On the Power Spectrum Estimator
i. Maximum Likelihood   ii. Quadratic

Zheng Zhang

Abstract

This note reviews power spectrum estimation with maximum likelihood (Jacobian/Hessian formalism), with an explicit focus on error propagation and covariance matrix construction, and extends methods for quadratic power spectrum estimation, including context-dependent optimal quadratic estimators.

Contents

1 Maximum likelihood power spectrum estimation

1.1 ML Estimate

1.1.1 Likelihood function

- The likelihood function reads

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

  where x is the data vector and C is the covariance matrix of the data vector.

- For convenience we define the log-likelihood 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 $D = xx^\dagger$.

1.1.2 Parameterisation with band power

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.

1.1.3 Jacobian and Hessian

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

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

Rewrite the perturbation in terms of δp_α:

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

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

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

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

1.1.4 Maximum likelihood estimate

The maximum likelihood estimate of the band power values can be obtained using the derived Jacobian and Hessian. A general tip is to make the most of the block diagonal structure in C (e.g., m-mode visibility of drift-scan surveys - assuming the m-homogeneous statistics), then different additive terms in the log-likelihood can be calculated in parallel.

1.2 Error in the estimation

How accurately can we estimate model parameters from a given data set? Fisher answered in 1935. He laid the foundation for understanding the accuracy of parameter estimation through the concept of Fisher information, which quantifies the amount of information a dataset provides about unknown model parameters. Here are some key concepts:

- **Fisher Information Matrix (FIM)**:

$$I(\theta)_{ij} = -\mathbb{E}\left[\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log f(X;\theta)\right]$$

Alternatively expressed as the covariance of the score function:

$$I(\theta) = \mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\log f(X;\theta)\right)\left(\frac{\partial}{\partial\theta}\log f(X;\theta)\right)^{\top}\right]$$

- **Cramér-Rao Lower Bound (CRLB)**: For any unbiased estimator $\hat{\theta}$:

  $$\mathrm{Var}(\hat{\theta}) \geq \frac{1}{I(\theta)}$$

  For multiple parameters, the covariance matrix satisfies:

  $$\mathrm{Cov}(\hat{\theta}) \succcurlyeq I^{-1}(\theta)$$

  where $\succcurlyeq$ means that the matrix $\mathrm{Cov}(\hat{\theta}) - I^{-1}(\theta)$ is positive semi-definite.

- **Maximum Likelihood Estimation (MLE)**: Asymptotic properties:

  $$\hat{\theta}_{\mathrm{MLE}} \xrightarrow[n\to\infty]{d} \mathcal{N}\left(\theta,\, I^{-1}(\theta)\right)$$

  where $\xrightarrow{d}$ denotes convergence in distribution. In other words, MLE achieves the CRLB asymptotically under regularity conditions.

- **Accuracy and Sample Size**: Fisher information scales linearly with sample size n for independent data:

  $$I(\theta) \propto n \quad \Rightarrow \quad \mathrm{SE}(\hat{\theta}) \propto \frac{1}{\sqrt{n}}$$

  where 'SE' stands for 'Standard Error'.

1.2.1 Error in band power

Calculating the curvature or Hessian could be numerically cumbersome, whereas we can effectively calculate the quadratic form of the first-order derivatives instead. To prove this, we start with the normalisation condition

$$\int \mathcal{L}(\boldsymbol{x};\boldsymbol{p})\,\mathrm{d}^{\mathrm{n}}x = 1, \tag{10}$$

where $\boldsymbol{p} = p_1, \ldots, p_m$. Differentiating this equation gives

$$\int \frac{\partial \mathcal{L}(\boldsymbol{x};\boldsymbol{p})}{\partial p_i} \,\mathrm{d}^{\mathrm{n}}x = 0 \tag{11}$$

which can be rewritten as

$$\int \frac{\partial \ln \mathcal{L}}{\partial p_i}\mathcal{L} \,\mathrm{d}^{\mathrm{n}}x = 0 \quad \Leftrightarrow \quad \left\langle \frac{\partial \ln \mathcal{L}(\boldsymbol{x};\boldsymbol{p})}{\partial p_i} \right\rangle = 0. \tag{12}$$

Differentiating it one more time we obtain

$$\int \frac{\partial^2 \ln \mathcal{L}}{\partial p_i \partial p_j}\mathcal{L} + \frac{\partial \ln \mathcal{L}}{\partial p_i}\frac{\partial \ln \mathcal{L}}{\partial p_j}\mathcal{L} \,\mathrm{d}^{\mathrm{n}}x = 0 \tag{13}$$

or equivalently

$$\left\langle \frac{\partial^2 \ln \mathcal{L}}{\partial p_i \partial p_j} \right\rangle = -\left\langle \frac{\partial \ln \mathcal{L}}{\partial p_i}\frac{\partial \ln \mathcal{L}}{\partial p_j} \right\rangle. \tag{14}$$

This equation states that the expected curvature of the likelihood function is equivalent to the expected value of the quadratic form of the first order derivatives.

1.2.2 Error in derived parameter

Since the model parameters (say the band power values) fully characterise the statistics, any quantity derived from the statistics can be estimated by viewing it as a function of these parameters.

Suppose the set of quantities to be estimated is $\boldsymbol{\Omega}(\boldsymbol{p})$, for which $\hat{\Omega}$ is the set of unbiased estimators, which means

$$\left\langle \hat{\Omega} \right\rangle = \int \hat{\Omega}\mathcal{L} \,\mathrm{d}^{\mathrm{n}}x = \boldsymbol{\Omega}(\boldsymbol{p}). \tag{15}$$

Differentiating this equation gives

$$\frac{\partial \boldsymbol{\Omega}(\boldsymbol{p})}{\partial p_i} = \int \hat{\Omega}\frac{\partial \mathcal{L}}{\partial p_i} \,\mathrm{d}^{\mathrm{n}}x = \int \hat{\Omega}\frac{\partial \ln \mathcal{L}}{\partial p_i}\mathcal{L} \,\mathrm{d}^{\mathrm{n}}x = \left\langle \hat{\Omega}\frac{\partial \ln \mathcal{L}}{\partial p_i} \right\rangle. \tag{16}$$

There is no harm to rewrite this equation as

$$\frac{\partial \boldsymbol{\Omega}(\boldsymbol{p})}{\partial p_i} = \left\langle (\hat{\Omega} - \boldsymbol{\Omega})\frac{\partial \ln \mathcal{L}}{\partial p_i} \right\rangle \tag{17}$$

because $\left\langle \mathbf{\Omega} \frac{\partial \ln \mathcal{L}}{\partial p_i} \right\rangle = \mathbf{\Omega} \left\langle \frac{\partial \ln \mathcal{L}}{\partial p_i} \right\rangle = 0$.

To establish a lower bound on the covariance of the estimators, we compute

$$\frac{\partial \mathbf{\Omega}}{\partial p_i} \frac{\partial \mathbf{\Omega}}{\partial p_j} = \left\langle (\hat{\Omega} - \mathbf{\Omega}) \frac{\partial \ln \mathcal{L}}{\partial p_i} \right\rangle \left\langle (\hat{\Omega} - \mathbf{\Omega}) \frac{\partial \ln \mathcal{L}}{\partial p_j} \right\rangle \tag{18}$$

Remembering the Cauchy Inequality that the inner product is no greater than the product of the modules, i.e.,

$$(f \cdot g)^2 \leq f^2 g^2, \tag{19}$$

or explicitly,

$$\left(\int f g \mathcal{L} \, \mathrm{d}^{\mathrm{n}} x \right)^2 \leq \left(\int f^2 \mathcal{L} \, \mathrm{d}^{\mathrm{n}} x \right) \left(\int g^2 \mathcal{L} \, \mathrm{d}^{\mathrm{n}} x \right), \tag{20}$$

thus

$$\int f g \mathcal{L} \, \mathrm{d}^{\mathrm{n}} x \leq \sqrt{\int f^2 \mathcal{L} \, \mathrm{d}^{\mathrm{n}} x \int g^2 \mathcal{L} \, \mathrm{d}^{\mathrm{n}} x}, \tag{21}$$

which then reads

$$\langle f g \rangle \leq \sqrt{\langle f^2 \rangle \langle g^2 \rangle}. \tag{22}$$

Now we can obtain a lower bound for the estimator covariance:

$$\begin{aligned} \frac{\partial \mathbf{\Omega}}{\partial p_i} \frac{\partial \mathbf{\Omega}}{\partial p_j} &= \left\langle (\hat{\Omega} - \mathbf{\Omega}) \frac{\partial \ln \mathcal{L}}{\partial p_i} \right\rangle \left\langle (\hat{\Omega} - \mathbf{\Omega}) \frac{\partial \ln \mathcal{L}}{\partial p_j} \right\rangle \\ &\leq \left\langle (\hat{\Omega} - \mathbf{\Omega})^2 \right\rangle \sqrt{\left\langle \left(\frac{\partial \ln \mathcal{L}}{\partial p_i} \right)^2 \right\rangle \left\langle \left(\frac{\partial \ln \mathcal{L}}{\partial p_j} \right)^2 \right\rangle}. \end{aligned} \tag{23}$$

2 Quadratic Estimation of the Power Spectrum

2.1 Overview

Quadratic estimators provide an optimal framework for measuring the power spectrum from noisy cosmological data, particularly in 21 cm cosmology where foreground contamination dominates. The key concepts are summarised as follows:

- **Data Vector Formalism**: The observed data vector $\mathbf{x}$ relates to the underlying signal through:
$$\mathbf{x} = \mathbf{s} + \mathbf{n} \tag{24}$$
where $\mathbf{s}$ is the cosmological signal and $\mathbf{n}$ represents noise/systematics.

- **Covariance Matrix Structure**: The data covariance matrix $\mathbf{C} \equiv \langle \mathbf{x}\mathbf{x}^\dagger \rangle$ decomposes as:
$$\mathbf{C} = \mathbf{N} + \sum_\alpha p_\alpha \mathbf{Q}_\alpha \tag{25}$$
where $\mathbf{N}$ is the noise covariance, $p_\alpha \equiv P(k_\alpha)$ are bandpowers, and $\mathbf{Q}_\alpha \equiv \frac{\partial \mathbf{C}}{\partial p_\alpha}$ is the response matrix.

- **Estimator Definition**: The quadratic estimator for bandpower p_α takes the form:
$$\hat{p}_\alpha \propto \mathbf{x}^\dagger \mathbf{E}^\alpha \mathbf{x} - b^\alpha \tag{26}$$
where $\mathbf{E}^\alpha$ encodes Fourier mode weighting and $b^\alpha \equiv \mathrm{Tr}(\mathbf{E}^\alpha \mathbf{N})$ removes noise bias.

- **Window Functions**: The estimator's spectral resolution is characterized by:
$$W_{\alpha\beta} = \mathrm{Tr}(\mathbf{E}^\alpha \mathbf{Q}_\beta) \tag{27}$$
quantifying leakage between k-bins.

2.2 A linear mapping perspective

- Data space and signal space:

The data vector ($\boldsymbol{x}$) and its components ($\boldsymbol{x}_s$ and $\boldsymbol{x}_n$) are vectors in the data space, $\mathcal{D}$:

$$\boldsymbol{x} = \boldsymbol{x}_s + \boldsymbol{x}_n \tag{28}$$
$$\boldsymbol{x}, \boldsymbol{x}_s, \boldsymbol{x}_n \in \mathcal{D} \tag{29}$$

where the data space signal component $\boldsymbol{x}_s$ can be understood as the linear map of a vector $\boldsymbol{s}$ in the signal space ($\mathcal{S}$):

$$\begin{aligned} \boldsymbol{x}_s &= U\boldsymbol{s} \qquad & \mathcal{S} &\equiv \operatorname{span}\{\hat{\boldsymbol{w}}_1, \cdots, \hat{\boldsymbol{w}}_{d_s}\} \\ \boldsymbol{s} &\in \mathcal{S} & U &: \mathcal{S} \mapsto \mathcal{D} \end{aligned}$$

where we interpret $\hat{\boldsymbol{w}}_\alpha$ as a unit-normalized, binned Fourier mode of the signal field. Correspondingly, the response matrix $\mathbf{Q}_\alpha$ is given (in linear algebra) by

$$\mathbf{Q}_\alpha \equiv U\hat{\boldsymbol{w}}_\alpha \hat{\boldsymbol{w}}_\alpha^\dagger U^\dagger. \tag{30}$$

- What is the band power p_α?

  Signal space viewpoint: the ensemble average of the squared projection of $\boldsymbol{s}$ on $\hat{\boldsymbol{w}}_\alpha$,

$$p_\alpha \equiv \langle |\boldsymbol{s}^\dagger \hat{\boldsymbol{w}}_\alpha|^2 \rangle = \langle \boldsymbol{s}^\dagger \hat{\boldsymbol{w}}_\alpha \hat{\boldsymbol{w}}_\alpha^\dagger \boldsymbol{s} \rangle. \tag{31}$$

  Data space viewpoint: Similar to the above, but now projecting onto the image of $\hat{\boldsymbol{w}}_\alpha$ in the data space, [1]

$$\frac{\langle |\boldsymbol{x}_s^\dagger U \hat{\boldsymbol{w}}_\alpha|^2 \rangle}{|U\hat{\boldsymbol{w}}_\alpha|^2} = \frac{\langle \boldsymbol{x}_s^\dagger U \hat{\boldsymbol{w}}_\alpha \hat{\boldsymbol{w}}_\alpha^\dagger U^\dagger \boldsymbol{x}_s \rangle}{\hat{\boldsymbol{w}}_\alpha^\dagger U^\dagger U \hat{\boldsymbol{w}}_\alpha} = \frac{\operatorname{Tr}(\mathbf{Q}_\alpha \mathbf{C}_s)}{\operatorname{Tr}(\mathbf{Q}_\alpha)} = \sum_\beta W_{\alpha\beta} p_\beta, \tag{32}$$

  where $\mathbf{C}_s \equiv \langle \boldsymbol{x}_s \boldsymbol{x}_s^\dagger \rangle$ is the covariance matrix of the signal component, and $W_{\alpha\beta} \equiv \operatorname{Tr}(\mathbf{Q}_\alpha \mathbf{Q}_\beta)/\operatorname{Tr}(\mathbf{Q}_\alpha)$ is the window function. Note that $W_{\alpha\beta}$ doesn't vanish in general. To avoid the mixing effect, one can first project data onto the space orthogonal to other modes, and then project onto the data space signal mode $U\hat{\boldsymbol{w}}_\alpha$.

2.3 Unbiased estimate with vector space structure

- Unbiased estimate of the windowed band power, $q_\alpha = \sum_\beta W_{\alpha\beta} p_\beta$:

$$\hat{q}_\alpha = \frac{\operatorname{Tr}(\mathbf{Q}_\alpha \hat{\mathbf{C}}_s)}{\operatorname{Tr}(\mathbf{Q}_\alpha)}. \tag{33}$$

  - Theoretically, we have

$$\mathbf{C}_s = \mathbf{C} - \mathbf{N}, \tag{34}$$

    but neither $\mathbf{C}$ nor $\mathbf{N}$ are known.

[1] Note that this normalization assigns a coefficient of 1 to the α-mode, rather than scaling the coefficients to sum to 1.

- A derived estimate of $\mathbf{C}_s$ is

$$\hat{\mathbf{C}}_s = \boldsymbol{x}\boldsymbol{x}^\dagger - \hat{\mathbf{N}}, \tag{35}$$

where $\boldsymbol{x}\boldsymbol{x}^\dagger$ is an instance/estimate[2] of $\mathbf{C}$ and $\hat{\mathbf{N}}$ is an estimate of $\hat{\mathbf{N}}$. So, if $\hat{\mathbf{N}}$ constitutes an unbiased estimator for the noise covariance matrix, the derived signal covariance estimate $\hat{\mathbf{C}}_s$ will consequently preserve this unbiasedness.

- Unbiased estimate with arbitrary weights:
  - Assume that we performed further data projections in the data space and we are using the projected data for the power spectrum estimation:

$$\boldsymbol{x}' = \mathbf{E}\boldsymbol{x}. \tag{36}$$

The windowed band power is then understood as

$$q'_\alpha = \frac{\mathrm{Tr}(\mathbf{E}^\dagger\mathbf{E}\mathbf{Q}_\alpha\mathbf{E}^\dagger\mathbf{E}\mathbf{C}_s)}{\mathrm{Tr}(\mathbf{E}\mathbf{Q}_\alpha\mathbf{E}^\dagger)}, \tag{37}$$

or written in the more familiar form,

$$q'_\alpha = \frac{\mathrm{Tr}(\tilde{\mathbf{E}}\mathbf{Q}_\alpha\tilde{\mathbf{E}}\mathbf{C}_s)}{\mathrm{Tr}(\mathbf{Q}_\alpha\tilde{\mathbf{E}})}, \qquad \tilde{\mathbf{E}} \equiv \mathbf{E}^\dagger\mathbf{E}. \tag{38}$$

Here we have used the "$'$" to denote the changed window:

$$q'_\alpha = \sum_\beta W'_{\alpha\beta}p_\beta, \qquad W_{\alpha\beta} \equiv \mathrm{Tr}(\tilde{\mathbf{E}}\mathbf{Q}_\alpha\tilde{\mathbf{E}}\mathbf{Q}_\beta)/\mathrm{Tr}(\mathbf{Q}_\alpha\tilde{\mathbf{E}}). \tag{39}$$

  - Consequently, the unbiased estimate is given by

$$\hat{p}_\alpha = \frac{\mathrm{Tr}\left[\tilde{\mathbf{E}}\mathbf{Q}_\alpha\tilde{\mathbf{E}}(\boldsymbol{x}\boldsymbol{x}^\dagger - \hat{\mathbf{N}})\right]}{\mathrm{Tr}(\mathbf{Q}_\alpha\tilde{\mathbf{E}})}, \tag{40}$$

which, again, conditional on the unbiasedness of $\hat{\mathbf{N}}$.

[2] Ideally, even though a data set captures a single instance of the sky field, it contains many redundant modes of the same statistic, so that the sample variance of the statistical parameters is suppressed.

2.4 Optimal estimate in different contexts

Thus far our analysis has utilized only the vector space properties inherent in the linear framework. However, the weighting matrix in unbiased estimation contains sufficient degrees of freedom to enable the construction of optimal estimators under different statistical criteria.

2.4.1 Minimum variance with Hilbert space structure

We now extend this foundation by endowing the linear spaces with metric structure - specifically through the formal introduction of Hilbert space inner products:

This enriched framework enables the rigorous definition of minimum-variance estimators, which are mathematically equivalent to solving generalized least squares optimization problems.

But wait, even "minimum variance" could mean different things:

- If we want to minimise the total variance, including the sampling variance of the signal modes as well as the noise variance:

$$\mathbf{E} = \mathbf{C}^{-\frac{1}{2}} \tag{41}$$

- If only to minimise the total noise variance, the weighting matrix will be

$$\mathbf{E} = \mathbf{N}^{-\frac{1}{2}} \tag{42}$$

- If necessary, we can further distinguish the foreground variance, instrumental variance, etc...

2.4.2 Optimal estimate in different contexts

The formalism generalizes to arbitrary linear projections, not limited to minimum-variance estimators. Two illustrative cases demonstrate this flexibility:

1. Projection onto the null space of systematics

   Given a eigendecomposed form of $\mathbf{N}$

$$\mathbf{N} = n_i \sum_{i=1}^{r} \hat{\boldsymbol{\mu}}_i \hat{\boldsymbol{\mu}}_i^{\dagger}, \tag{43}$$

then the "minimal systematic" projection is given by

$$\mathbf{E} = E_i \sum_{i=r+1}^{N_d} \hat{\boldsymbol{\mu}}_i \hat{\boldsymbol{\mu}}_i^\dagger, \tag{44}$$

where E_i are non-zero values that can be optimised to minimise variance.

2. Project onto the non-principle modes (in PCA analysis)

   PCA method in power spectrum estimation can also be understood in this framework. It effectively regards the linear space spanned by the non-principle modes of $\mathbf{C}$ as a signal-dominated space:[3]

$$\mathbf{E} = 1 - \sum_{i=1}^{n_{\rm PCA}} \hat{e}_i \hat{e}_i^\dagger = \sum_{i=n_{\rm PCA}}^{N_d} \hat{e}_i \hat{e}_i^\dagger \tag{46}$$

Note that different "optimal" operations (minimum variance, systematic avoidance, PCA filter, etc.) can be combined to serve multiple purposes.

2.5 Summary

- **Estimate**: a band power estimate can be simply understood as the squared projection onto the mapped signal basis vector.
- **Window function**: other modes could have nontrivial projections.
- **Weighting**: weights and additional projection can be multi-purpose: the PCA method of MeerKAT is just a special case.
- **Optimal**: when noise covariance is given, we can obtain minimum variance, that's understood as general least squares. . .

[3] This projector possesses several nice properties, Hermitian and the 'power idempotence':

$$\mathbf{E}^\dagger = \mathbf{E}, \mathbf{R}^n = \mathbf{R}, \ \textbf{for} \ n = 1, 2, \ldots. \tag{45}$$

Its Fourier-space counterpart, $\tilde{\mathbf{R}}$, inherits all of these properties (up to the normalisation of the Fourier transform).

A Cosmological Observables

The Baryon Acoustic Oscillations (BAO) in the radial and tangential directions derive different cosmological observables. The radial BAO measures the Hubble parameter, $H \equiv \dot{a}/a$, where a is the scale factor. The BAO in the tangential direction provide measurements of the angular diameter distance, D_A.

The Friedmann equation is

$$H^2(z) = \frac{8\pi G}{3}\left[\rho(z) + \frac{\rho_{cr} - \rho_0}{a^2(z)}\right] \tag{47}$$

where G is the Newton's constant, $\rho(z)$ is the energy density in the universe at redshift z with ρ_0 its present value, $\rho_{cr} = 3H_0^2/(8\pi G)$ is the critical density.

If the mean mass density is dominated by non-relativistic matter, the Friedmann equation gives

$$E(z) \equiv \frac{H(z)}{H_0} = \sqrt{\Omega_m(1+z)^3 + \Omega_R(1+z)^2 + \Omega_\Lambda} \tag{48}$$

where Ω_m, Ω_R and Ω_Λ are fractional contributions to the present value of Hubble's constant H_0 by the present mean mass density ρ_0, the radius of curvature $a_0 R$ and the cosmological constant Λ, which reads

$$\Omega_m = \frac{8\pi G\rho_0}{3H_0^2}, \quad \Omega_R = \frac{1}{(H_0 a_0 R)^2}, \quad \Omega_\Lambda = \frac{\Lambda}{3H_0^2}. \tag{49}$$

The way to constrain cosmological observables using BAO are essentially through distance measurements rather than direct measurement of the Hubble parameter.

A.1 Distances

Below we formulate the distances with explicit dependence on the frequency ν:

1. **Comoving distance (radial)**

   redshift z_ν:

$$1 + z_\nu = \frac{\nu_{21}}{\nu} \tag{50}$$

   radial comoving distance D_C (or r_ν):

$$\begin{aligned} r_\nu &= D_C(z_\nu) \\ &\equiv \frac{c}{H_0}\int_0^{z_\nu} \frac{\mathrm{d}z}{E(z)} \\ &= \frac{c}{H_0}\int_0^{z_\nu} \frac{\mathrm{d}z}{\sqrt{\Omega_m(1+z)^3 + \Omega_R(1+z)^2 + \Omega_\Lambda}} \end{aligned} \tag{51}$$

2. **Comoving distance (transverse)**

$$D_M = \begin{cases} D_H \frac{1}{\sqrt{\Omega_R}} \sinh\left[\sqrt{\Omega_R} D_C/D_H\right], & \text{for } \Omega_R > 0 \\ D_C, & \text{for } \Omega_R = 0 \\ D_H \frac{1}{\sqrt{|\Omega_R|}} \sin\left[\sqrt{|\Omega_R|} D_C/D_H\right], & \text{for } \Omega_R < 0 \end{cases} \tag{52}$$

where $D_H = c/H_0$ is the Hubble distance, and D_C is the comoving distance.

3. **Angular diameter distance**

$$D_A = \frac{D_M}{1+z} \tag{53}$$

where D_M is the transverse comoving distance.

A.2 Comoving space PS to configuration space variance

The two point correlation function of the signal field in the configuration space can be represented in terms of the cosmological power spectrum:

$$\begin{aligned} &\langle \mathbf{I}_{21}(\boldsymbol{\Theta}, r_\nu + \Delta_r) \mathbf{I}_{21}^*(\boldsymbol{\Theta}', r_\nu + \Delta_r') \rangle \\ &= \int \mathrm{d}^3k \; P_{21}(\mathbf{k}_\perp, k_\parallel) \exp\left[i\mathbf{k}_\perp \cdot \mathbf{r}_\perp + i k_\parallel r_\parallel\right] \\ &= \int \mathrm{d}^3k \; P_{21}(\mathbf{k}_\perp, k_\parallel) \exp\left[i\mathbf{k}_\perp \cdot D_A(\boldsymbol{\Theta}' - \boldsymbol{\Theta}) + i k_\parallel (\Delta_r' - \Delta_r)\right] \\ &= \int \mathrm{d}^3k \; P_{21}(\mathbf{k}_\perp, k_\parallel) \exp\left[i\mathbf{k}_\perp \cdot D_A(\boldsymbol{\Theta}' - \boldsymbol{\Theta}) + i k_\parallel \alpha (\Delta_\nu' - \Delta_\nu)\right] \end{aligned} \tag{54}$$

where, for a flat universe,

$$D_A = \frac{D_M}{1+z} = \frac{D_C}{1+z} = \frac{r_\nu}{1+z_\nu}, \tag{55}$$

and

$$\alpha = \frac{\mathrm{d}r_\nu}{\mathrm{d}\nu} = \frac{\mathrm{d}r_\nu}{\mathrm{d}z}\frac{\mathrm{d}z}{\mathrm{d}\nu} = \frac{-\nu_{21} c/H_0 \nu^2}{\sqrt{\Omega_\Lambda + \Omega_m (1+z_\nu)^3}}. \tag{56}$$