Notes on analytical marginalization

Hernán E. Noriega
henoriega@icf.unam.mx
Website: henoriega.github.io

The purpose of these notes is to demonstrate a potential approach to the marginalization of linear nuisance parameters in an Effective Field Theory of Large-Scale Structure. In particular, the idea is to show that with the current model and pipeline of the `FOLPS`ν code,[1] it is possible to analytically marginalize over both effective and stochastic parameters.

1 Perturbative model

We employ an Eulerian Perturbation Theory (EPT) up to 1-loop, which includes standard ingredients such as non-linear biasing, Effective Field Theory (EFT) counterterms, stochastic noise, and IR resummations. Therefore, the redshift space power spectrum can be expressed as [1]

$$\begin{aligned} P_s^{\text{EFT}}(k,\mu) = &P_{\delta\delta}(k) + 2f_0\mu^2 P_{\delta\theta}(k) + f_0^2\mu^4 P_{\theta\theta}(k) + A^{\text{TNS}}(k,\mu) + D(k,\mu) \\ &+ (\alpha_0 + \alpha_2\mu^2 + \alpha_4\mu^4)k^2 P_L(k) + \tilde{c}(f_0\sigma_v k\mu)^4 P_s^K(k,\mu) \\ &+ P_{\text{shot}}\big[\alpha_0^{shot} + \alpha_2^{shot}(k\mu)^2\big], \end{aligned} \tag{1.1}$$

where the functions $P_{\delta\delta}, P_{\delta\theta}$, and $P_{\theta\theta}$ are the tracers 1-loop real space power spectra for the velocity and density fields. The function $A^{\text{TNS}}(k,\mu)$ is equivalent to the one introduced in the TNS paper [2], but with a growth function $f(k,t)$ that depends on both scale and time. On large scales, this function reduces to $f_0(t) = f(k \to 0, t)$. The function $D(k,\mu)$ arises from the correlation between the density and velocity fields at fourth order. In fact, this function is a generalization of the $B(k,\mu)$ function introduced in the TNS paper. The second line is conformed by the EFT contribution and the linear Kaiser effect $P_s^K(k,\mu)$, while the stochastic terms are included in the last line.

All of these functions introduce a total of 11 nuisance parameters, where b_1, b_2, b_{s^2}, and b_{3nl} account for biasing. Meanwhile, α_0, α_2, and α_4 are EFT counterterms that model the backreaction of small scales over large scales and the non-linear map between real and redshift spaces. Additionally, in the line-of-sight direction, 2-point statistics are dominated by Fingers of God, which are a non-linear coupling between the velocity and density fields, with a characteristic scale given by the velocity dispersion σ_v. This yields the next-to-leading order counterterm $\tilde{c}$ [3].

The third line in eq. (1.1) includes stochastic parameters that are uncorrelated with long-wavelength fluctuations. The constant P_{shot} can be set equal to the Poisson process shot noise, which is given by $P_{\text{Poisson}} = 1/\bar{n}_X$, where $\bar{n}_X$ is the number density of tracers. Alternatively, P_{shot} can be set to any other constant value that is not relevant since it is completely degenerate with α_0^{shot}. We have also introduced a tilt α_2^{shot} proportional to $(k\mu)^2$.

Regarding the set of nuisance parameters introduced above, not all of them are necessary at the same time. As previously mentioned, the value of P_{shot} is not relevant and only alters

[1] https://github.com/henoriega/FOLPS-nu

the numerical values of α_0^{shot} and α_2^{shot}, so it can take any value. Here, we set it to the Poissonian shot noise value. On the other hand, the counterterm α_4 becomes redundant when only fitting the monopole and quadrupole. Therefore, this counterterm is employed only when including the hexadecapole in the analysis. Finally, the functional dependence introduced by $\tilde{c}$ is $k^4 P_L(k)$, which is approximately proportional to k^2 at high-k and hence degenerate with α_2^{shot}. Therefore, it is common to choose between the use of either $\tilde{c}$ or α_2^{shot}. In this work, we fixed $\tilde{c} = 0$ and used the latter as a nuisance parameter.

Large-scale displacements lead to non-linear damping of spatially localized features in the power spectrum, such as the BAO, which are not naturally captured by the EPT approach. Therefore, to adequately describe the spread and degradation of BAO oscillations within EPT, it is imperative to employ IR resummation methods [4–6]. Since large-scale displacements affect only the BAO wiggles, a convenient approach is to split the linear power spectrum as $P_L(k) = P_{\rm nw}(k) + P_{\rm w}(k)$, where $P_{\rm nw}(k)$ is the smooth component (non-wiggle) and $P_{\rm w}(k)$ is the wiggle component of the linear power spectrum that contains the BAO information.

As a result of the linear power spectrum splitting described above, the 1-loop IR-resummed power spectrum in redshift space $P_s^{\rm IR}(k,\mu)$ becomes [1],

$$P_s^{\rm IR}(k,\mu) = e^{-k^2\Sigma^2_{\rm tot}(k,\mu)} P_s^{\rm EFT}(k,\mu) + \left(1 - e^{-k^2\Sigma^2_{\rm tot}(k,\mu)}\right) P_{s,nw}^{\rm EFT}(k,\mu) + e^{-k^2\Sigma^2_{\rm tot}(k,\mu)} P_w(k) k^2 \Sigma^2_{\rm tot}(k,\mu), \tag{1.2}$$

where $P_s^{\rm EFT}(k,\mu)$ is given by eq. (1.1) and is to be calculated using the linear power spectrum as input, and $P_{s,\rm nw}^{\rm EFT}(k,\mu)$ is computed in the same manner but using as input the non-wiggle linear power spectrum $P_{\rm nw}(k)$. In addition, the angle-dependent damping function $\Sigma^2_{\rm tot}(k,\mu)$ is given by [6],

$$\Sigma^2_{\rm tot}(k,\mu) = \left[1 + f\mu^2(2+f)\right]\Sigma^2 + f^2\mu^2\left(\mu^2 - 1\right)\delta\Sigma^2, \tag{1.3}$$

with

$$\Sigma^2 = \frac{1}{6\pi^2}\int_0^{k_s} dp\, P_{nw}(p)\left[1 - j_0\left(p\,\ell_{\rm BAO}\right) + 2 j_2\left(p\,\ell_{\rm BAO}\right)\right], \tag{1.4}$$

$$\delta\Sigma^2 = \frac{1}{2\pi^2}\int_0^{k_s} dp\, P_{nw}(p) j_2\left(p\,\ell_{\rm BAO}\right), \tag{1.5}$$

where $\ell_{\rm BAO} \simeq 105\,h^{-1}{\rm Mpc}$ corresponds to the BAO peak scale, j_n represent the spherical Bessel functions of degree n. The scale k_s splits the long and short modes, whose choice is somewhat arbitrary. We use the value $k_s = 0.4\,h^{-1}{\rm Mpc}$.

We have already presented the full 1-loop power spectrum (1.2). However, to compare with data from simulations or observations, it is more convenient to decompose the information in multipoles

$$P_\ell(k) = \frac{2\ell+1}{2}\int_{-1}^{1} d\mu\, P_s^{\rm IR}(k,\mu)\mathcal{L}_\ell(\mu), \tag{1.6}$$

where $\mathcal{L}_\ell$ are the Legendre polynomial of degree ℓ.

2 Marginalization over linear parameters

Given a dataset $\boldsymbol{d}$ and a theoretical model $\mathcal{M}$ parameterized by a set of parameters $\boldsymbol{\theta}$, the likelihood function $\mathcal{L}(\boldsymbol{d}|\boldsymbol{\theta},\mathcal{M})$ describes the conditional probability distribution of the data.

Specifically, for the model in relevance, the likelihood can be expressed as follows

$$\mathcal{L}(\boldsymbol{d}|\boldsymbol{\theta}) = \mathrm{Exp}\left[-\frac{1}{2}\left(P_\ell^{(m)}(k,\boldsymbol{\theta}) - P_\ell^{(d)}(k)\right)^{\mathrm{T}} \mathrm{Cov}^{-1}\left(P_\ell^{(m)}(k,\boldsymbol{\theta}) - P_\ell^{(d)}(k)\right)\right], \tag{2.1}$$

where $P_\ell^{(d)}(k)$ are the multipoles extracted from the data, $P_\ell^{(m)}(k,\boldsymbol{\theta})$ are the multipoles given by the model, and Cov^{-1} is the inverse covariance matrix. The parameters $\boldsymbol{\theta} = \{\boldsymbol{\Omega}, \boldsymbol{n}\}$ are the combination of the cosmological parameters $\boldsymbol{\Omega}$ and the nuisances parameters $\boldsymbol{n} = \{b_1, b_2, b_{s^2}, b_{3nl}, \alpha_0, \alpha_2, \alpha_4, \alpha_0^{shot}, \alpha_2^{shot}\}$.[2]

On the other hand, we can divide the nuisance parameters into two types: bias parameters $\boldsymbol{b} = \{b_1, b_2, b_{s^2}, b_{3nl}\}$, and effective and stochastic parameters $\boldsymbol{\alpha} = \{\alpha_0, \alpha_2, \alpha_4, \alpha_0^{shot}, \alpha_2^{shot}\}$. The latter have the characteristic that they are linear-order parameters at the level of the power spectrum multipoles. This characteristic allows us to rewrite the multipoles as

$$P_\ell^{(m)}(k,\boldsymbol{\theta}) = \sum_i \alpha_i\, P_{\ell,i}^{(m)}(k,\{\boldsymbol{\Omega},\boldsymbol{b}\}) + P_{\ell,\mathrm{const}}^{(m)}(k,\{\boldsymbol{\Omega},\boldsymbol{b}\}), \tag{2.2}$$

where $P_{\ell,i}^{(m)}(k,\{\boldsymbol{\Omega},\boldsymbol{b}\}) \equiv \frac{\partial P_\ell^{(m)}(k,\boldsymbol{\theta})}{\partial \alpha_i}$ and $P_{\ell,\mathrm{const}}^{(m)}(k,\{\boldsymbol{\Omega},\boldsymbol{b}\}) \equiv P_\ell^{(m)}(k,\boldsymbol{\theta})\Big|_{\boldsymbol{\alpha}\to 0}$.

Marginalizing over the linear nuisances parameters $\boldsymbol{\alpha}$ and using eqs. (2.1) and (2.2), we found

$$\begin{aligned}
\mathcal{L}(\boldsymbol{d}|\{\boldsymbol{\Omega},\boldsymbol{b}\}) &= \int d\alpha\, \mathcal{L}(\boldsymbol{d}|\boldsymbol{\theta} = \{\boldsymbol{\Omega},\boldsymbol{b},\boldsymbol{\alpha}\}) \\
&= \int d\alpha\, e^{-\frac{1}{2}\left(\sum_i \alpha_i P_{\ell,i}^{(m)} + P_{\ell,\mathrm{const}}^{(m)} - P_\ell^{(d)}\right)^{\mathrm{T}} \mathrm{Cov}^{-1}\left(\sum_j \alpha_j P_{\ell,j}^{(m)} + P_{\ell,\mathrm{const}}^{(m)} - P_\ell^{(d)}\right)} \\
&= \int d\alpha\, e^{-\frac{1}{2}\sum_{i,j}\alpha_i\alpha_j \mathcal{L}_{2,ij} + \sum_i \alpha_i \mathcal{L}_{1,i} + \mathcal{L}_0},
\end{aligned} \tag{2.3}$$

with

$$\mathcal{L}_0 = -\frac{1}{2}\mathcal{D}_{\mathrm{const}} \star \mathcal{D}_{\mathrm{const}}, \tag{2.4}$$

$$\mathcal{L}_{1,i} = -P_{\ell,i}^{(m)} \star \mathcal{D}_{\mathrm{const}}, \tag{2.5}$$

$$\mathcal{L}_{2,ij} = P_{\ell,i}^{(m)} \star P_{\ell,j}^{(m)}, \tag{2.6}$$

where $\mathcal{D}_{\mathrm{const}} \equiv P_{\ell,\mathrm{const}}^{(m)} - P_\ell^{(d)}$ is the residual between the model multipoles (for the constant part) and the data vector, also note we used the shorthand notation $A \star B \equiv A^T\, \mathrm{Cov}^{-1}\, B$.

Using the multivariate Gaussian integral

$$\int d^n x\, e^{-\frac{1}{2}\sum_{i,j} x_i x_j A_{ij} + \sum_i x_i B_i} = \sqrt{\frac{(2\pi)^n}{\det A}}\, e^{\frac{1}{2}B_i A_{ij}^{-1} B_j}, \tag{2.7}$$

we found that the marginalized Likelihood takes the form

$$\ln \mathcal{L}(d|\{\boldsymbol{\Omega},\boldsymbol{b}\}) = \mathcal{L}_0 + \frac{1}{2}\mathcal{L}_{1,i} \cdot \mathcal{L}_{2,ij}^{-1} \cdot \mathcal{L}_{1,j} - \frac{1}{2}\ln\left[\det(\mathcal{L}_2)\right], \tag{2.8}$$

where irrelevant constants were omitted.

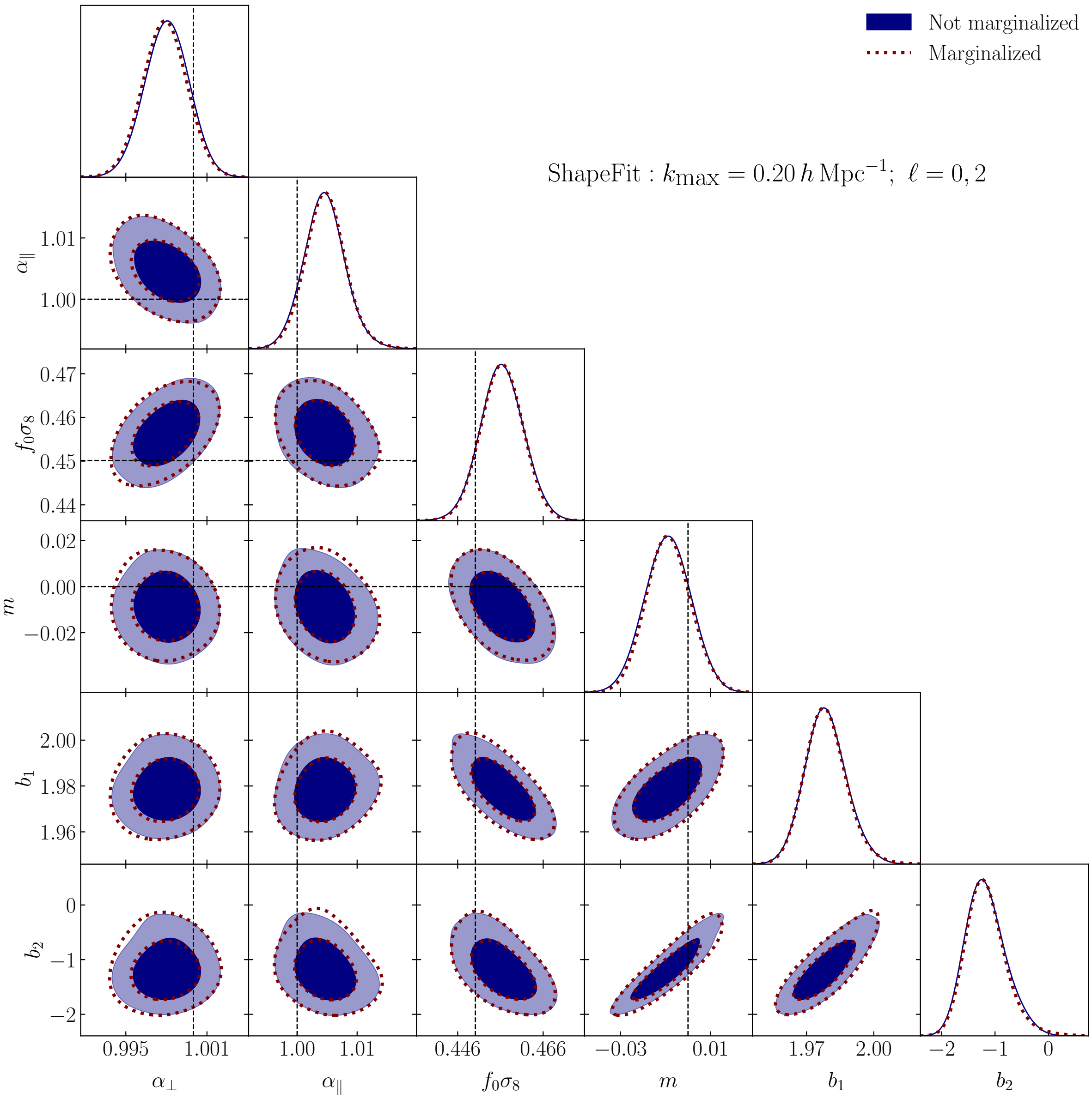

Figure 1. Comparison for ShapeFit.

3 Results

References

[1] H. E. Noriega, A. Aviles, S. Fromenteau and M. Vargas-Magaña, *Fast computation of non-linear power spectrum in cosmologies with massive neutrinos*, *JCAP* **11** (2022) 038, [2208.02791].

[2] A. Taruya, T. Nishimichi and S. Saito, *Baryon Acoustic Oscillations in 2D: Modeling*

[2]In this work we set $\tilde{c} = 0$ and $P_{\rm shot}$ to its Poisson shot noise value.

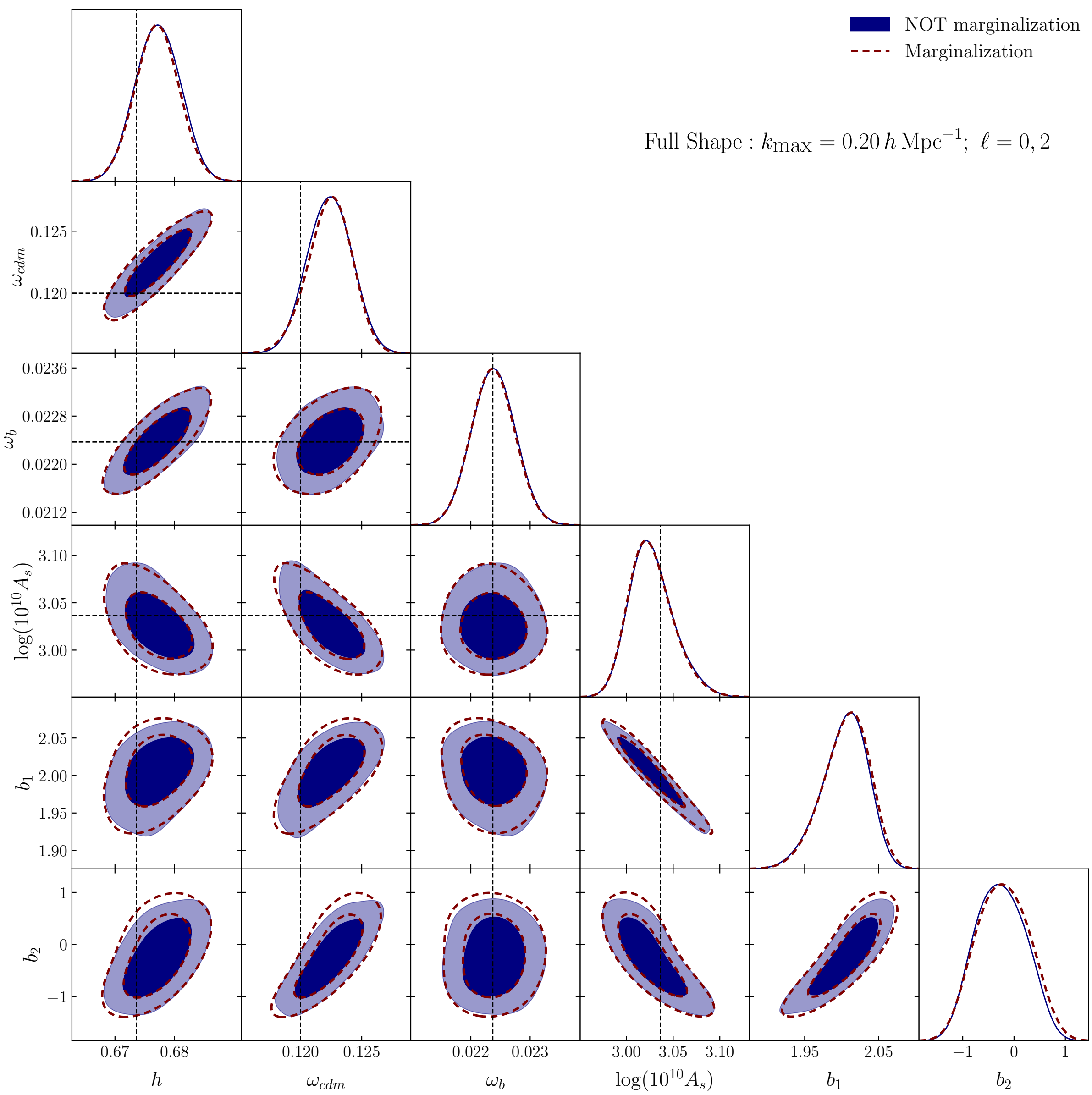

Figure 2. Comparison for Full Shape.

Redshift-space Power Spectrum from Perturbation Theory, *Phys. Rev. D* **82** (2010) 063522, [1006.0699].

[3] M. M. Ivanov, M. Simonović and M. Zaldarriaga, *Cosmological Parameters from the BOSS Galaxy Power Spectrum*, *JCAP* **05** (2020) 042, [1909.05277].

[4] T. Baldauf, M. Mirbabayi, M. Simonović and M. Zaldarriaga, *Equivalence Principle and the Baryon Acoustic Peak*, *Phys. Rev. D* **92** (2015) 043514, [1504.04366].

[5] L. Senatore and M. Zaldarriaga, *The IR-resummed Effective Field Theory of Large Scale Structures*, *JCAP* **02** (2015) 013, [1404.5954].

[6] D. Blas, M. Garny, M. M. Ivanov and S. Sibiryakov, *Time-Sliced Perturbation Theory II: Baryon Acoustic Oscillations and Infrared Resummation*, *JCAP* **07** (2016) 028, [1605.02149].