm{Tr}\left[C^{-1}Q_{\alpha}(I - C^{-1}D)\right] \\ &+ \sum_{\alpha,\beta} \delta p_{\alpha}\delta p_{\beta} \,\mathrm{Tr}\left[C^{-1}Q_{\alpha}C^{-1}Q_{\beta}(C^{-1}D - \frac{1}{2}I)\right] \\ &+ O\left(\delta p^3\right) \end{aligned} \tag{7}$$

The above expression has given the first and second order perturbations, which are

$$\text{1st order derivatives (Jacobian): } J_{\alpha} = \mathrm{Tr}\left[C^{-1}Q_{\alpha}(I - C^{-1}D)\right] \tag{8}$$

$$\text{2nd order derivatives (Hessian): } H_{\alpha\beta} = 2\,\mathrm{Tr}\left[C^{-1}Q_{\alpha}C^{-1}Q_{\beta}(C^{-1}D - \frac{1}{2}I)\right] \tag{9}$$

5 M-mode formalism

Statistical homogeneity, or translational invariance, of a field implies that different Fourier modes are uncorrelated. Here in our case, we know that the statistics of the cosmological field is time transnational invariant, as we assume statistically isotropic cosmology. Thus, different m modes, as Fourier modes of the time stream (with the FG removed), independently follow the statistics described above. So the joint likelihood function is just the product of all m-mode likelihoods. Similarly, the total log-likelihood, as well as the corresponding Jocobian and Hessian, is just the sum of all m modes:

$$\mathcal{L} = \sum_m \mathcal{L}^{(m)} \tag{10}$$

$$J_\alpha = \sum_m J_\alpha^{(m)} \tag{11}$$

$$H_{\alpha\beta} = \sum_m H_{\alpha\beta}^{(m)} \tag{12}$$

6 General grasp

1. Inputs:

   Just to fill the lines 12-18.

   "configfile": the yaml file you used in driftscan

   "data_path": The path to the data in K-L basis, which can be generated by draco.

   "kpar_start, kpar_end, kpar_dim, kperp_start, kperp_end, kperp_dim": self-illustrated k space settings.

   "kltrans_name": the name of the K-L product, which should appear on your yaml file. Example: "dk_1thresh_fg_3thresh".

   "outputname": the name of the output file.

2. Outputs:

   In ".HDF5" format with following keywords:

   "ps_result" : numpy array of shape (n,), where n is the number of total bands. Band power.

   "k_parallel" : numpy array of shape (n,), where n is the number of total bands. The $k_\parallel$ of corresponding band power.

   "k_perp" : numpy array of shape (n,), where n is the number of total bands. The $|\mathbf{k}_\perp|$ of corresponding band power.

   "k_centers" : numpy array of shape (n,), where n is the number of total bands. The $|\mathbf{k}|$ of corresponding band power.

“first guess” : The initial guess of the optimization.

“theory” : Theoretical guess of the results given by CORA.

3. Scaling: $\mathcal{L}(p) = \mathcal{L}(\cosh x - 1)$

   It is strongly recommended to turn this option on!

   The original variables are $(k_{\|}, k_{\perp})$-band power parameters, p_α, which are positive real numbers.

   One can choose to rescale the function by substitute p_α with $\cosh x - 1$. Now the domain with respect to x becomes the whole real region. Furthermore, the optimization converges faster since the scaling also improves the convexity of the optimized function.

   In the case that the solution of some parameter p_* is very small, x numerically approximates 0, subsequently $\cosh x - 1$ also approximates 0. So the numerical estimate of p_* can also approximate the real solution.

4. Regularization: $\mathcal{L}' = \mathcal{L}(\cosh x - 1) + |x|$

   Optional.

   It can push those extremely flat variables to 0. So you can notice these bands and filter them out in your analyses.

   Its impact on other non-trivial parameters should be very tiny. One can also use Fisher matrix to evaluate the impact.

5. Detailed explanation about why regularisation.

   - The variables p_α are band power of the cosmological signal in the $(k_{\|}, k_{\perp})$ space.
   - The covariance matrices are first constructed in sky space, and then projected on KL basis, where the residual noise covariance should become very small.
   - Owing to the flat sky approximation, we can sum over these $k-$space terms to get the angular power spectrum $C_l(\nu_1, \nu_2)$.
   - Conceptually, low frequency and high $k_\perp$ modes correspond to high multipole angular power spectrum. So a reasonable binning of the $k-$space is relevant to the choice of l_{max} and the frequency range.

- An example for the above point: bands with $k_\perp = 0.15\text{MPC}^{-1}$ contribute nothing to $C_l(\nu_1, \nu_2)$ for $l \leq 200$ and $\nu < 800$MHz. Thus the corresponding response matrices are zero. These zero response bands are filtered in the program.
- There could also be some bands with extremely small response. They will make the log-likelihood function extremely flat with respect to those band power parameters. We can also do regularization to avoid problems in convergence and for flagging these parameters. The band power of these trivial modes will be pushed towards 1 if rescaled, or to 0 if not rescaled.
- mpi+openmpi setting: I recommend you use as many parallel processes as the number of m modes, if possible. Number of m-modes per process affects the running time a lot.