Gradient Accelerated Cosmology

Jaime Ruiz-Zapatero
St Cross College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy

Hilary 2024

A mis abuelas, María y Delfina.

DECLARATION

I declare that no part of this thesis has been accepted, or is currently being submitted, for any degree or diploma or certificate or any other qualification in this University or elsewhere.

The chapters of this thesis are based on already published works. Chapter 2 is based on Ruiz-Zapatero *et al.* [1]. In this work I led the development of the method as well as performing the presented numerical analyses. Chapter 3 is based on the work Ruiz-Zapatero *et al.* [2] where I led the extension of the methods developed in Ruiz-Zapatero *et al.* [1] to modified gravity. Chapter 4 is based on Ruiz-Zapatero *et al.* [3]. In this work, I led the development of an open-source library for auto-differentiable methods for angular power spectra analysis. Finally, chapter 5 is based on Hadzhiyska *et al.* [4] and Ruiz-Zapatero *et al.* [5]. In the former work, I played a supporting role to my collaborators who investigated the application of the Laplace approximation to 3x2-pt analyses. In the later work, I led a study on the validity of the methodology when applied specifically to photometric redshift uncertainties in cosmic shear analyses.

The work contained in this thesis was supported by the UK's STFC as well as St Cross College, Oxford.

Jaime Ruiz Zapatero

April 1, 2024

Acknowledgements

No podría empezar esta tesis sin expresar mi profundo agradecimiento a mis padres, Luz y Gabriel, por todo su amor, paciencia y dedicación, sin los que nada de esto hubiera sido posible. También me gustaría agradecer a mi novia, Valeria, sujetar mi mano y apoyarme durante estos años.

I would like to thank my supervisors, David Alonso and Pedro Ferreira, whom I have happily come to regard as esteemed friends. The work in this thesis is a direct result of their insightful guidance and unflinching optimising. Equally, I would like to thank Uroš Seljak and Hong Ge for hosting me during my greatly enriching stays at Berkeley and Cambridge. I would also like to thank Carlos García-García who has in no small part also directed this thesis as well as greatly contributed to it. Similarly, I would like to thank all the people that have contributed to this thesis: Susanna Azzoni, Marco Bonici, Richard Grumitt, Boryana Hadzhiyska, Andrina Nicola, Arrykrishna Mootoovaloo, Kevin Wolz, Anže Slosar and Jamie Sullivan. Moreover, I would like to express my endless gratitude to Ashling Gordon without whose unparalleled diligence and disposition I would still be trying to enrol on this course. Equally vital were the heroic efforts of Jonathan Patterson that kept Glamdring, the Oxford computer cluster on which this thesis was ran, working like a Swiss watch.

Finally, I would also like to thank all my friends and colleagues for the innumerable coffee breaks, lunch excursions and Spoons's visits that fuelled me through this PhD. I would like to thank all the members of la BIPAC for providing me with a cult-like feeling of being welcomed. A special place in my heart is reserved to my office mates, Alex Andersson, Casey Cragg, Eric Liang, Curro Rodriguez and Madalina Tudorache without whose contribution this thesis would be twice as long but infinitely less pleasurable. I would also like to thank my flat mates for bearing with me, specially during the COVID lockdowns. Moreover, I would like to thank Roy and Mateo for hosting me at Green Fairies Farm during the rainiest winter in California.

Contents

1 Introduction 3

1.1 Cosmology 3

1.1.1 Background 3

1.1.2 Inhomogeneities 9

1.1.3 Power Spectra 12

1.1.4 Angular Power Spectra 17

1.1.5 Tomography and Systematics 23

1.2 Statistical Inference 29

1.2.1 Traditional Inference Algorithms 29

1.2.2 Gradient-Based Inference Algorithms 36

1.2.3 The Laplace Approximation 42

1.3 Computer Differentiation 50

1.3.1 Finite Differences 50

1.3.2 Auto-Differentiation 51

1.4 Gaussian Processes 55

1.4.1 Motivation 55

1.4.2 Formalism 56

1.4.3 Interpretation 60

2 Model-independent constraints on Ω_m and $H(z)$ from the link between geometry and growth 62

2.1 Introduction 62

2.2 Methods 64

2.2.1 A Gaussian Process for $H(z)$ 64

2.2.2 Likelihood 67

2.3 Observables and Data sets 69

2.3.1 Cosmic Chronometers 71

2.3.2 Supernovae 71

2.3.3 Baryon Acoustic Oscillations 72
2.3.4 Redshift Space Distortions 73
2.3.5 Synthetic Stage-IV Data 74
2.4 Results 74
2.4.1 Current constraints 74
2.4.2 Forecasts 80
2.5 Conclusions 82

3 The impact of the Universe's expansion rate on constraints on modified growth of structure 85
3.1 Introduction 85
3.2 Methods 87
3.2.1 A Gaussian Process for $\mu(z)$ 87
3.2.2 Likelihood 88
3.3 Observables and Data sets 92
3.4 Results 94
3.4.1 Model-dependent constraints 94
3.4.2 Model-independent constraints 99
3.5 Conclusions 103

4 LimberJack: Auto-differentiable methods for angular power spectra analyses 108
4.1 Introduction 108
4.2 Methods 110
4.2.1 LimberJack.jl 110
4.3 Results 118
4.3.1 DES Y1 3x2-pt 118
4.3.2 Growth Factor Reconstruction 124
4.4 Conclusions 140

5 Analytical Marginalisation over Nuisance Parameters in Tomographic Analyses of Large Scale Structure 145
5.1 Introduction 145
5.2 Methods 146
5.2.1 Calibratable systematics 147
5.2.2 Non-calibratable systematics 149
5.3 Data 151

5.3.1 DES Y1 151
5.3.2 Synthetic Stage-IV data 153
5.4 Results 153
5.4.1 DES Y1 3x2-pt 153
5.4.2 DES-Y1 Shear Photometric Redshift Uncertainties 160
5.5 Conclusions 168

6 Conclusions **172**

Bibliography **176**

Acronyms

AD Auto-Differentiation.

BOSS Baryonic Oscillation Specstroscopic Survey.

CCL Core Cosmology Library.

CDM Cold Dark Matter.

CMB Cosmic Microwave Background.

DES Dark Energy Survey.

DESI Dark Energy Spectroscopic Instrument.

eBOSS Extended Baryonic Oscillation Specstroscopic Survey.

FLRW Friedmann Lemaître Robertson Walker.

GC Galaxy Clustering.

GP Gaussian Process.

GR General Relativity.

HMC Hamiltonian Monte Carlo.

KiDS Kilo Degree Survey.

LSST Legacy Survey of Space and Time.

MCMC Monte Carlo Markov Chain.

MH Metropolis Hastings.

OO Operator Overload.

SO Simons Observatory.

ST Source Transformation.

WL Weak Lensing.

Abstract

Cosmology is currently sitting on a ticking time bomb that will result in an unprecedented explosion in the quantity and quality of data. In preparation, physicists are starting to incorporate into their theoretical predictions more of the physical, observational and instrumental effects which, until now, could be overlooked. In practice, this translates into a dramatic increase in the number of parameters that future analyses will have to consider. This combination of large data sets with complex models will (and in many cases already does) overwhelm the inference methods we currently use to constrain the values of these parameters. In this thesis, we propose two solutions to this problem. First, we show how gradient-based inference algorithms can dramatically speed up the numerical marginalisation of high dimensional parameter spaces. Second, we show how analytical marginalisation schemes, such as the Laplace approximation, can achieve similar speed increases. Crucially, both these methods rely on having access to computationally affordable gradients of the cosmological models, stressing the importance of developing differentiable analyses pipelines for future cosmological surveys.

Chapter 1

Introduction

1.1 Cosmology

1.1.1 Background

Cosmology is the study of the Universe as a whole. On such large scales (megaparsecs [Mpc] to gigaparsecs [Gpc]), the primary form of interaction is the force of gravity. While the exact place of gravity alongside the other fundamental forces in the standard model is uncertain, cosmologists have come up with a very successful operational understanding of gravity, General Relativity (GR). GR proposes that the observed richness of gravitational interactions boils down to bodies finding their shortest path through a warped space-time whose curvature is sourced by the mass of the very same bodies [6–8] (See Fig. 1.1).

In the absence of matter, the intertwining of space and time is given by the Minkowski metric:

$$\eta_{\mu\nu} = \begin{pmatrix} -c^2 & 0 \\ 0 & \delta_{ij} \end{pmatrix} , \tag{1.1}$$

where c is the speed of light and δ_{ij} is the Kronecker delta symbol which provides the metric in traditional Euclidean spaces. Thus, distances on this manifold are given by:

$$ds^2 = dx^\nu \eta_{\nu\mu} dx^\mu = -c^2 dt^2 + dx^2 + dy^2 + dz^2 \, . \tag{1.2}$$

From now on we set the speed of light to one ($c = 1$) for convenience.

When matter is present, the metric needs to be found by solving the Einstein Field Equations (EFE) which establish the relationship between matter and curvature:

$$\boxed{G_{\mu\nu} + \Lambda g_{\mu\nu} = \kappa T_{\mu\nu}} \, . \tag{1.3}$$

The left-hand side of the EFE is known as the geometry term, while the right-hand side is known as matter term. The geometry term captures the geometrical properties

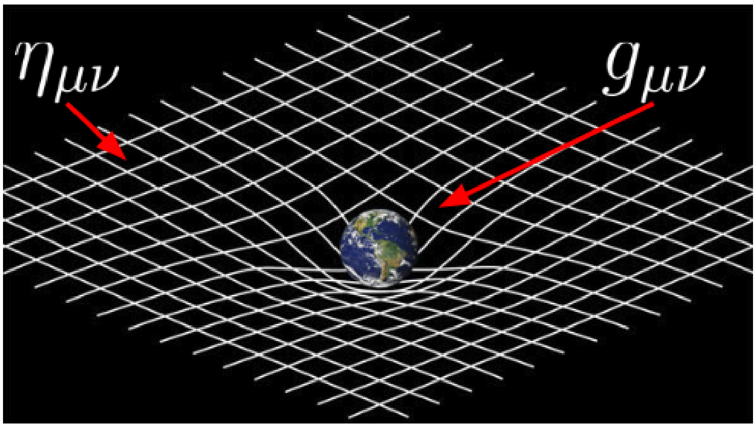

Figure 1.1: Visualisation of GR. In the absence of matter the space-time metric is given by Eq. 1.1. When matter is introduced the space-time curves following the EFE. Original picture credit: NASA.

of the space-time manifold given by the Einstein tensor $G_{\mu\nu}$, the cosmological constant Λ and the space-time metric $g_{\mu\nu}$. The matter term features the stress-energy tensor $T_{\mu\nu}$ which describes the energy-matter content of the Universe and κ, a proportionality constant tuned to $\kappa \equiv 8\pi G$ where G is Newton's constant such that Newtonian gravity is returned in the small mass limit.

The EFE describe a complicated system of non-linear, coupled, tensorial equations which can only be solved numerically even for fairly simple distributions of matter. Thus, cosmologists rely on assumptions to simplify Eq. 1.3 and solve for the space-time metric. In Cosmology, the most important assumption is the Copernican principle, which states we are not a *privileged* observer of the Universe and that what we see must be the same as what other observers see, independent of their location. For this to be the case, it follows that the Universe must be homogeneous and isotropic on large enough scales. In other words, the Universe must be invariant under translations and rotations. These homogeneity and isotropy requirements imply that, for a given time, the associated space-like hypersurface must possess constant curvature. This set of assumptions is so vital to Cosmology that it is also often referred to as the *cosmological principle*.

By combining the different implications of the Copernican principle we can construct an ansatz for what the form of the solution to Eq. 1.3 should be, known as the Friedmann, Lemaître, Robertson, and Walker (FLRW) metric [9, 10]:

$$g_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & \frac{a^2(t)}{1-kr^2} & 0 & 0 \\ 0 & 0 & a^2(t)r^2 & 0 \\ 0 & 0 & 0 & a^2(t)r^2\sin^2\theta \end{pmatrix}, \tag{1.4}$$

where we have chosen spherical coordinates, $\{r, \theta, \phi\}$, to describe the spatial part of the metric, $a(t)$ is an unknown and k is the curvature constant of the space hypersurface. The role of $a(t)$ is best understood by considering the line element of the FLRW metric:

$$ds^2 = -dt^2 + a^2(t)\left[\frac{dr^2}{1-kr^2} + r^2 d\Omega^2\right] , \tag{1.5}$$

where $d\Omega^2 = d\theta^2 + \sin^2\theta d\phi^2$. Thus $a(t)$ rescales the spatial sector of the metric at any given time, hence, it is known as the scale factor. Note that it is common to reparameteriseEq. 1.4 in terms of the *comoving radial distance*, χ, such that the line element is given by:

$$ds^2 = -dt^2 + a^2(t)\left[d\chi^2 + \mathcal{E}(\chi, k)d\Omega^2\right] , \tag{1.6}$$

where $\mathcal{E}(\chi, k)$ is a function of curvature. In the absence of curvature $r \equiv \chi$ and $\mathcal{E}(\chi, k) \equiv \chi^2$ such that the two parametrisations become equivalent. Thus, when discussing flat models we will use the two coordinates indistinguishably.

The cosmological principle also implies that the distribution of matter in the Universe must be equally invariant to translations and rotations. Assuming that matter in the Universe is described by a perfect fluid, the stress-energy tensor can then be written as

$$\boxed{T_{\mu\nu} = (\rho + p)U_\mu U_\nu + pg_{\mu\nu}} , \tag{1.7}$$

where $p(t) \equiv p$ and $\rho(t) \equiv \rho$ are the mean pressure and energy density of Universe, U_μ is the four-velocity of the fluid and $g_{\mu\nu}$ is the aforementioned metric.

Using Eq. 1.4 to compute the Einstein tensor and substituting Eq. 1.7 into the right-hand side of Eq. 1.3 results in the Friedmann Field Equations (FFE), given by:

$$\boxed{\frac{\dot{a}^2}{a^2} = \frac{8\pi G}{3}\rho + \frac{\Lambda}{3} - \frac{k}{a^2}} \tag{1.8}$$

$$\boxed{\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3p) + \frac{\Lambda}{3}} \tag{1.9}$$

where Eq. 1.8 is also known as the Hubble rate, $H(a) \equiv \dot{a}/a$. Thus, the FLRW metric describes a Universe with a dynamical size that evolves according to the FFE, where $H > 0$ describes an expanding Universe, $H < 0$ describes a contracting Universe and finally $H = 0$ corresponds to a static Universe.

Unfortunately, Eqs. 1.8 and 1.9 define a system of two differential equations with three unknown variables: $a(t)$, $p(t)$ and $\rho(t)$, meaning that their solution is underdetermined. Thus, we must provide an additional equation. A good initial guess is to enforce conservation of energy and momentum; i.e. $\nabla_\alpha T^\alpha_\beta = 0$, to derive a third continuity equation. Doing this leads to the following conservation equation:

$$\boxed{\dot{\rho} + 3\frac{\dot{a}}{a}(p+\rho) = 0}\,. \tag{1.10}$$

However, conservation of energy and momentum is already implied in the FFE by the construction of the Einstein tensor through the twice contraction of the Bianchi identities. Thus, Eq. 1.10 is not independent of the FFE. Instead, we must specify the matter content in the Universe. In order to do so, we must introduce an equation of state relating our matter density and pressure:

$$p = w\rho\,, \tag{1.11}$$

where w is that equation of state parameter and can be understood as the speed of sound through the specific fluid squared[1]. Plugging Eq. 1.11 into Eq. 1.10 allows us to find solutions for the evolution of the matter content of the Universe depending on its nature, i.e. its value of w.

The current cosmological paradigm, the Λ - Cold Dark Matter (ΛCDM) model [11], proposes that the Universe is composed of four different fluids: baryonic matter, dark matter, radiation and dark energy. Baryonic matter corresponds to traditional matter as described by the standard model of particle physics. Dark matter is an enigmatic form of matter which is *dark*, meaning that it does not interact electromagnetically, only gravitationally. Radiation encompasses the photons and other relativistic species (including massless neutrinos) travelling through the Universe. Finally, dark energy is another mysterious electromagnetically dark fluid which is believed to be the contribution of the cosmological constant to the Hubble rate. Both baryonic and dark matter are described by $w = 0$, such that the evolution of the total matter density is given by:

$$\rho_m(t) = \rho_{m\,0}\,a(t)^{-3}\,, \tag{1.12}$$

where $\rho_{m\,0}$ is matter density today. Radiation is described by $w = 1/3$. Hence, radiation density evolves according to:

$$\rho_r(t) = \rho_{r\,0}\,a(t)^{-4}\,, \tag{1.13}$$

where $\rho_{r\,0}$ is radiation density today. Finally, one can straightforwardly see that the Λ contribution to the FFE is equivalent to that of a exotic form of matter with state parameter $w = -1$.

---

[1] Note that interpreting w as the square of the speed of sound through the fluid is not so straight forward when w takes negative values

Substituting Eqs. 1.12 and 1.13 into Eq. 1.8 results into the ΛCDM expression for the Hubble rate:

$$H^2(a) = \frac{8\pi G\rho_{m\,0}}{3a^3} + \frac{8\pi G\rho_{r\,0}}{3a^4} + \frac{\Lambda}{3} - \frac{k}{a^2}\,, \tag{1.14}$$

which is now a fully determined ordinary differential equation (ODE) from which we can solve the scale factor to specify the FLRW metric in terms of the contents of the Universe. However, it is possible to further simplify the expression in Eq. 1.14. In order to do so, let us start by dividing both sides of Eq. 1.14 by $H^2(a)$ to obtain:

$$1 = \frac{8\pi G\rho_{\mathrm{m}\,0}}{3H^2}a^{-3} + \frac{8\pi G\rho_{r\,0}}{3H^2}a^{-4} + \frac{\Lambda}{3H^2} - \frac{k}{H^2}a^{-2}\,, \tag{1.15}$$

Now setting $t = t_0$ and rearranging we get:

$$1 + \Omega_{\mathrm{k}} = \Omega_{\mathrm{cdm}} + \Omega_{\mathrm{b}} + \Omega_{\mathrm{r}} + \Omega_{\Lambda}\,, \tag{1.16}$$

where $\Omega_{\mathrm{m}} \equiv 8\pi G\rho_{m0}/3H_0^2$, $\Omega_{\mathrm{r}} \equiv 8\pi G\rho_{r0}/3H_0^2$, $\Omega_\Lambda \equiv \Lambda/3H^2$ and $\Omega_k \equiv \frac{k}{H_0^2}$ are matter, radiation, dark energy and curvature cosmological densities. The cosmological curvature density has been repeatedly been measured to be consistent with $\Omega_k = 0$ [12, 13]. Using this knowledge, we can use Eq. 1.16 to write Ω_Λ in terms of the other cosmological densities leading to the following expression for the Hubble rate in the flat ΛCDM model:

$$\boxed{H(a) = H_0\sqrt{\Omega_{\mathrm{m}}a^{-3} + \Omega_{\mathrm{r}}a^{-4} + (1 - \Omega_{\mathrm{m}} - \Omega_{\mathrm{r}})}\,}\,. \tag{1.17}$$

Armed with Eq. 1.17 we can answer one of the simplest, yet non-trivial, questions in Cosmology: *How far are objects in the cosmos?* Assuming that we observe them through light, the photons that reached us did so by following the radial null geodesics of the manifold described by the FLRW metric (Eq. 1.4). This means that the 4-modulus of the distance must be zero and that the only non-zero terms are the time and the radial coordinate, leading us to the following expression:

$$ds^2 = dx^\nu g_{\nu\mu} dx^\mu = -dt^2 + a^2(t)d\chi^2 = 0\,. \tag{1.18}$$

Thus our radial separation to the source is given by:

$$\chi(t) = \int_t^0 \frac{dt}{a(t)}\,, \tag{1.19}$$

where t is the time at which the photons were emitted. However, observers rarely have access to the time of the emission of the photons. Instead, we normally measure the redshift of the source, z, i.e. how much the wavelength of the original light emission has dilated due to the expansion of the Universe along its trajectory

$$z(t) \equiv \frac{\lambda_0 - \lambda(t)}{\lambda(t)} = \frac{\lambda_0(1 - a(t))}{\lambda_0 a(t)} = \frac{1}{a(t)} - 1\,, \tag{1.20}$$

where λ_0 is the observed wavelength. We can then use the following substitution to express Eq. 1.19 as a function of redshift:

$$dt = dz\frac{dt}{da}\frac{da}{dz} = \frac{-dz}{H(z)(1+z)}\,. \tag{1.21}$$

Thus, in a flat ΛCDM model, the comoving distance (i.e. accounting for the expansion of the Universe) is given by:

$$\boxed{\chi(z) = \frac{1}{H_0}\int_0^z \frac{dz}{\sqrt{\Omega_\mathrm{m}(1+z)^3 + \Omega_\mathrm{r}(1+z)^4 + (1 - \Omega_\mathrm{m} - \Omega_\mathrm{r})}}}\,. \tag{1.22}$$

Astronomers rarely observe the comoving distance itself. Instead, we often observe proxies of it. One of the most common ways in which distances are determined in astronomy is by measuring the angle subtended by an object of known physical size d. Assuming the angle subtended, θ, is small, the angular diameter distance to the object is given by:

$$D_A = \frac{d}{\theta}\,. \tag{1.23}$$

However, in an expanding Universe, the size of the object scales as d/a, where a is the scale factor. Inserting this into Eq. 1.23 and solving for θ we get:

$$\theta = \frac{d/a}{\chi}\,, \tag{1.24}$$

where χ is the previously derived comoving distance. Thus we can see that in an expanding Universe, the angular diameter distance to an object is given by:

$$\boxed{D_A \equiv a\chi = \frac{\chi}{1+z}}\,. \tag{1.25}$$

Another common proxy is the so called luminosity distance. The luminosity distance is related to the flux of photons (F) through a comoving spherical shell irradiated by an object of known luminosity (L):

$$F = \frac{L}{4\pi\chi^2}\,. \tag{1.26}$$

In an expanding Universe, Eq. 1.26 needs to be modified to account for two effects. First, the energy of the photons crossing the spherical shell will by reduced as a consequence of redshift. Second, the interval between the arrival of different photons will also be dilated by the expansion of the Universe. Adding these two effects into Eq. 1.26 we obtain:

$$F = \frac{La^2}{4\pi\chi^2} \,. \tag{1.27}$$

Thus, in an expanding Universe the luminosity distance (D_L) is given by:

$$\boxed{D_L \equiv \frac{\chi}{a} = (1+z)\chi} \,. \tag{1.28}$$

1.1.2 Inhomogeneities

So far we have concerned ourselves with the background dynamics of space-time due to the presence of different species of matter as described by the EFE/FFE. In order to do so, we have assumed the distributions of these different components to be perfectly homogeneous and isotropic at every point in the history of the Universe. However, galaxy surveys have uncovered that our Universe possess a large scale structure made of objects that span entire fractions of the observable Universe such as the Sloan Great Wall [14] or the Coma Cluster [15]. This large scale structure, often referred to as the cosmic web [16], is believed to be the result of the gravitational collapse of primordial matter inhomogeneities [17]. However, the origin of these primordial inhomogeneities is still subject of speculation [18].

At first, it might seem like the presence of this large scale structure might violate the cosmological principle. However, the cosmological principle only requires the Universe to be *statistically* homogeneous and isotropic. Meaning that, on large enough scales, the Universe possess the same statistical properties for all observers. This means that on such scales the inhomogeneities can be considered to be *small.* Thus, we can study the formation of this structure as a linear perturbation on top of a FLRW background. In such a scenario, and on small enough scales, the gravitational fields involved are weak enough that we can, for the most part, do away with GR and follow the Newtonian dynamics of the fluid [7]. The dynamics of a non-relativistic fluid are then determined by three equations:

$$\frac{D\boldsymbol{v}}{Dt} = -\frac{\nabla p}{\rho} - \nabla\Phi \,, \tag{1.29}$$

$$\frac{D\rho}{Dt} = -\rho\nabla\cdot\boldsymbol{v} \,, \tag{1.30}$$

$$\nabla^2\Phi = 4\pi G\rho \,, \tag{1.31}$$

where $D/Dt = \partial_t + \boldsymbol{v} \cdot \nabla$ is the Lagrangian derivative of the fluid. The first equation is the Euler equation and it encapsulates conservation of momentum of the fluid. The second equation describes the conservation of energy in the fluid. Finally, the last equation is the Poisson equation for gravitational collapse. In a fully relativistic set up, these equations would correspond to the space-like and time-like components of the divergence stress energy tensor of a perfect fluid and the time-like FFE.

Our goal now is to linearise these equations by assuming small perturbations in the density, velocity and gravitational fields:

$$\rho(t, \boldsymbol{x}) = \bar{\rho}(t) + \delta\rho(t, \boldsymbol{x}) \,, \tag{1.32}$$

$$\boldsymbol{v}(t, \boldsymbol{x}) = \bar{\boldsymbol{v}}(t) + \delta\boldsymbol{v}(t, \boldsymbol{x}) \,, \tag{1.33}$$

$$\Phi(t, \mathbf{x}) = \bar{\Phi}(t) + \delta\Phi(t, \mathbf{x}) \,, \tag{1.34}$$

where $\bar{\rho}(t)$ is the background FLRW density, $\bar{\Phi}(t)$ is its associated gravitational potential and $\bar{\boldsymbol{v}}(t)$ is the velocity due to the expansion rate of the Universe given by Hubble's law $\bar{\boldsymbol{v}}(t) = H\boldsymbol{x}$.

Since the perturbations are *small* we can set the quadratic terms to zero. Moreover, the zero-th term cancels due to the FFE. Thus, one arrives at the following linearised equations for the dynamics of the perturbations:

$$\frac{d}{dt}(\delta\boldsymbol{v}) = -\frac{\nabla\delta p}{\bar{\rho}} - \nabla\delta\Phi - \delta\boldsymbol{v} \cdot H \,, \tag{1.35}$$

$$\frac{d}{dt}\delta = -\nabla \cdot \delta\boldsymbol{v} \,, \tag{1.36}$$

$$\nabla^2\delta\Phi = 4\pi G\bar{\rho}\delta \,, \tag{1.37}$$

where $d/dt = \partial_t + \bar{\boldsymbol{v}} \cdot \nabla$ and $\delta \equiv \rho/\bar{\rho} - 1$.

Given that we are interested in the gravitational collapse dynamics on top of the background expansion of the Universe, it is useful to make the expansion implicit in our coordinates such that the equations of motion only explicitly describe the perturbed dynamics. We can achieve this by introducing a series of comoving coordinates that move along the Hubble flow:

$$\boldsymbol{x}(t) \equiv a(t)\boldsymbol{\chi}(t) \,, \tag{1.38}$$

$$\delta\boldsymbol{v}(t) \equiv a(t)\boldsymbol{u}(t) \,, \tag{1.39}$$

$$\nabla_x \equiv \frac{1}{a}\nabla_\chi \,, \tag{1.40}$$

$$(\partial_t)_x = (\partial_t)_\chi - H\,\chi\,\nabla_\chi \,. \tag{1.41}$$

From now on we will assume that ∂_t and ∇ indicate the comoving time and spatial derivatives (i.e. with respect to χ) and drop the subscripts. Thus, for an observer moving alongside the expansion of the Universe, Eqs. 1.35 and 1.36 appear as:

$$\dot{\boldsymbol{u}} + 2H\boldsymbol{u} = \frac{\nabla \delta\Phi}{a^2} - \frac{\nabla \delta p}{\bar{\rho}} \, , \tag{1.42}$$

$$\dot{\delta} = -\nabla \cdot \boldsymbol{u} \; , \tag{1.43}$$

$$\nabla^2 \delta\Phi = 4\pi G a^2 \bar{\rho} \delta \, , \tag{1.44}$$

where we have introduced $\dot{} \equiv d/dt$ for simplicity.

We find ourselves in a similar situation to the one we were in when studying the FFE. The conservation equations of the fluid and its Poisson equation defines a system of 3 coupled ODEs with 4 variables: δ, $\boldsymbol{u}$, $\delta\Phi$ and δp. Similarly to then, the solution is to bring the state equation of the fluid in. Since we are interested in the large scales, we can assume that the fluid is pressureless. This because, generally speaking, the baryonic effects that would source said pressure have a much shorter effective range than gravitationally effects and thus only become relevant at small scales. Thus, we can get rid of the last term of Eq. 1.42. We can then construct a fully determined ODE for the amplitude of the matter inhomogeneities. In order to do so, let us take the gradient of Eq. 1.42:

$$\nabla \cdot \dot{\boldsymbol{u}} + 2H \nabla \cdot \boldsymbol{u} = \frac{\nabla^2 \delta\Phi}{a^2} \, . \tag{1.45}$$

Then we can take time derivative of Eq. 1.43 to obtain:

$$\ddot{\delta} = -\nabla \cdot \dot{\boldsymbol{u}} \, . \tag{1.46}$$

Thus substituting Eqs. 1.43 and 1.46 into Eq. 1.45 we can remove the dependency on $\boldsymbol{u}$. Moreover, we can use the Poisson equation to express the the right-hand side of Eq. 1.45 in terms of the background density. Putting all of this together, we arrive at the following second order, linear, homogeneous equation for amplitude of the matter inhomogeneities:

$$\ddot{\delta} + 2H\dot{\delta} - \frac{3}{2}\Omega_{\rm m}(z) H^2 \delta = 0 \, , \tag{1.47}$$

known as the Jeans equation. In order to relate the predictions of the Jeans equation to observations it is useful to rewrite it in terms of the variable $f = d\ln\delta / d\ln a$, known as the growth rate, leading to the following formulation:

$$\boxed{f' + f^2 + \left(1 + \frac{d\ln aH}{d\ln a}\right) f = \frac{3}{2}\Omega_m(z)} \, , \tag{1.48}$$

where $f' \equiv d\,f/d\ln a$ and $\Omega_m(z) = (\Omega_m H_0^2)/(a^3 H^2)$.

Since during our derivation we considered matter to be presureless fluid, our expression for the Jeans equation does not involve any scale-dependent terms [2]. Thus, inhogeneities evolve equally regardless of their scale. In other words, the evolution of matter inhomogeneities is self-similar. Therefore, the solution to the Jeans equation is a function exclusively of redshift:

$$\boxed{D(z) = \frac{\delta(z, \mathbf{x})}{\delta(0, \mathbf{x})}}\,. \tag{1.49}$$

This function is known as the linear growth factor since it describes the evolution of the matter inhomogeneities under gravitational collapse in the linear regime. Moreover, the linear growth factor can be directly related to the growth rate as:

$$\boxed{f(z) = \frac{d\ln D(z)}{d\ln a}}\,. \tag{1.50}$$

Thus it also possible, and sometimes convenient, to write to Eq. 1.48 as a system of ODEs on $D(z)$ and $f(z)$ as we will see in Sect. 2.

1.1.3 Power Spectra

In the last section, we showed how matter inhomogeneities exist and evolve in an statistically homogeneous and isotropic Universe. As a matter of fact, most of modern observational cosmology is concerned with studying the statistical properties of these inhomogeneities.

The primordial inhomogeneities onto which surrounding matter started accreting are believed to be Gaussianly distributed [7, 8]. This belief is strongly supported by observations of the Cosmic Microwave Background (CMB) [18, 19]. The CMB is the background of photons that were first able to travel freely after recombination, 380,000 years after the big bang. The inhomogeneities in the temperature and polarisation of the CMB are sourced by perturbations in the density of photons and in the gravitational potential at early times, and thus provide a picture of the state of matter in the early Universe. From then, the action of gravity has had two effects on this Gaussian field. On the one hand, the accretion of matter onto small inhomogeneities increased the overall variance of the field over time. On the other hand, since the density of a region cannot go below zero gravity, the distribution of the inhomogeneities is skewed to be dominated by empty regions or voids, leading to a

---

[2] Scale-dependent terms would come about in the form of pressure gradients

non-Gaussian distribution at late times. Thus the Universe today displays a richly non-Gaussian structure with great degree of variance.

If we stick to large enough scales the distribution of matter is still roughly Gaussian. This is because, as we saw in Sect. 1.1.2, in this regime the action of gravity is effectively linear. Since the linear combination of Gaussian variables is another Gaussianly distributed variable, linear evolution preserves the Gaussianity of the matter density field. For a Gaussian field, all its statistical information is contained in its first two moments: the mean and the variance (which will be defined formally in Sect. 1.2). Of these two statistics, we are particularly interested in studying the variance of the field for two reasons. First, the variance of the inhomogeneities directly traces the action of gravity. Second, by construction, the mean of the inhomogeneities is zero.

Before we go any further, it is important to notice that studying the variance of the matter density field of the Universe is not a particularly well posed question. This is because the concept of variance is defined as an average over a multitude of different realisations. However, there is only one Universe. Once again, the cosmological principle is key to wiggle ourselves out of this hurdle. Since we expect the Universe to be homogeneous and isotropic over large enough scales, we can assume that by averaging over large enough volumes we return the true underlying statistics of the Universe. In other words, as the volume of integration increases, we expect the volume average to resemble the ensemble average over many Universes. This equivalence between volume and ensemble averages is known as ergodicity and will come back later in the context of Monte Carlo Markov chains.

Now let us start by defining the two point correlation function of the matter density field,

$$\xi(\mathbf{x}, \mathbf{y}) = \langle \delta(\boldsymbol{x}) \delta(\boldsymbol{y}) \rangle \, , \tag{1.51}$$

where $\boldsymbol{x}$ and $\boldsymbol{y}$ are two arbitrary position vectors and $\langle \cdot \rangle$ denotes the ensemble average over all possible Universes. Since the matter density field is statistically homogeneous, it can be shown that the correlation function can only depend on the separation vector between two two points, $\mathbf{r} = \mathbf{x} - \mathbf{y}$. Moreover, the statistical isotropy of matter density field implies that there cannot be a preferred direction to the correlation function. Thus, the correlation function can only depend on the magnitude of the separation vector matters, $r \equiv |\mathbf{r}|$.

Separating the large linear scales, where the cosmological principle holds, from the small problematic scales in real space can be very challenging. An effective way of doing so is to think of the observed density field as a continuous[3] superposition of different Fourier modes:

$$\delta_{\mathbf{k}} \equiv \int d\mathbf{x}\, e^{-i\,\mathbf{k}\cdot\mathbf{x}}\delta(\mathbf{x}) \;\;\leftrightarrow\;\; \delta(\mathbf{x}) \equiv \int \frac{d\mathbf{k}}{(2\pi)^3} e^{i\mathbf{k}\cdot\mathbf{x}}\,\delta_{\mathbf{k}}. \tag{1.52}$$

Thus, we can write the previous correlation function as a Fourier expansion [20]:

$$\xi(\mathbf{r}) = \langle \delta(\boldsymbol{x})\delta(\boldsymbol{x}+\boldsymbol{r})\rangle = \left\langle \int \frac{d\mathbf{k}}{(2\pi)^3} \int \frac{d\mathbf{q}}{(2\pi)^3} \delta_{\boldsymbol{k}}\delta^*_{\boldsymbol{q}} e^{i(\boldsymbol{k}-\boldsymbol{q})\boldsymbol{x}} e^{-i\boldsymbol{k}\boldsymbol{r}} \right\rangle . \tag{1.53}$$

There are a number of important simplifications we can make to Eq. 1.53. First, we can use the completeness of the Fourier basis to invoke a Dirac delta function and remove one of the summations in Eq. 1.53 leading to:

$$\xi(\mathbf{r}) = \frac{1}{(2\pi)^3} \int dk^3\, P_{\mathbf{k}}\, e^{i\mathbf{k}\cdot\mathbf{r}} \,, \tag{1.54}$$

where

$$\boxed{\langle \delta_{\mathbf{k}}\delta_{\mathbf{q}}\rangle \equiv (2\pi)^3 \delta^{\mathcal{D}}(\mathbf{k}+\mathbf{q}) P(\mathbf{k})} \,, \tag{1.55}$$

is the matter spectral density of Fourier modes often referred to as *matter power spectrum* for brevity and $\delta^{\mathcal{D}}$ denotes the Dirac delta function. Second, we can bring back the isotropy requirement to make Eq. 1.54 a function of the magnitude of the separation vector only. Similarly, since the field is isotropic its power spectrum can only depend on the magnitude of the wave-vectors, k. Incorporating these into Eq. 1.54 we obtain:

$$\boxed{\xi(\boldsymbol{r}) = \frac{1}{2\pi^2} \int P(\mathbf{k}) \frac{\sin kr}{kr} k^2 dk} \,. \tag{1.56}$$

The question is now: what is the shape of $P(k)$? The current leading theory for the early expansion of the Universe, Inflation, predicts that the primordial (i.e. at times shortly after the beginning of the Universe) matter power spectrum should be a featureless power law[4]:

$$P(k) \propto k^{n_s} \,, \tag{1.57}$$

where n_s is known as the spectral index and controls the prominence of small versus large scales. Moreover, Inflation requires a roughly *scale invariant* primordial matter power spectrum, meaning $n_s \sim 1$. Recent observations of the CMB [13, 19] have found the spectral index to be just below one, as Inflation would suggest.

---

[3] The expression in Eq. 1.52 becomes continuous as a result of imposing periodic boundary conditions in the limit of infinitely large volumes. This is done to incorporate fact the Universe is indeed infinite!

[4] This is because the accelerated expansion of Inflation has no preferred scale, i.e. all the scales evolve for the same amount of time.

Fortunately for careers in Cosmology, the matter power spectrum is not a completely featureless power law. While Eq. 1.57 is a good starting point, $P(k)$ actually possess a rich range of features due to the interactions of dark and baryonic matter as well as radiation at different stages of the evolution of the Universe. Moreover, we expect the matter power spectrum to evolve in time as gravity progressively shifts power from the larger scales into the smaller scales.

Predicting the original shape of these features and their evolution until today is a theoretically challenging task. To correctly include large scales, we must work with relativistic perturbation theory. Furthermore, treating matter as a perfect fluid is no longer accurate. Instead, we have to consider the phase space distribution of matter $W(\boldsymbol{x}, \boldsymbol{p})$ which describes the probability of finding a number of particles at a given position $\boldsymbol{x}$ with a given momentum $\boldsymbol{p}$. The evolution of $W(\boldsymbol{x}, \boldsymbol{p})$ is given by the Boltzmann-Einstein equation:

$$\left(p^\mu \frac{\partial}{\partial x^\mu} - \Gamma^\mu_{\alpha\beta} p^\alpha p^\beta\right) W = C\,, \tag{1.58}$$

where Γ is the Levi-Civita connection of the space-time metric and C contains the scattering physics of the matter species being evolved (Compton scattering, breaking radiation... etc). Solving this system involves a set of coupled partial differential equations, and can only be done numerically [21]. Computer codes that solve the Einstein-Boltzmann equations are known as Boltzmann codes. Currently the two most popular cosmological codes are CAMB [22] and CLASS [23].

Hopefully the paragraph above has been intimidating enough to dissuade the reader from seeking a thorough description of how $P(k, z)$ is actually computed since we will not provide one. As a matter of fact, even the approach we described above is not correct on sufficiently small scales. Modern cosmological analyses often rely on running actual N-body simulations of the different matter species of the Universe from which $P(k, z)$ is measured as opposed to directly predicting the matter power spectrum.

However, it turns out that against all odds we can make an educated guess on the shape of $P(k, z)$ that is good enough to fit a significant fraction current data in Cosmology without the need of Boltzmann codes or N-body simulations. As we saw in Sect. 1.1.2, right after Inflation the matter power spectrum was a featureless power law. However, once Inflation was over, the different matter species re-entered the horizon decaying at different rates. This induced a series of series of features in the matter power spectrum. In order to capture this effects, we need to multiply the original matter power spectrum by a *Transfer* function, that bridges the gap between the end of Inflation and today. Thus, the matter power spectrum today can be approximated as:

$$\boxed{P_0(k) \propto \left(\frac{k}{H_0}\right)^{n_s} T^2(k)}\,, \tag{1.59}$$

where $T^2(k)$ is the aforementioned Transfer function. The most popular prescriptions for $T(k)$ are Bardeen *et al.* [24] and Eisenstein and Hu [25, 26].

In order to compute the matter power spectrum at a given redshift we can then make use of the fact that the evolution of the matter inhomogeneities in the linear regime is scale-independent (See Sect. 1.1.2) such that:

$$P(k, z) \propto P_0(k) D^2(z)\,, \tag{1.60}$$

where $D(z)$ is the linear growth factor defined in Sect. 1.1.2.

The only thing left is to find the overall amplitude of $P_0(k)$ which is directly linked to the amplitude of the primordial inhomogeneities. Since we don't have access to this information a popular approach is to normalise $P_0(k)$ at an arbitrary scale and then rescale it by the measured value from our observations. It has become customary in Cosmology to normalise the matter power spectrum by its variance within spheres of $8/h$[Mpc] radius where h is the dimensionless Hubble parameter defined as $H_0 = h * (100\,[\mathrm{km/s/Mpc}])$. Thus, this normalisation constant is known as σ_8. In order to do so we need to compute the filtered variance using the convolution theorem:

$$\sigma_8^2 \equiv \frac{1}{(2\pi)^2} \int dk\, k^2 P_0(k) |W_k|^2\,, \tag{1.61}$$

where W_k is the Fourier transform of the filter. In our case the filter is a top-hat function given by:

$$W_k = \frac{3}{y^3} (\sin(y) - y \cos y)\,, \tag{1.62}$$

where $y \equiv k \cdot 8/h$[Mpc].

Putting all of this together, we have derived a linear approximation to the real matter power spectrum, accurate on the largest scales. However, modern surveys observe the matter power spectrum well beyond the linear regime. In order to use Eq. 1.59 to match modern observations we must include non-linear effects. Thankfully, the literature contains a plethora of phenomenological prescriptions of what the impact of the non-linear terms should be on Eq. 1.59. The most popular prescription is the Halofit formula [27, 28] which was derived by fitting the measured matter power spectrum in N-body simulations. We show the impact of this non-linear effects on the linear matter power spectrum in Fig. 1.2. As one might expect, including these

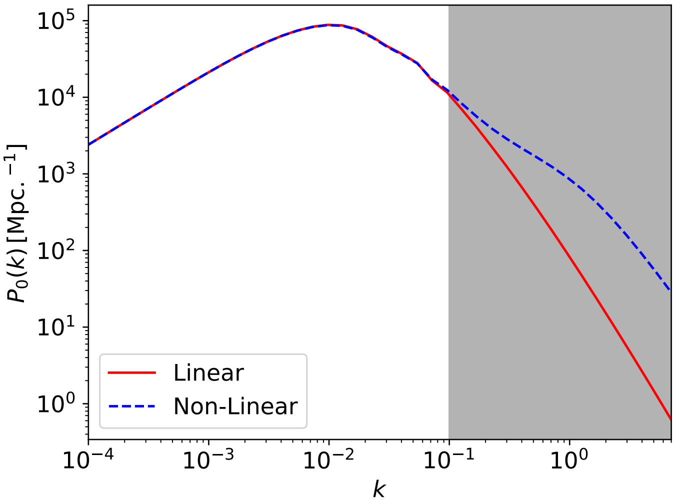

Figure 1.2: The impact of non-linear effects on the linear matter power spectrum as described by the Halofit formula. The grey zone represents the region beyond which the linear approximation fails.

non-linear interactions raises the power of the smallest scales at which gravitational accretion is enhanced. Moreover, we can observe that the linear approximation cannot be trusted for $k < 0.1$ [Mpc^{-1}].

1.1.4 Angular Power Spectra

In the previous section, we discussed how to build theoretical predictions for the expected statistical properties of the matter density field. The question now is: how can we relate these theoretical predictions to actual observations? In order to answer this question, there is a fundamental obstacle that we have to overcome: our telescopes do not directly observe the matter density field. Thus, cosmologists often talk of *tracers* of the matter density field.

A natural way of thinking of these tracers is as sampling 3D fluctuations of the matter density field. While fully 3D approaches have been developed [29], the current uncertainty in the redshift of the objects that constitute the tracer often renders this approach unfeasible. Thus it is common for cosmological analyses to project the observed fields on the celestial sphere by integrating over the line-of-sight with a radial selection function $q_f(\chi)$ [30]:

$$f(\hat{\mathbf{n}}) = \int d\chi \, q_f(\chi) \, F(\chi\hat{\mathbf{n}}, z), \tag{1.63}$$

where $\hat{\mathbf{n}}$ is a unit vector on the sphere and z is the redshift corresponding to the comoving distance χ such that the integral is taken along the lightcone (i.e. it accounts for the time evolution of the field along the photons' paths). From now on, we will refer to the radial selection functions as the tracers' *kernels* since they are responsible for establishing the relationship between the observed field and the underlying matter density field.

Following this formalism we can derive an expression for the *angular* correlation function (ξ_{fg}), meaning the correlation function of two projected quantities, given a 3D correlation function (ξ_{FG}):

$$\xi_{fg}(\theta) = \int d\chi_1 q_f(\chi_1) \int d\chi_2 q_g(\chi_2)\, \xi_{FG}\left(\sqrt{\chi_1^2 + \chi_2^2 - 2\chi_1\chi_2\cos\theta}\right), \tag{1.64}$$

where q_f and q_g are radial selection functions of the two fields being correlated. As shown in Eq. 1.56, we can perform a Fourier expansion of the correlation function to derive a similar expression for the *angular* power spectrum:

$$C_\ell^{fg} = \frac{2}{\pi} \int d\chi_1 q_f(\chi_1) \int d\chi_2\, q_g(\chi_2) \int_0^\infty dk\, k^2 j_\ell(k\chi_1) j_\ell(k\chi_2)\, P_{FG}(k). \tag{1.65}$$

Note that since we are working now on the sphere, instead of a Fourrier expansion, we need to perform an expansion in spherical harmonics. Thus, Eq. 1.65 features j_ℓ, the spherical Bessel functions. Since the radial dimension has been integrated, Eq. 1.65 is often referred to as the *angular* power spectrum.

Thus computing 2D projections of 3D fields involves performing a costly series of nested integrals over highly oscillatory functions, involving $j_\ell(x)$. We can, though, leverage the properties of the integrands in Eq. 1.64 and 1.65 to simplify this process. The fundamental realisation is that the value of the correlation function is only significant when the points being correlated are *close by.* Moreover, for most cosmological fields, the radial selection functions are much smoother than the correlation function or the Bessel functions, meaning that they are approximately constant across small intervals. This set of assumptions is known as the Limber approximation [5] [31, 32].

In order to formalise our assumptions, allow us to reparameteriseEq. 1.64 in terms of $\chi_{1,2} = \chi \pm r/2$:

$$\xi_{fg}(\theta) = \int d\chi\, dr\, q_f(\chi - r/2)\, q_g(\chi + r/2)\, \xi_{FG}\left(\sqrt{2\chi^2(1-\cos\theta) + r^2(1+\cos\theta)/2}\right). \tag{1.66}$$

---

[5] Note that these are two conflicting assumptions. On the one hand, we want sufficiently large scales such that we can trust our estimate of the non-linear matter power spectrum. However, if the scales become too large we can no-longer trust the the Limber approximation. Thus, these assumptions define an upper and lower bound for the scales our models can actually fit.

Now, taking the limit in which $r, \theta \to 0$, the tracers' kernels are approximately constant meaning that $q_{f,g}(\chi \pm r/2) \simeq q_{f,g}(\chi)$. Moreover, in this limit $\cos\theta \simeq 1 - \theta^2/2$. Plugging these into our previous expression we get:

$$\xi_{fg}(\theta) \simeq \int_0^{\infty} d\chi q_f(\chi) q_g(\chi) \int_{-\infty}^{\infty} dr\, \xi_{FG}\left(\sqrt{\chi^2\theta^2 + r^2}\right). \tag{1.67}$$

If we now focus on the power spectrum, the spherical Bessel functions $j_\ell(x)$ at high ℓ (i.e. for close separations in real space) are sharply peaked around $x \sim \ell + 1/2$. Thus we can approximate them as:

$$j_\ell(x) \simeq \sqrt{\frac{\pi}{2\ell+1}}\delta^{\mathcal{D}}(\ell + 1/2 - x), \tag{1.68}$$

where the prefactor is added to preserve the normalisation of the Bessel functions as a basis and $\delta^{\mathcal{D}}$ is the Dirac delta function. Applying this to Eq. 1.65 leads us to the much simpler expression for the angular power spectrum:

$$\boxed{C_\ell^{fg} = \int \frac{d\chi}{\chi^2} q_f(\chi) q_g(\chi) P_{FG}\left(\frac{\ell + 1/2}{\chi}, z\right)}. \tag{1.69}$$

Having derived an expression for the angular power spectrum in terms of two arbitrary tracers, our goal now is to specify what these tracers look like. For the purposes of this thesis we will focus on two particular tracers of the matter density field, galaxy clustering and weak gravitational lensing.

1.1.4.1 Galaxy Clustering

Galaxy Clustering (GC) carries information about the inhomogeneity of the underlying matter distribution. The fundamental idea is that the formation of galaxies requires the presence of deep gravitational wells. Thus galaxies form in the densest regions of the matter density field. Hence galaxies constitute a biased tracer of the matter density distribution. On large scales, we expect such bias to be linear [33]. Thus we can establish the following relationship between the measured 3D galaxy-galaxy power spectrum and the matter power spectrum:

$$\delta_g(\boldsymbol{x}) \simeq b_g \delta(\boldsymbol{x}), \tag{1.70}$$

where δ_g is the galaxy overdensity field and b_g is a linear bias parameter.

Incorporating this linear bias into the tracer kernel, the GC radial selection function is given by:

$$\boxed{q_g(\chi) = b_g p(z) \frac{dz}{d\chi}}, \tag{1.71}$$

where $p(z)$ is the distribution of galaxies along the line of sight normalised to one and $dz/d\chi$ relates the distributions of galaxies defined as a function of redshift to the tracer kernel defined as a function of radial distance.

1.1.4.2 Weak Lensing

Gravitational lensing is the phenomenon in which the light emitted by an object is deflected by the gravitational field of matter across its path. Weak gravitational lensing, or Weak Lensing (WL) for short, refers to the regime where said effects contribute a mere few percentages to the observed image. Since WL is associated with weak gravitational fields we can model it by considering the linear perturbations of the FLRW metric.

Within this framework, it can be shown that the travel time of a photon propagating through an inhomogenous Universe is given by:

$$t = \int a(\chi)(1 - \Phi(\chi))d\chi \, . \tag{1.72}$$

Analogous to traditional optics, the potential in Eq. 1.72 acts as a medium with variable refractive index, deflecting the direction of light across its path. Hence, the phenomenon is known as gravitational lensing. Applying Fermat's principle to Eq. 1.72, one can derive an equation for the deflection of the angle:

$$\hat{\boldsymbol{\alpha}} = -\int \nabla_\perp \Phi d\chi \, , \tag{1.73}$$

where $\nabla_\perp$ is the gradient along the perpendicular direction to the light path.

In order to derive an expression for the weak lensing kernel, let us consider the separation vector between two photons converging onto an observer which have been distorted by a source at radial distance χ' at a perceived angle $\boldsymbol{\theta}$. This separation will be given by a combination of the distance suggested by the apparent angle (assuming no distortion) minus the lensing distortion:

$$\boldsymbol{x}(\chi) = \boldsymbol{x}_0(\chi) + \int_0^\chi (\chi - \chi')d\hat{\boldsymbol{\alpha}} \tag{1.74}$$

$$= \chi\boldsymbol{\theta} - \int_0^\chi (\chi - \chi')\nabla_\perp \Phi(\boldsymbol{\theta}, \chi')d\chi' \, . \tag{1.75}$$

Dividing the previous expression by the radial distance we can write the following lensing equation:

$$\boldsymbol{\beta} = \boldsymbol{\theta} - \boldsymbol{\alpha} \, , \tag{1.76}$$

where $\boldsymbol{\theta}$ is the observed angular separation between the sources, $\boldsymbol{\beta}$ is the true angular separation and $\boldsymbol{\alpha}$ is the scaled deflection angle given by the following integrated effect:

$$\boldsymbol{\alpha}(\chi) = \int_0^{\chi} \frac{(\chi - \chi')}{\chi} \nabla_{\perp}\Phi(\boldsymbol{\theta}, \chi')d\chi' \,. \tag{1.77}$$

Therefore, gravitational lensing can be understood as a differentiable map between real and the distorted images given by the Jacobian $\boldsymbol{A} = \partial\boldsymbol{\beta}/\partial\boldsymbol{\theta}$. This transformation is often represented as an amplification matrix with components:

$$A_{ij} = \frac{\partial\beta_i}{\partial\theta_j} = \delta_{ij}^{\mathcal{D}} - \frac{\partial\alpha_i}{\partial\theta_j} = \tag{1.78}$$

$$\delta_{ij}^{\mathcal{D}} - \partial_i\partial_j\psi \,, \tag{1.79}$$

where $\partial_\theta = \frac{1}{\chi'}\partial_x$ and we have introduced the lensing potential:

$$\psi(\boldsymbol{\theta}, \chi) = \int_0^{\chi} \frac{(\chi - \chi')}{\chi\chi'} \Phi(\boldsymbol{\theta}, \chi')d\chi' \,. \tag{1.80}$$

Thus we can write the components of the amplification matrix explicitly in terms of the lensing potential:

$$A_{ij} = \begin{pmatrix} 1 - \kappa - \gamma_1 & -\gamma_2 \\ -\gamma_2 & 1 - \kappa - \gamma_1 \end{pmatrix} \,, \tag{1.81}$$

where

$$\kappa = \frac{1}{2}(\partial_1\partial_1 + \partial_2\partial_2)\psi \,, \tag{1.82}$$

$$\gamma_1 = \frac{1}{2}(\partial_1\partial_1 - \partial_2\partial_2)\psi \,\text{and} \tag{1.83}$$

$$\gamma_2 = \partial_1\partial_2\psi \,. \tag{1.84}$$

Therefore WL can be broken down in two distinct effects. On the one hand, we have the convergence, κ, responsible for the isotropic magnification or reduction of the size of the observed objects. On the other hand, we have γ, the shear, responsible for shape distortions. Shear is often written an as complex quantity given by $\gamma = \gamma_1 + i\gamma_2 = |\gamma|e^{2i\phi}$ where ϕ is a complex phase. Thus it is easy to see that shear transforms as a spin-2 quantity which is to be expected given the symmetry properties of the elliptical shapes of galaxies.

Our goal now is to relate the previously derived expression for the convergence and shear components to the underlying matter density field sourcing the lensing.

In principle, the two lensing components need to be treated differently since convergence is a scalar quantity while shear is a spin-2 quantity[6]. While scalars are rotationally invariant. However, it turns out that we can model the two components identically. Any spin-s quantity can be associated with a scalar quantity, the E-mode, and a pseudo-scalar quantity, the B-mode. Since lensing is the result of the second derivative of a scalar quantity (i.e. the lensing potential defined in Eq. 1.80), the convergence is identical to the E-mode of the shear. Moreover and for the same reason, the B-mode is exactly zero. Thus both convergence and shear can be modelled indistinguishably [34].

Armed with this knowledge, let us consider the convergence definition in Eq. 1.82. Moving the Laplace operator inside the lensing potential we obtain:

$$\kappa(\boldsymbol{\theta}, \chi) = \int_0^{\chi} \frac{(\chi - \chi')}{\chi\chi'} \nabla_{\perp}^2 \Phi(\boldsymbol{\theta}, \chi) d\chi' \,. \tag{1.85}$$

We can then use the Newton-Poisson equation (Eq. 1.44) to relate the Laplacian of the gravitational potential to the matter inhomogeneities directly [7]. Thus, we obtain the following expression.

$$\kappa(\boldsymbol{\theta}, \chi) = \frac{3H_0^2\Omega_{\mathrm{m}}}{2} G_\ell \int_0^{\chi} \frac{d\chi'}{a(\chi')} \frac{(\chi - \chi')}{\chi} \chi' \delta(\boldsymbol{\theta}\chi', \chi') \,, \tag{1.86}$$

where

$$G_\ell \equiv \sqrt{\frac{(\ell+2)!}{(\ell-2)!} \frac{1}{(\ell+1/2)^2}} \,, \tag{1.87}$$

is a scale dependent prefactor needed to transform the 2D angular Laplacian in the lensing equation to the 3D Laplacian featured in the Newton-Poisson equation. However, Eq. 1.86 is a 3D expression. In order to obtain a projected quantity we need to average Eq. 1.86 over the distribution of sources along the line of sight:

$$\kappa(\boldsymbol{\theta}) = \int_{z(\chi)}^{\infty} p(z) \kappa(\boldsymbol{\theta}, \chi(z)) dz \,. \tag{1.88}$$

Exchanging the order of the integrals in Eq. 1.88, we can derive an expression for the convergence tracer as a projection of the 3D matter density field:

$$\kappa(\boldsymbol{\theta}) = \int_0^{\chi} q_\gamma(\chi') \delta(\boldsymbol{\theta}\chi', \chi') d\chi' \,, \tag{1.89}$$

---

[6] The spin of a quantity describes its properties under rotations. Scalar quantities (such as point quantities) are invariant under rotations and possess spin-0. Spin-n quantities are invariant under $2\pi/n$ radians rotations.

[7] The Laplacian inside Eq. 1.85 is the transverse Laplacian rather than full Laplacian featured in the Poisson equation. However, we can seamlessly exchange the two since the radial component of the full Laplacian integrates out to zero by parts.

where q_γ is the kernel of the convergence/shear tracer defined as:

$$\boxed{q_\gamma(\chi) \equiv G_\ell \frac{3H_0^2\Omega_\mathrm{m}\chi}{2a(\chi)} \int_{z(\chi)}^{\infty} dz' n(z') \frac{\chi(z')-\chi}{\chi(z')}} \,. \tag{1.90}$$

From now on we will refer to this particular kernel as the *cosmic shear* kernel since, observationally, gravitational lensing is associated with measurements of shape distortions of galaxies.

A special case of the gravitational lensing effect is the distortion of the CMB photons. As these photons propagate through the Universe towards us, their trajectories are deflected via lensing, giving rise to secondary distortions in the CMB. These can be used to reconstruct the intervening gravitational potential.

Modelling the gravitational lensing of the CMB (also known as CMB lensing) is remarkably similar to modelling the gravitational lensing of photons emitted by nearby galaxies. The key difference is that in the case of CMB lensing the distribution of sources along the line of sight is given by a delta function:

$$p(z) = \delta^{\mathcal{D}}(\chi(z) - \chi^*(z)) \,, \tag{1.91}$$

where χ^* is the radial distance associated with the redshift at which photons decoupled also known as the surface of last scattering. Thus, substituting Eq. 1.91 into Eq. 1.90 we can derive the following expression for the CMB lensing tracer:

$$\boxed{q_\kappa(\chi) \equiv K_\ell \frac{3H_0^2\Omega_\mathrm{m}\chi}{2a(\chi)} \frac{\chi^*-\chi}{\chi^*}} \,, \tag{1.92}$$

where

$$K_\ell \equiv \frac{\ell(\ell+1)}{(\ell+1/2)^2} \tag{1.93}$$

is the equivalent term to G_ℓ in Eq. 1.90.

Similarly to how galaxy gravitational lensing is observationally associated with the shear component, CMB lensing is observationally associated with the convergence component since it is measured in the form of the magnifications in the CMB inhomogeneities. Accordingly we refer to Eq. 1.92 as the CMB convergence tracer.

1.1.5 Tomography and Systematics

As we saw in the last section, even when treated as 2D projected quantities, both GC and WL are related to the observed distribution of galaxies along the line of sight.

However, determining this radial distribution is a difficult task since we often only have access to inaccurate estimates of the redshifts of these sources.

Astronomers produce two different types of redshifts in galaxy catalogues: spectroscopic redshifts and photometric redshifts. The distinction between spectroscopy and photometry can be summarised as quality versus quantity [35]. On the one hand, spectroscopic catalogues measure the spectral energy distribution (SED) of the light of each source allowing them to accurately determine the associated redshift. However, the exposure time required to do so means that only a relatively small number of sources can be resolved. Moreover, spectroscopic redshifts tend be biased towards certain galaxy populations in terms of stellar mass and magnitude, as well as galaxies with emission lines. On the other hand, photometric catalogues are composed of images filtered through a small number of colour bands, needing a much smaller exposure time to resolve each source than spectroscopic catalogues. Thus, they contain orders of magnitude more objects. However, the measurements of the SEDs are much coarser, leading to a far greater uncertainty in the sources' redshifts.

Cosmology, by its nature, is bound to the realm of large numbers in order to effectively beat shot noise. This is specially true in WL analyses where very large number of galaxies are necessary to beat the shape measurement noise. Thus, a large fraction of the catalogues used in modern Cosmology are photometric. This means that there is a large degree of uncertainty in the observed $p(z)$.

One way in which cosmological analyses address this issue is by binning the distribution of galaxies in redshift bins, known as tomographic bins. However, it is often difficult to calibrate said bins (meaning to assign each galaxy to its correct bin) due to degeneracies between the photometric redshift estimate and other properties of galaxies. Since WL is already an integrated effect, broad photometric bins, that are easier to calibrate but that wash off the radial information, are often used. In the case of GC, narrower bins are normally beneficial. This allows us to measure the effective value of the angular power spectra at the mean redshift of the tomographic bin. Moroever, given the large uncertainties on $p(z)$, most of the information is also captured [36].

Binning the data by approximate redshift also allows us to model systematic effects in a much more tractable way. The first and most common systematic we have to address is the uncertainty in the radial distribution of galaxies. The way this uncertainty is most often captured by cosmological analyses is by introducing a shift parameter between the observed and the true distribution of galaxies per tomographic bin:

$$p(z) = \tilde{p}(z + \Delta z)\,, \tag{1.94}$$

where $\tilde{p}(z)$ is the observed radial distribution of galaxies and Δz is a free shift parameter. The range of Δz can be estimated by comparing our photometric estimate of $p(z)$ with smaller samples of spectroscopic galaxies via direct calibration, [37, 38], clustering redshifts [39–42] or and shear ratios [43, 44] among other techniques. However it has been argued that this methodology fails to adequately capture the uncertainty in the radial distribution of galaxies, specially at high redshift where spectroscopic observations are extremely sparse [45, 46].

There are reasons to believe that one single shift parameter per tomographic bin might not be enough to encapsulate the whole of our uncertainty on $p(z)$ [5]. For example, models with an additional width parameter, w_z, have been studied [4]:

$$p(z) = \tilde{p}(z_c + w_z(z - z_c) + \Delta z), \tag{1.95}$$

where z_c is the redshift at which $\tilde{p}(z)$ peaks. However, as we will see in Sect. 5, more complex models can also be considered.

In addition to this, there are systematics specific to each tracer. In Sect. 1.1.4.1, we saw that in Eq. 1.70 the galaxy distribution was related to the matter distribution by a simple linear bias parameter. In the context of tomography, this implies that we consider one bias parameter per tomographic bin in a clustering analysis. Such a linear parameter was argued to be enough when the analysis was limited to large scales ($k < 0.15$ [Mpc^{-1}]). However, future surveys will go well beyond this limit.

In order to rise to this challenge, cosmological analyses have come up with two strategies. The first method consists on going beyond the linear bias by considering an effective field theory of galaxy formation [47]. Over mildly non-linear scales, one can relate Δ_g with scalar combinations of the Hessian of the gravitational potential $\partial_i \partial_j \Phi$. Following McDonald and Roy [48] and [49], the second order bias expansion is given by:

$$\Delta_g(\boldsymbol{x}) = b_1 \delta(\boldsymbol{x}) + \frac{b_2}{2!}(\delta(\boldsymbol{x})^2 - \langle \delta(\boldsymbol{x})^2 \rangle) + \frac{b_s}{2!}(s^2 - \langle s^2 \rangle) + b_\nabla \nabla^2 \delta(\boldsymbol{x}). \tag{1.96}$$

Here $s^2 \equiv s_{ij} s^{ij}$ is the trace of the squared tidal tensor, where $s_{ij} \equiv \partial_i \partial_j \Phi - \nabla^2 \Phi / 3$. The quantities b_1, b_2, b_s, and b_∇ are the so-called “linear”, “quadratic”, “tidal”, and “non-local” bias parameters, which characterise the response of the galaxy overdensity to the corresponding terms in perturbation theory.

The second approach uses N-body simulations of galaxy formation to establish an empirical relationship between the observed galaxy field and the underlying matter density. Then an emulator, i.e. an easier to evaluate approximate model to the original function, can be built relating the two quantities [50]. Both approaches however increase the number of parameters that need to considered in each tomographic bin.

Moving on to weak lensing, the first systematic that we have to think about is the potential mismatch between the observed galaxy image and the actual shape of the galaxy. This is often taken into account in the form of one multiplicative bias parameter per tomographic bin that modifies the WL radial kernel as:

$$(1+m)q_\gamma(\chi)\,, \tag{1.97}$$

where q_γ is given by Eq. 1.90. Given current error budgets, the contribution of multiplicative biases to our theoretical predictions is negligible [51]. Hence, there is currently no need to consider more complex models.

More fundamental to WL is the phenomenon of intrinsic alignments. Galaxy shapes are not only correlated by cosmic shear but also by intrinsic alignments (IAs) in the orientation of galaxies due to local interactions. Within the so-called Non-Linear Alignment [NLA, 52] model this can be accounted for by adding an extra contribution to the final shear kernel given by:

$$q_I(\chi) = A_{\rm IA}(z)H(z)p(z), \tag{1.98}$$

where $A_{\rm IA}(z)$ is:

$$A_{\rm IA}(z) = A_{\rm IA,0}\left(\frac{1+z}{1+z_0}\right)^{\eta_{\rm IA}}\frac{0.0139\Omega_{\rm m}}{D(z)}\,, \tag{1.99}$$

where $A_{\rm IA,0}$ and $\eta_{\rm IA}$ are two free parameters, z_0 is a redshift pivot (which we fix to $z_0 = 0.62$ as in [51, 53]), and $D(z)$ the linear growth factor. Thus, the final WL radial kernel is given by:

$$q_I + (1+m)q_\gamma(\chi)\,. \tag{1.100}$$

It is worth noting that even when the analysis is performed tomographically, only a set of intrinsic alignment parameters are shared between all the bins.

One can imagine that the contribution of all these systematic effects obfuscates the role of the cosmological parameters of interest. One way in which modern cosmological analyses overcome this issue is by analysing together cross-correlations of different tracers. The reason why joint analyses are helpful is because, while different tracers have different systematics, they all depend on the same underlying cosmological parameters. This allows us to isolate the impact of the different contributions. The most popular of these combinations is the so called 3x2-pt analysis

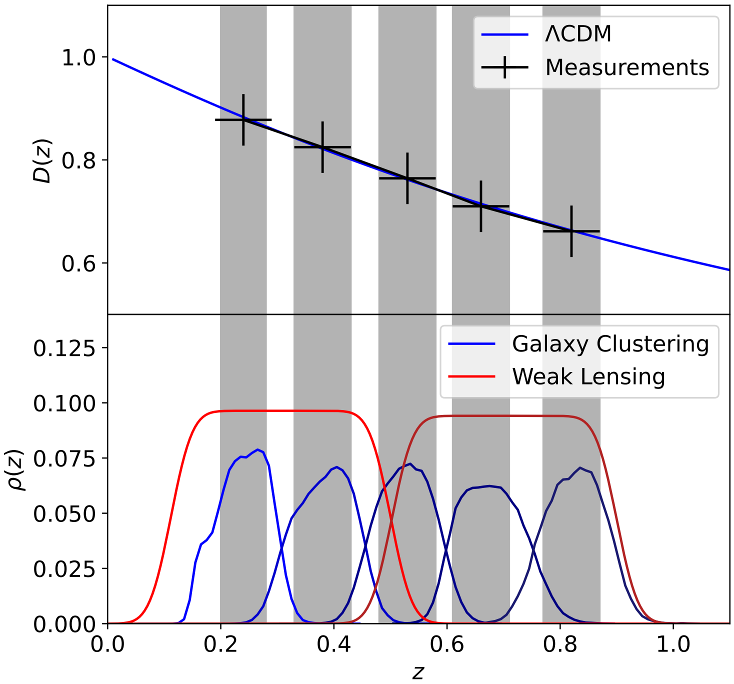

Figure 1.3: Bottom panel shows the binned $p(z)$'s of an example galaxy clustering (blue) and weak lensing (red) survey. Top panel shows the ΛCDM prediction for the growth factor, $D(z)$ (See Eq. 1.49), as well as the measurements from the different tomographic bins of the survey.

where WL auto-correlations are analysed together with GC auto-correlations as well as the cross-correlation between GC and WL.

This style of tomographic analysis has very interesting properties. Not only do they allow us to combine extremely large numbers of data and disentangle systematic effects, they also allow us to measure phenomena across cosmic time. In Fig. 1.3, we show an example of how tomography can be used to measure cosmological functions as a function of redshift. Bottom panel shows the binned $p(z)$'s of an example GC (blue) and WL (red) survey. Top panel shows the ΛCDM prediction for the growth factor, $D(z)$ (See Eq. 1.49), as well as the measurements from the different tomographic bins of the survey. Precise measurements can be obtained even from WL data despite their broad bins by cross-correlating them narrower GC bins. This has been done to great success for a variety of cosmic phenomena [54–57].

However, performing this combination of data also presents major challenges. The first challenge is how to combine all the different tracers in a statistically consistent manner [20, 58, 59]. Effectively, what we want to avoid is *double counting* the same galaxies. In practice, this is a rather convoluted exercise since each survey observes a limited fraction of the sky with a non-trivial shape which may or may not overlap, partially or completely, with other surveys. Thus the cross-covariance between the different fields has to be carefully computed.

The second challenge is brought about by the very same number of tomographic bins involved in these analyses. The problem stems from the fact that each tomographic bin will have a number of free parameters associated with its different systematics. Parameters associated with the modelling of systematics are often referred to by cosmologists as nuisance parameters. This is because their value is often uninteresting but they must be included in the analysis so as not to overestimate our confidence in the parameters we ultimately care about.

Current 3x2-pt analyses manage of order of 5 to 10 tomographic bins which lead to a total 20 to 30 free parameters when including both cosmological and nuisance parameters [13, 51, 60, 61]. The next generation of surveys, known as stage IV, have to consider several folds more parameters for two reasons. First Stage-IV surveys will provide data of unprecedented *quality*. This means that all the models considered in this section will have to be extended to encompass smaller scales to higher degree of precision. These complex models will inexorably bring along higher numbers of nuisance parameters per tomographic bin. In addition to this, Stage-IV will also bless us with unprecedented *quantities* of data. Thus, we expect a several fold increase in the number of tomographic bins considered by each individual survey. The combination of more tomographic bins with larger numbers of nuisance parameters per bin will result in 3x2-pt analyses of Stage-IV surveys having to consider unprecedented numbers of free parameters. Finally, note that one does not need to stop at 3x2-pt and that larger numbers of correlations have and will be considered [54, 62]. Managing these large number of parameters will be one of the greatest challenges of Cosmology in the future.

1.2 Statistical Inference

1.2.1 Traditional Inference Algorithms

As we saw towards the end of Sect. 1.1.5, Cosmology is currently sitting on a ticking time bomb that will result in an unprecedented explosion in the quantity and quality of data. In order to match the quality of the data, physicists are starting to incorporate into their theoretical predictions more of the physical, observational and instrumental effects which, until now, could be overlooked. In practice, this translates into a dramatic increase in the number of parameters that future analyses will have to consider. This combination of large data sets with complex models will (and in many cases already does) overwhelm the inference methods we currently use to constrain the values of these parameters.

The fundamental problem of Cosmology is that there is only one Universe [8]. To date, cosmologists have not figured out how to reproduce the birth, life and death of the Universe in a laboratory. This lack of reproducible experiments means that the only way we can learn about our Universe is through the limited observations our telescopes provide us with.

Statistical inference is the exercise of learning the properties of a population given a limited sample [63]. In Cosmology, this translates into learning the properties of the Universe from a small fraction (millions) of galaxies. These galaxies, however, all stem from the same Universe *realisation.* One can imagine a distribution over all possible Universes in terms of some cosmological parameters, $P(\boldsymbol{\theta})$. Within this framework, our Universe is nothing more than a particular draw from such a distribution [9]. If one had access to $P(\boldsymbol{\theta})$, it would be possible to quantify the likelihood of the particular properties of our Universe. This is the end goal of Cosmology.

If the Universe were reproducible, we could directly learn $P(\boldsymbol{\theta})$ using a frequentist approach by repeating such a hypothetical experiment many times. The limitations brought about the Universe's singular nature make Cosmology a profoundly Bayesian discipline. The key notion of Bayesian inference is that one can make up for the lack of draws by coming up with a model for the phenomenon being observed [64, 65]. In Cosmology we often refer to such models as cosmological models.

One can then derive an expression for $P(\boldsymbol{\theta})$ by considering the joint probability of the parameters and the data, $P(\boldsymbol{\theta}, \mathbf{d})$, and expanding it in terms of conditional

---

[8] We briefly discussed this issue in the context of computing the variance of the matter density field in Sect. 1.1.3

[9] The idea of a distribution over possible Universes was originally proposed by Leibniz who argued we must live in the best possible one of them.

probabilities. Since, by definition, $P(\boldsymbol{\theta}, \mathbf{d}) = P(\mathbf{d}, \boldsymbol{\theta})$, it then follows that:

$$P(\boldsymbol{\theta}, \mathbf{d}) \equiv P(\boldsymbol{\theta}|\mathbf{d})P(\mathbf{d}) = P(\mathbf{d}|\boldsymbol{\theta})P(\boldsymbol{\theta})\,. \tag{1.101}$$

Isolating $P(\boldsymbol{\theta}|\mathbf{d})$ and renaming the different probabilities directly leads to the famous Bayes's theorem:

$$\boxed{P(\boldsymbol{\theta}|\boldsymbol{d}) = \frac{\mathcal{L}(\boldsymbol{d}|\boldsymbol{\theta})\Pi(\boldsymbol{\theta})}{\mathcal{Z}(\boldsymbol{d})}}\,. \tag{1.102}$$

In this equation, $P(\boldsymbol{\theta}|\boldsymbol{d})$ is the distribution of the parameters, $\boldsymbol{\theta}$, given the data $\boldsymbol{d}$ and is known as the posterior. $\mathcal{L}(\boldsymbol{d}|\boldsymbol{\theta})$ is the likelihood of the data given the parameters and is commonly referred to as the likelihood. Note that the likelihood hides a dependency on the model chosen to fit the observations, $\mathcal{L}(\boldsymbol{d}|m(\boldsymbol{\theta})) \equiv \mathcal{L}(\boldsymbol{d}|\boldsymbol{\theta})$. $\Pi(\boldsymbol{\theta})$ is our belief in the distribution of parameters prior to the observations and is referred to as the prior. Finally,

$$\mathcal{Z} = \int \mathcal{L}(\boldsymbol{d}|\boldsymbol{\theta})\Pi(\boldsymbol{\theta})d\boldsymbol{\theta}\,, \tag{1.103}$$

is known as the evidence and it is an integral over all the possibilities of the model. The evidence is often ignored since computing it is very computationally challenging and its contribution amounts to a normalisation factor on the posterior. However, it is very useful when performing comparisons between different models. Inside Eq. 1.103 there are two competing factors. On the one hand, the contribution of the likelihood to the integral will increase, the better the model fits the data. On the other hand, the contribution of the prior will decrease the more complex the model is. This means that the evidence is maximum for models that strike a balance between goodness of fit and simplicity[10].

Cosmologists thus use Bayesian statistics to make statements about the properties of the Universe. As in any other probabilistic problem, anything that can be said about the values of cosmological parameters must be spoken in the language of expectation values also known as moments:

$$\boxed{\mathbb{E}(\boldsymbol{\theta}_{\mathrm{i}}^{n}) = \int \theta_{\mathrm{i}}^{n} dP(\boldsymbol{\theta}|\boldsymbol{d}) = \int \theta_{\mathrm{i}}^{n} P(\boldsymbol{\theta}|\boldsymbol{d}) d\boldsymbol{\theta}} \qquad \forall \theta_{\mathrm{i}} \in \boldsymbol{\theta}\,, \tag{1.104}$$

where $\mathbb{E}(\theta_{\mathrm{i}}^{n})$ is the $\mathrm{n^{th}}$ moment of a particular parameter θ_{i} in the set of cosmological parameters $\boldsymbol{\theta}$ distributed as $P(\boldsymbol{\theta}|\boldsymbol{d})$. The most sought-after moments are the first ($n = 1$) and second moments ($n = 2$) since they relate to the mean, $\langle \theta_i \rangle = \mathbb{E}(\theta_{\mathrm{i}}^{1})$, and

---

[10] This is the same logic behind Occam's razor or Leibniz's sufficient reason arguments.

the variance, $\sigma^2(\theta_i) = \mathbb{E}(\theta_\mathrm{i}^2) - \mathbb{E}(\theta_\mathrm{i}^1)^2$, of the distribution[11]. This is because, if $P(\boldsymbol{\theta}|\boldsymbol{d})$ is Gaussian, then the mean and the variance contain all the information about the distribution.[63].

Ideally, one would have a close form for the posterior such that Eq. 1.104 can be computed analytically. However, we started under the premise that we don't know $P(\boldsymbol{\theta}|\boldsymbol{d})$. As we will see, this idea will make a comeback in Sect. 1.2.3. Until then, we are bound to numerical methods. Naively, we may try to map $P(\boldsymbol{\theta}|\boldsymbol{d})$ using sufficiently fine grids such that the posteriors can be fully characterised. However, given the average cost of evaluating cosmological models this approach becomes computationally unfeasible once the number of dimensions surpasses three. For example, given a likelihood that costs one second to evaluate and a hundred nodes grid, mapping a one-dimensional posterior takes a minute an a half, doing so for two dimensions takes nearly three hours and raising the number to three dimensions costs just short of two weeks.

Computing moments on a grid is not only very costly but also extremely inefficient. In order to understand why we have to return to Eq. 1.104 and note that only a relatively small fraction of the domain of $\boldsymbol{\theta}$, the one in which $P(\boldsymbol{\theta}|\boldsymbol{d})$ is distinctively non-zero, contributes most to the integrating volume $d\boldsymbol{\theta}$ [66]. This domain of interest is known as the typical set. However, finding the typical set of $P(\boldsymbol{\theta}|\boldsymbol{d})$ if $P(\boldsymbol{\theta}|\boldsymbol{d})$ itself is unknown to us can be very challenging. Thankfully, there is a literature full of algorithms specifically designed for this purpose. This is not surprising given how ubiquitous statistical inference is as a problem. The algorithms designed to find the typical set in an inference problem are known as inference algorithms or samplers since they output samples of the typical set.

Most of the inference algorithms currently used in cosmology map $P(\boldsymbol{\theta}|\boldsymbol{d})$ through a variety of stochastic processes that diffuse towards the typical set. In a Bayesian setting, the starting point of the process is often sampled from the prior. However, it is not uncommon to run an optimisation algorithm such that the process starts from the maximum of the posterior, known as the maximum a posteriori (MAP) point. This usually results in efficiency gains since the MAP often is already in the typical set[12]. The properties of the process that follows are dictated by a transition kernel that defines the probability of the next sample. Often, these processes are chosen to be Markovian and thus the transition kernel only depends on the latest sample such

---

[11] Note that in the case of multivariate distributions the variance is promoted to the covariance matrix whose entries are given by $\mathrm{Cov}_{ij} - = \mathbb{E}(\theta_\mathrm{i}\theta_\mathrm{j}) - \mathbb{E}(\theta_\mathrm{i}^1)\mathbb{E}(\theta_\mathrm{j}^1)$.

[12] Note that the MAP can be outside the typical set for multimodal or sharply peaked distributions.

that the transition kernel is given by $Q(\boldsymbol{\theta}_{i+1}|\boldsymbol{\theta}_i)$. The result is a chain of samples of the typical set. Expectation values can then be computed using straight forward Monte Carlo integration [67]. Hence these methods are known as Monte Carlo Markov Chains (MCMCs).

MCMCs are extremely popular due to their asymptotic properties. Given enough integration time, a stochastic process is guaranteed to visit every point of the distribution. In other words, it is ergodic. Thus, for a given moment of the posterior distribution, the ensemble average (what we are after) is the same as the time average over the MCMC samples:

$$\boxed{\mathbb{E}(\theta_{\mathrm{i}}^{n}) \equiv \int \theta_{\mathrm{i}}^{n} P(\boldsymbol{\theta}|\boldsymbol{d}) d\boldsymbol{\theta} = \lim_{T\to\infty} \frac{1}{T} \int \boldsymbol{\theta}(t) dt} \,. \tag{1.105}$$

Moreover, by looking at the trace (i.e. the distribution of samples over time) it is possible to observe whether or not the algorithm has found the typical set. For example one would expect the mean of the variance, $\bar{\sigma}^2(\boldsymbol{\theta})$, to be constant across integration time. Similarly, the variance of the mean, $\sigma^2(\bar{\boldsymbol{\theta}})$, should be zero hence why the typical set is also commonly referred to as the *stationary* distribution. Formally, the posterior distribution is said to be stationary if it satisfies *detailed balance* [68]:

$$\boxed{P(\boldsymbol{\theta}|\mathbf{d})Q(\boldsymbol{\theta}|\boldsymbol{\theta}') = P(\boldsymbol{\theta}'|\mathbf{d})Q(\boldsymbol{\theta}'|\boldsymbol{\theta})} \,, \tag{1.106}$$

meaning that there's the same probability of the chain transitioning from $\boldsymbol{\theta}$ to $\boldsymbol{\theta}'$ as from $\boldsymbol{\theta}'$ to $\boldsymbol{\theta}$. Therefore, the MCMC of a stationary distribution is said to be *reversible.*

The most common metric to assess whether the process has spent enough time in the stationary distribution is the Gelman-Rubin statistic[13] [69]. The Gelman-Rubin statistic is based around the fact that for a sufficiently long chain one expects:

$$\lim_{N\to\infty} \sigma^2(\bar{\theta}_i) = \lim_{N\to\infty} \left(\frac{1}{N}\bar{\sigma}^2(\theta_i)\right) = 0 \,. \tag{1.107}$$

The Gelman-Rubin statistic is then given by:

$$\boxed{R_i = \sqrt{1 + \frac{\sigma^2(\bar{\theta}_i) - \frac{1}{N}\bar{\sigma}^2(\theta_i)}{\bar{\sigma}^2(\theta_i)}}} \,. \tag{1.108}$$

If the typical set is sufficiently well explored, one would expect $R_i \approx 1$ for all parameters of the chain. Thus, it is common to stop the inference process once

$$R_i - 1 \leq 0.01 \qquad \forall \theta_i \in \boldsymbol{\theta} \,. \tag{1.109}$$

---

[13] Florian Beutler has a nice practical tutorial on how to compute R.

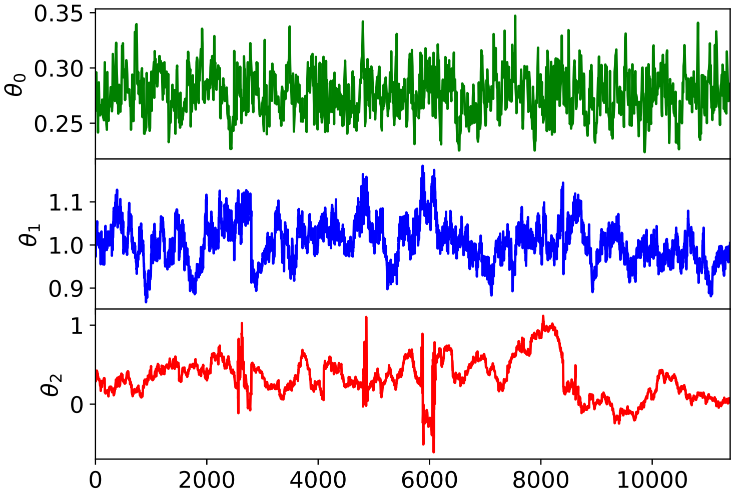

Figure 1.4: Depiction of the concentration of measure problem. In high-dimensions, the volume term, $d\boldsymbol{\theta}$, dominates over the probability density, $P(\boldsymbol{\theta})$, dragging the typical set away from $P(\boldsymbol{\theta})$'s mean.

Once this criterion is met, the chain is said to have *converged* to the stationary distribution. In Fig. 1.4 we show the traces of three different parameters in a MCMC. By visually observing the three different stochastic processes and noting the variability of their mean and variance it is possible to see that that θ_0 is more converged than θ_1 which is more converged than θ_2. This end condition is one of the most attractive properties of MCMC methods.

The most popular of these MCMC methods is the Metropolis Hastings (MH) algorithm [70]. In a MH algorithm samples from the transition kernel, known as candidates, are not directly accepted into the MCMC. Instead, candidates have a probability of being accepted into the MCMC given by:

$$\alpha = \min\left(1, \frac{\mathrm{Q}(\boldsymbol{\theta}_{\mathrm{i}}|\boldsymbol{\theta}_{\mathrm{i+1}})\mathrm{P}(\boldsymbol{\theta_{i+1}}|\mathrm{d})}{\mathrm{Q}(\boldsymbol{\theta}_{\mathrm{i+1}}|\boldsymbol{\theta}_{\mathrm{i}})\mathrm{P}(\boldsymbol{\theta_{i}}|\mathrm{d})}\right) . \tag{1.110}$$

When the transition kernel is symmetric Eq. 1.110 reduces to:

$$\boxed{\alpha = \min\left(1, \frac{\mathrm{P}(\boldsymbol{\theta_{i+1}}|\mathrm{d})}{\mathrm{P}(\boldsymbol{\theta_{i}}|\mathrm{d})}\right)} . \tag{1.111}$$

This acceptance process is known as the Metropolis adjustment and is responsible for two critical properties of MH. First, if the next sample has higher probability it will always be accepted, meaning that the random walk process will always eventually dissipate towards the typical set. Second, the Metropolis adjustment guarantees that the samples that make it into the chain are distributed as $P(\boldsymbol{\theta}|\boldsymbol{d})$. Thus the MH algorithm will always converge to the true posterior distribution as opposed to some biased stationary distribution. Putting these notions together we can explicitly write the MH algorithm as Alg. 1 where $\boldsymbol{\theta}_0$ is the starting point, ϵ is the step size and N is the number of steps.

Algorithm 1 Metropolis Hastings

```
Require: $\boldsymbol{\theta}_0$, $\epsilon$, $N$
  for i=1 to N do
      Set $\boldsymbol{\theta}_i \leftarrow \boldsymbol{\theta}_{i-1}$
      Sample $\tilde{\boldsymbol{\theta}} \sim \mathcal{N}(\boldsymbol{\theta}_i, \epsilon\mathbb{I})$                ▷ Propose sample
      With probability $\alpha = \min(1, \frac{\mathrm{P}(\tilde{\boldsymbol{\theta}})}{\mathrm{P}(\boldsymbol{\theta}_{\mathrm{i-1}})})$, set $\boldsymbol{\theta}_i \leftarrow \tilde{\boldsymbol{\theta}}$    ▷ Metropolis adjustment
  end for
```

It is easy to see then that the success of the MH algorithm depends on how likely the proposal distribution is to generate an accepted sample (i.e. a sample of the typical set). Far from being a problem specific to MH, choosing the particular form of the proposal distribution is the fundamental problem of all MCMC methods. Most approaches involve assuming that samples from the typical set live close to one another and thus they assume a Gaussian transition kernel of the form:

$$\boxed{Q(\boldsymbol{\theta}_{i+1}|\boldsymbol{\theta}_i) = \mathcal{N}(\boldsymbol{\theta}_i, \epsilon\mathbb{I})}\,, \tag{1.112}$$

where ϵ defines the standard deviation of the proposal distribution, acting as the step size of the process.

In the absence of multi-modality, one should indeed expect samples of the typical set to live close by. However, it is crucial to quantify how far and, specially, in what direction. Too large of a step size will result in the algorithm proposing too many samples outside of the typical set which will be rejected. Simultaneously, a step size too small will lead to samples with a high degree of auto-correlation. While this correlation doesn't bias our estimates of the posterior, it nonetheless means that we require more samples to obtain a representative picture. Therefore, in MCMC methods, it is common to talk about the effective sample size (ESS) meaning the number of uncorrelated samples accepted into the MCMC [71]. The ESS of a MCMC can be estimated[14] as:

$$\boxed{\mathrm{ESS} \equiv \frac{\mathrm{N}}{1 + 2\sum_{\mathrm{t}=0}^{\infty} \rho_{\mathrm{t}}}}\,, \tag{1.113}$$

---

[14] Note that the actual ESS cannot be computed from a finite chain, thus Eq. 1.113 is an approximate estimate [72].

where N is the number of samples in the chain and ρ_t is the auto-correlation of the stochastic process at given lag t:

$$\rho_t = \frac{1}{\sigma^2} \int \boldsymbol{\theta}^{(n)} \boldsymbol{\theta}^{(n+t)} d\boldsymbol{\theta} \,, \tag{1.114}$$

where σ^2 is the variance (or covariance matrix) of the samples [72]. Thus, adequately tuning the value of ϵ is vital for the efficiency of the MH algorithm. Moreover, it is easy to see that Eq. 1.112 will struggle for non-isotropic typical sets where, at a given point, there's a preferred direction along which one would ideally propose samples.

Despite the success of this methodology [see 13, 51, 60, 61, 73, among others.], MH becomes very inefficient at high dimensions. In order to understand why, we have to revise our intuition of what happens to distributions as we increase the number of dimensions. Most cosmologists have an intuition that *things become nice* as we increase the number of dimensions due to the central limit theorem. This is true to some extent. The central limit theorem makes it such that the mean of the posterior distribution $P(\boldsymbol{\theta}|\boldsymbol{d})$ becomes Gaussianly distributed as the size of $\boldsymbol{\theta}$ increases. This is why marginalised posterior distributions become Gaussian even if $P(\boldsymbol{\theta}|\boldsymbol{d})$ is not Gaussian itself.

However, the central limit theorem says nothing about what $P(\boldsymbol{\theta}|\boldsymbol{d})$ looks like and helps us very little finding the typical set. To understand why things become *not so nice* when we increase the number of dimensions, we must note that the typical set is located in the regions of high probability **mass**, not probability **density**. The probability mass is given by the zeroth moment:

$$\mathbb{E}(\boldsymbol{\theta}_{\boldsymbol{i}}^{\boldsymbol{0}}) = \int P(\boldsymbol{\theta}|\boldsymbol{d}) d\boldsymbol{\theta} \,. \tag{1.115}$$

Inside Eq. 1.115 there is two competing forces. For the sake of argument, let us consider a set of parameters for which $P(\boldsymbol{\theta}|\boldsymbol{d})$ is Gaussian such that there is a well defined set of radial coordinates centred at the mean of $P(\boldsymbol{\theta}|\boldsymbol{d})$. In such scenario, $P(\boldsymbol{\theta}|\boldsymbol{d})$ is a cloud of probability centred around its mean. However, $d\boldsymbol{\theta}$ increases as we move further away from the posterior's mean. In order to understand this, it is useful to consider how the volume inside a solid angle increases as the radial coordinates l. In low dimensions, the contribution of $P(\boldsymbol{\theta}|\boldsymbol{d})$ dominates and the typical set is distributed around the distribution's mean. However, in high dimensions, the contribution of $d\boldsymbol{\theta}$ will start to take over, shifting the mode away from the mean, dragging along the typical set with it. Thus, for a high-dimension Gaussian, most of the probability mass lives in an increasingly thin shell around the distribution's mean.

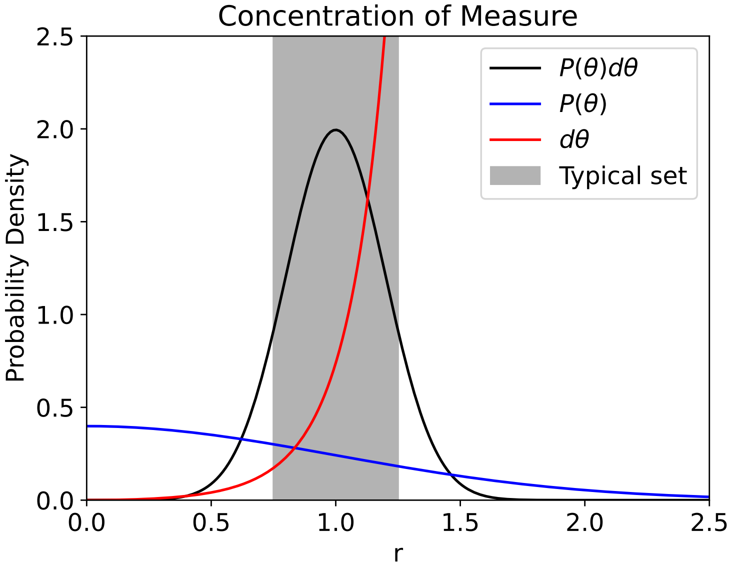

Figure 1.5: Depiction of the concentration of measure problem. In high-dimensions, the volume term, $d\boldsymbol{\theta}$, dominates over the probability density, $P(\boldsymbol{\theta})$, dragging the typical set away from $P(\boldsymbol{\theta})$'s mean.

This phenomenon is known as the concentration of measure [66] and it is depicted in Fig. 1.5.

The concentration of measure has disastrous consequences for the efficiency of the MH algorithm at high dimensions. This is because it directly invalidates the assumptions made when constructing the transition kernel shown in Eq. 1.112. First, since the typical set is now an extremely thin shell, the chances of randomly stepping into it plummet as the dimensionality increases. Second, even if one were to find one sample of the typical set, other points are no longer homogenously distributed around the first as the covariance of Eq. 1.112 would suggest. Instead, the transition kernel would need to follow the probability mass shell, something which is extremely challenging using random guesses. In combination these two effects drastically reduce the acceptance rate of samples into the MCMC. Eventually, the number of likelihood calls (and thus the computing time) required to obtain enough samples representative of the posterior becomes unfeasible. This effect is known as the "curse of dimensionality".

1.2.2 Gradient-Based Inference Algorithms

In the previous section we saw that in order to perform statistical inference in more than a handful of dimensions we rely on MCMC methods. However, MCMC methods live and die by the ability of their transition kernels to track the typical set. The question is then laid as follows: is there a systematic way of constructing $Q(\boldsymbol{\theta}_{i+1}|\boldsymbol{\theta}_i)$ such that it tracks the typical set?

The intuitive idea is to use some information about $P(\boldsymbol{\theta}|\boldsymbol{d})$ (which we will refer to as $P(\boldsymbol{\theta})$ from now on for simplicity) to construct a better transition kernel at every point. Ensemble samplers [74] such as `EMCEE` [75] are based on running multiple MCMCs (referred to as particles) in parallel. The position of the particles furthest away from the typical set are moved towards those closest. However, this approach has a couple of drawbacks. On the one hand, it introduces correlations between the different chains obscuring their interpretation. On the other hand, it assumes that at least some particles find the typical set, which is not guaranteed.

Ensemble samplers fundamentally attempt to track the rate of change of the posterior by using several points of $P(\boldsymbol{\theta})$. Thus one might wonder why not use the gradient of the posterior, $\nabla P(\boldsymbol{\theta})$, directly. We will refer to inference algorithms that use the gradient of posterior to construct their transition kernels, $Q(\boldsymbol{\theta}_{i+1}|\boldsymbol{\theta}_i, \nabla P(\boldsymbol{\theta}_i))$, as gradient-based inference algorithms [76–80]. However, how to specifically use $\nabla P(\boldsymbol{\theta})$ is not a straight forward question. The gradient of the posterior defines a vector field at every $\boldsymbol{\theta}$ that points towards the regions of highest probability **density**, not **mass** which is what we are after. Thus, blindly following $\nabla P(\boldsymbol{\theta})$ will make us descend to the minimum of the posterior, throwing us off the thin shell where the typical set lives. Therefore, in order to efficiently explore the typical set we have to provide $\nabla P(\boldsymbol{\theta})$ with additional structure.

1.2.2.1 Hamiltonian Monte Carlo

Hamiltonian Monte Carlo (HMC) [66, 76] is a MCMC algorithm that explores a parameter space by simulating the dynamics of a Hamiltonian system. This is done by introducing a set of auxiliary momenta variables, $\boldsymbol{p}$, that are independent of the target distribution (i.e. the posterior). The joint distribution of the position (*i.e.* the original parameters) and momentum variables is then defined by a Hamiltonian function which governs the dynamics of the system. The Hamiltonian is typically chosen to be the sum of the potential energy, defined as the negative logarithm of probability density of the likelihood, and the canonical kinetic energy of the momentum variables:

$$\boxed{H(\boldsymbol{\theta}, \boldsymbol{p}) \equiv K(\boldsymbol{p}) + U(\boldsymbol{\theta}) = \frac{1}{2}\mathbf{p}^T \Lambda^{-1} \mathbf{p} - \log P(\boldsymbol{\theta})}\,, \tag{1.116}$$

where $U(\boldsymbol{\theta})$ is the potential energy and Λ is the mass matrix of the momenta variables, i.e. a positive definite matrix which acts as covariance matrix of the momenta variables. The dynamics associated with the Hamiltonian Eq. 1.116 are given by the following Eqs. of motion:

$$\dot{\boldsymbol{\theta}} = \frac{\partial H}{\partial \boldsymbol{p}} = \Lambda^{-1}\boldsymbol{p} \tag{1.117}$$

$$\dot{\boldsymbol{p}} = -\frac{\partial H}{\partial \boldsymbol{\theta}} = \nabla \log P(\boldsymbol{\theta}) \tag{1.118}$$

At each HMC iteration, a new sample is proposed by simulating the dynamics of the system for a fixed number of steps using a numerical integration scheme. Traditionally, Leapfrog integration is used to evolve the Hamiltonian dynamics given its symplectic properties which ensure that the volume of the phase-space is preserved along the trajectory. The acceptance probability of the new sample is then computed using a Metropolis adjustment based on the Hamiltonian energy difference between the current sample and the proposal.

Thus, the full HMC inference algorithm can be written as shown in Alg. 2, where $\boldsymbol{\theta}_0$ is the starting point, ϵ is the step size of the Leapfrog integrator, M is the number of steps in the Hamiltonian trajectory and N is the number of samples. In this algorithm, it is possible to see that the Hamiltonian dynamics act as a form of deterministic transition kernel. However, the overall process remains stochastic since the momentum is re-sampled every M steps. If we were to set $M = 1$, HMC returns to form of MH.

Algorithm 2 Hamiltonian Monte Carlo

```
Require: $\boldsymbol{\theta}_0$, $\epsilon$, $M$, $N$
  for i=1 to N do
      Sample $\boldsymbol{p_0} \sim \mathcal{N}(0, \mathbb{I})$                          ▷ Sample new momentum
      Set $\boldsymbol{\theta}_i \leftarrow \boldsymbol{\theta}_{i-1}, \boldsymbol{p}_i \leftarrow \boldsymbol{p}_{i-1}$                      ▷ Set in case of rejection
      Set $\tilde{\boldsymbol{\theta}} \leftarrow \boldsymbol{\theta}_{i-1}, \tilde{\boldsymbol{p}} \leftarrow \boldsymbol{p}_0$                      ▷ Set variables for recursion
      for m=1 to M do                                        ▷ Hamiltonian trajectory
          Set $\tilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{p}} \leftarrow \mathrm{Leapfrog}(\tilde{\boldsymbol{\theta}}, \tilde{\mathrm{p}}, \epsilon)$
      end for                                                ▷ Metropolis adjustment
      With probability $\min(1, \frac{\exp(-\mathrm{H}(\tilde{\boldsymbol{\theta}}, \tilde{\mathrm{p}})}{\exp(-\mathrm{H}(\boldsymbol{\theta}_{\mathrm{i-1}}, \mathrm{p}_{\mathrm{i-1}}))})$, set $\boldsymbol{\theta}_i \leftarrow \tilde{\boldsymbol{\theta}}, \boldsymbol{p}_i \leftarrow \tilde{\boldsymbol{p}}$
  end for

  function Leapfrog($\boldsymbol{\theta}, \boldsymbol{p}, \epsilon$)
  Set $\tilde{\boldsymbol{p}} \leftarrow \boldsymbol{p} + \frac{\epsilon}{2}\Lambda\nabla U(\boldsymbol{\theta})$                          ▷ Half step in momentum
  Set $\tilde{\boldsymbol{\theta}} \leftarrow \boldsymbol{\theta} + \epsilon\Lambda^{-1}\tilde{\boldsymbol{p}}$                                  ▷ Step in space
  Set $\tilde{\boldsymbol{p}} \leftarrow \boldsymbol{p} + \frac{\epsilon}{2}\Lambda\nabla U(\tilde{\boldsymbol{\theta}})$                          ▷ Half step in momentum
  retrun $\tilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{p}}$
```

Finally, the so much sought-after posterior can be returned by marginalising the aforementioned Hamiltonian distribution over the momenta variables:

$$\boxed{P(\boldsymbol{\theta}) = \frac{1}{\mathcal{Z}} \int \exp(-H(\boldsymbol{\theta}, \boldsymbol{p}))\, d\boldsymbol{p}}\,, \tag{1.119}$$

where $\mathcal{Z}$ is the canonical partition function given by:

$$\mathcal{Z} = \int \exp\left(\frac{-1}{2}\mathbf{p}^T \Lambda^{-1} \mathbf{p}\right) d\boldsymbol{p}\,. \tag{1.120}$$

1.2.2.2 The No U-Turns Sampler

One of the challenges of HMC is choosing the Leapfrog step size and number of integration steps in Alg. 2, which can have a significant impact on the performance of the algorithm. On the one hand, a large step size can lead to the Hamiltonian energy not being conserved along the Hamiltonian trajectory when moving through regions where the gradient of the posterior is large. Conservation errors above a certain threshold will cause the sample to be automatically rejected, known as a divergence. A large number of divergences can lead the sampler over-rejecting samples in these regions leading to biased estimates of the posterior. On the other hand, a small step size will result in an inefficient exploration of most regions of the parameter space where the gradient of the posterior is small. The role of the number of integration steps is slightly more subtle. Choosing a small number of integration steps reduces the distance between the current sample and the proposed sample which increases the degree of auto-correlation between the samples in the MCMC. Choosing a large number of steps not only increases the number of likelihood calls (and hence the computing cost) but it also does not ensure that the new sample will be uncorrelated. This counter-intuitive phenomenon follows from the fact that all Hamiltonian trajectories eventually return to already explored parts of their phase space. Thus, HMC trajectories must be run for just the right number of steps. This raises the question, how can such number of steps be known?

To address these challenges, the No U-Turn Sampler [NUTS, 77] algorithm was proposed as an extension of HMC. NUTS introduces a recursive algorithm that determines the optimal number of trajectory steps for each step in the MCMC. The algorithm generates a *tree* of proposals by evolving the Hamiltonian dynamics forwards and backwards in time until one of the branches starts to turn on itself. At each step of the branching processes the proposed samples undergo a MH adjustment to be accepted to the tree. Once the turn occurs, the latest sample in the tree to

Algorithm 3 No U-Turns Sampler

```
Require: $\boldsymbol{\theta}_0$, $\epsilon$, $N$
  for i=1 to N do
      Sample $\boldsymbol{p_0} \sim \mathcal{N}(0, \mathbb{I})$                                          ▷ Sample momentum
      Set $\boldsymbol{\theta}_i \leftarrow \boldsymbol{\theta}_{i-1}\,, \boldsymbol{p}_i \leftarrow \boldsymbol{p}_{i-1}$                          ▷ Set in case of rejection
      Set $\boldsymbol{\theta}_- \leftarrow \boldsymbol{\theta}_{i-1}\,, \boldsymbol{\theta}_+ \leftarrow \boldsymbol{\theta}_{i-1}\,, \boldsymbol{p}_- \leftarrow \boldsymbol{p}_0\,, \boldsymbol{p}_+ \leftarrow \boldsymbol{p}_0$    ▷ Set variables for recursion
      Set $j \leftarrow 0\,, s \leftarrow 1$                                              ▷ Counter and *while* condition
      while s=1 do                                                     ▷ Branching
          $v \sim B(\{-1, 1\}, 0.5)$                                         ▷ Choose a direction
          $\boldsymbol{\theta}_-, \boldsymbol{p}_-, \boldsymbol{\theta}_+, \boldsymbol{p}_+ \leftarrow \mathrm{BuildTree}(\boldsymbol{\theta}_-, \mathrm{p}_-, \mathrm{v}, \mathrm{j}, \epsilon)$
          $s \leftarrow \tilde{s}\,\mathbb{I}\{(\boldsymbol{\theta}_+ - \boldsymbol{\theta}_-)\boldsymbol{p}_- \geq 0\}\mathbb{I}\{(\boldsymbol{\theta}_+ - \boldsymbol{\theta}_-)\boldsymbol{p}_+ \geq 0\}$            ▷ No U-turn
          $j \leftarrow j + 1$
      end while
  end for

  function BuildTree($\boldsymbol{\theta}, \boldsymbol{p}, v, j, \epsilon$)
  Recursion - Build left and right tree.
  $\boldsymbol{\theta}_-, \boldsymbol{p}_-, \boldsymbol{\theta}_+, \boldsymbol{p}_+ \leftarrow \boldsymbol{\theta}, \boldsymbol{p}, \boldsymbol{\theta}, \boldsymbol{p}$
  if $v = 1$ then
      $\boldsymbol{\theta}_-, \boldsymbol{p}_-, -, - \leftarrow \mathrm{Leapfrog}(\boldsymbol{\theta}_-, \mathrm{p}_-, \mathrm{v} \cdot \epsilon)$                          ▷ Backwards
      Set $\tilde{\boldsymbol{\theta}} \leftarrow \boldsymbol{\theta}_-\,, \tilde{\boldsymbol{p}} \leftarrow \boldsymbol{p}_-$
  else
      $-, -, \boldsymbol{\theta}_+, \boldsymbol{p}_+ \leftarrow \mathrm{Leapfrog}(\boldsymbol{\theta}_+, \mathrm{p}_+, \mathrm{v} \cdot \epsilon)$                          ▷ Forwards
      Set $\tilde{\boldsymbol{\theta}} \leftarrow \boldsymbol{\theta}_+\,, \tilde{\boldsymbol{p}} \leftarrow \boldsymbol{p}_+$
  end if                                                          ▷ Metropolis adjustment
  With probability $\min(1, \frac{\exp(-\mathrm{H}(\tilde{\boldsymbol{\theta}}, \tilde{\mathrm{p}})}{\exp(-\mathrm{H}(\boldsymbol{\theta}_{\mathrm{i-1}}, \mathrm{p}_{\mathrm{i-1}}))})$, set $\boldsymbol{\theta}_i \leftarrow \tilde{\boldsymbol{\theta}}\,, \boldsymbol{p}_i \leftarrow \tilde{\boldsymbol{p}}$
  return $\boldsymbol{\theta}_-, \boldsymbol{p}_-, \boldsymbol{\theta}_+, \boldsymbol{p}_+$

  function Leapfrog($\boldsymbol{\theta}, \boldsymbol{p}, \epsilon$)
  Set $\tilde{\boldsymbol{p}} \leftarrow \boldsymbol{p} + \frac{\epsilon}{2}\Lambda\nabla U(\boldsymbol{\theta})$                                    ▷ Half step in momentum
  Set $\tilde{\boldsymbol{\theta}} \leftarrow \boldsymbol{\theta} + \epsilon\Lambda^{-1}\tilde{\boldsymbol{p}}$                                           ▷ Step in space
  Set $\tilde{\boldsymbol{p}} \leftarrow \boldsymbol{p} + \frac{\epsilon}{2}\Lambda\nabla U(\tilde{\boldsymbol{\theta}})$                                    ▷ Half step in momentum
  return $\tilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{p}}$
```

pass the MH adjustment is accepted into the MCMC. Thus we can write the NUTS algorithm by appending this logic to the HMC algorithm as Alg. 3.

NUTS is normally paired with a series of prescriptions to tune the step size of the Leapfrog scheme. A good heuristic to avoid divergences is that ϵ should be at least twice as small as the smallest eigenvalue of the posterior covariance matrix. Since we rarely have access to such information prior to the inference, practitioners have had to come up with dynamical tuning routines that tweak ϵ along the process to avoid divergences. These tuning routines can be fundamentally classified in two groups depending on whether they target the error in the conservation of energy along the Hamiltonian trajectory or the acceptance rate of the MH adjustment. The first type of strategies establish a target energy conservation error and then update the value of step size at each trajectory step based on the difference between the target and the obtained energy conservation error. This updating is usually done using a dual averaging scheme. The second set of schemes establish a target acceptance probability (TAP) of the samples during the MH adjustment. Thus, if the acceptance rate is too low, the step size is lowered and vice-versa. The second strategy is the most popular of the two since it can be shown that for a Gaussian posterior there is an optimal acceptance rate equal to 65% which can be used as well-motivated initial guess for the TAP.

Finally, one can imagine that setting ϵ to be in the same order of magnitude as the smallest eigenvector of the covariance matrix of the posterior will significantly hamper performance if all the other dimensions have a much greater variance. HMC algorithms address this problem by tweaking the mass matrix, Λ, of the momenta variables. One can think of Λ as a transformation from the original parameter space to a secondary space where the variance of all the parameters is of order one. Thus the optimal mass matrix is given by:

$$\Lambda = \mathbb{E}(\boldsymbol{\theta}\boldsymbol{\theta}')^{-1/2}\,. \tag{1.121}$$

Of course, we don't have information on the variance of the posterior ahead of the inference. Thus the variance is periodically estimated from the samples accepted in the MCMC so far. One can imagine however that if we first tune ϵ and then adapt Λ the value of ϵ found will no longer be adequate. Thus it is common to tune Λ and ϵ iteratively until:

$$\mathbb{E}(\Lambda\boldsymbol{\theta}\Lambda\boldsymbol{\theta}') \approx \mathbb{I}\,. \tag{1.122}$$

1.2.3 The Laplace Approximation

1.2.3.1 Formalism

So far we have focused on how to use the gradient of the posterior to construct proposal distributions for high-dimensional MCMC methods. In this section, we will show how gradients can be used to reduce the dimensionality of the problem so that traditional MH becomes viable again.

One can imagine that the premise of running an inference process implies that for some parameters in our model we expect a rather *nasty* (i.e. non-Gaussian, in surprising ways) distribution that we hope to learn. However, we also expect to have parameters in our model which are there to only account for our own lack of knowledge about some menial aspect that we ultimately don't care about. In Cosmology, the former type of parameters are the cosmological parameters that describe the properties of the Universe. In addition to this, in recent years, the increase in the quality of data has fuelled the appearance of a large number of parameters of the latter kind. As discussed in Sect. 1.1.5, future cosmological analyses will have to consider progressively larger numbers of nuisance parameters whose sole purpose is to account for a plethora of theoretical and observational uncertainties that we can no longer ignore. Moreover, these parameters often remain unconstrained by the data, returning to their assigned prior distribution. Thus, nuisance parameters often act as a Bayesian propagation of errors of sorts.

Formally, the spaces of parameters can be decomposed as $\vec{\theta} = \{\boldsymbol{\Omega}, \boldsymbol{n}\}$, where $\boldsymbol{\Omega}$ is a set of cosmological parameters that we care about, and $\boldsymbol{n}$ is a vector of nuisance parameters, which are largely irrelevant to the fundamental question being explored. Thus, hypothetically we would only need to compute the contribution of these parameters to the expectation values of the cosmological parameters without computing their own expectation values. In other words, we only care about the marginalised posterior:

$$P(\boldsymbol{\Omega}) = \int P(\boldsymbol{\Omega}, \boldsymbol{n})\, d\boldsymbol{n}\,. \tag{1.123}$$

The Laplace approximation [4, 5, 81] is a gradient-based analytical marginalisation scheme based on building a Gaussian approximation to the posterior distribution such that the integral in Eq. 1.123 can be performed analytically. This is achieved by Taylor-expanding the log-posterior distribution around the best-fit values of the dimensions we wish to Gaussianise. By virtue of expanding around the best-fit values, the linear term of the expansion vanishes by definition, leaving only the zeroth term and the quadratic term. Since the zeroth term does not depend on n, the dependence of the log-posterior on n becomes quadratic, leading to a Gaussian form for the posterior.

In order to understand how this is done in practice, let us start by considering the best-fit value of the nuisance parameters having fixed $\boldsymbol{\Omega}$. That is, we define $\boldsymbol{n}_*(\boldsymbol{\Omega})$ as

$$\boldsymbol{n}_*(\boldsymbol{\Omega}) \equiv \arg\,\max_{\boldsymbol{n}} P(\boldsymbol{\Omega}, \boldsymbol{n})\,. \tag{1.124}$$

Assuming that the distribution is differentiable at all points, $\boldsymbol{n}_*$ then satisfies

$$\left.\frac{\partial \chi^2}{\partial \boldsymbol{n}}\right|_{\boldsymbol{n}_*} = 0, \tag{1.125}$$

where we have defined $\chi^2 = -\log P(\boldsymbol{\theta})$. Following [82], we can then approximate the distribution at each value of $\boldsymbol{\Omega}$ by expanding χ^2 to second order in $\boldsymbol{n}$ around $\boldsymbol{n}_*$, i.e.:

$$\boxed{\chi^2(\boldsymbol{\Omega}, \boldsymbol{n}) \simeq \chi^2_*(\boldsymbol{\Omega}) + \Delta\boldsymbol{n}^T \mathcal{F}_* \Delta\boldsymbol{n}}\,, \tag{1.126}$$

where $\chi^2_*(\boldsymbol{\Omega}) \equiv \chi^2(\boldsymbol{\Omega}, \boldsymbol{n}_*)$, $\Delta\boldsymbol{n} \equiv \boldsymbol{n} - \boldsymbol{n}_*$, and we have defined the matrix

$$\mathcal{F}_{*,ij} = \left.\frac{1}{2}\frac{\partial^2 \chi^2}{\partial n_i \partial n_j}\right|_{\boldsymbol{n}_*}\,. \tag{1.127}$$

In this limit, the distribution is locally (i.e. at each $\boldsymbol{\Omega}$) a multivariate normal distribution in $\boldsymbol{n}$, and thus the integral in Eq. (1.123) can be solved analytically. The resulting marginalised likelihood has a $\chi^2_m(\boldsymbol{\Omega}) \equiv -2\log p(\boldsymbol{\Omega}|\mathbf{d})$ given by

$$\boxed{\chi^2_m(\boldsymbol{\Omega}) \simeq \chi^2_*(\boldsymbol{\Omega}) + \log\left\{\det\left[\mathcal{F}_*(\boldsymbol{\Omega})\right]\right\} + \text{const.}}\,. \tag{1.128}$$

In what follows, we will label the two contributions in Eq. (1.128), χ^2_* and $\log\det\mathcal{F}_*$, as the *profile* and *Laplace* terms respectively:

1. The *profile* term is related to the *profile likelihood* [83], defined as

   $$P_{\rm prof}(\boldsymbol{\Omega}) \propto P(\boldsymbol{\Omega}, \boldsymbol{n}_*). \tag{1.129}$$

   The profile likelihood is a tool commonly used in frequentist parameter inference [84]. The advantage of the profile likelihood is that its maximum is, by definition, the global maximum of the joint distribution. Understanding this maximum as an estimator for $\boldsymbol{\Omega}$ given the data, constraints on $\boldsymbol{\Omega}$ can be obtained by calculating this maximum for random simulated realisations of the data. Additionally, the posterior profile likelihood is, by construction, centred on the best-fit parameters, and is thus free from volume effects associated with the choice of the nuisance parameters [19, 85, 86].

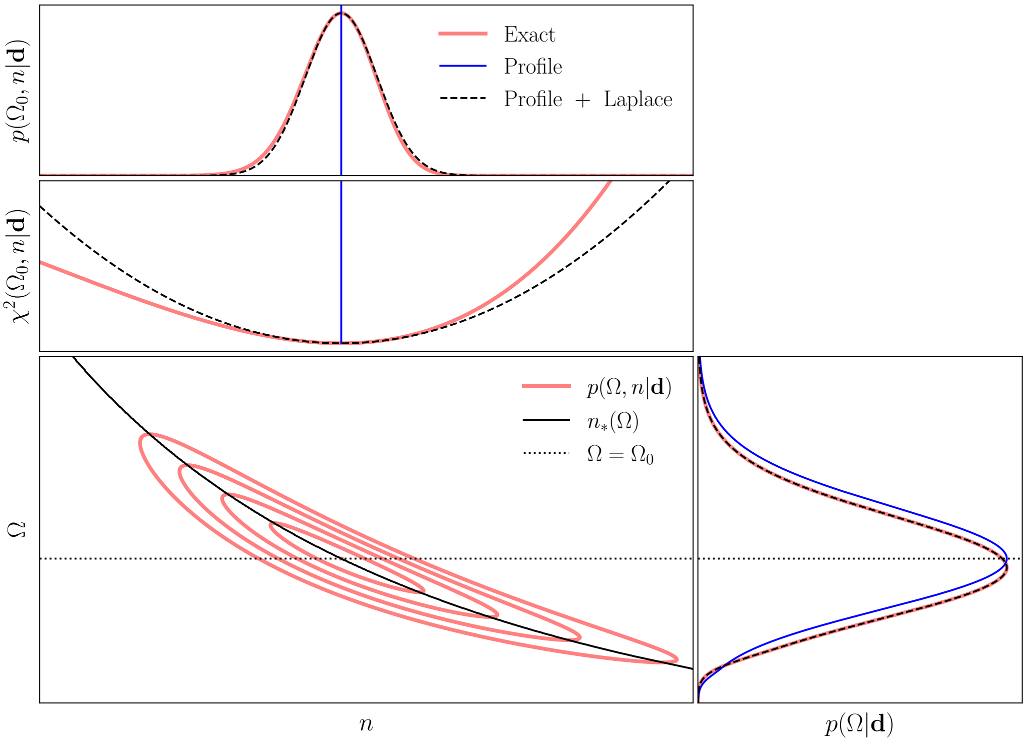

Figure 1.6: Joint posterior distribution on two parameters, Ω and n, with an approximate degeneracy of the form $n\Omega^{1.2} \sim$ const. The large bottom left panel shows the joint distribution as red contours, with the position of the best-fit value of n as a function of Ω in solid black. The central panel shows the χ^2 for a fixed value of $\Omega = \Omega_0$ (shown as the dotted line in the first panel). The red line shows the true χ^2, while the dashed black line shows the Laplace approximation. The top panel shows the probability distribution along $\Omega = \Omega_0$ (given by $p \propto \exp(-\chi^2/2)$), with the exact distribution and its Laplace approximation in red and dashed black respectively. The bottom right panel shows the distribution of Ω marginalised over n. The exact result is shown in red, with its Laplace approximation shown in dashed black. The blue line shows the marginalised profile likelihood obtained by simply maximising the joint likelihood over n for each Ω. The Laplace approximation provides an excellent description of the marginalised distribution, while the profile likelihood returns a distribution with very similar width but centred, by construction, on the best-fit value of Ω, avoiding volume effects.

2. The *Laplace* term, sometimes referred to as *Occam's razor* term, is associated with the quadratic contribution to the Laplace approximation (Eq. (1.128)), and accounts, to first order, for the volume in the space of nuisance parameters $\boldsymbol{n}$ that has been integrated over for fixed $\boldsymbol{\Omega}$ (i.e. the local curvature of the joint distribution at each $\boldsymbol{\Omega}$)[15]. The Laplace term is associated with volume effects, and is subdominant with respect to the profile term for sufficiently constraining data.

The role of the *profile* and *Laplace* terms is illustrated in Fig. 1.6. The figure shows a bivariate distribution for two parameters with an approximate degeneracy of the form $n\Omega^{1.2} \sim$ const. The contour levels of the true distribution are shown in red in the bottom left panel, while the black solid line shows the best-fit value of n as a function of Ω. The middle panel shows the exact χ^2 of the distribution as a function of n for a fixed $\Omega = \Omega_0$ (for convenience, we chose Ω_0 to be the maximum of the distribution, also shown as a dotted line in the bottom panel). The black dashed line shows the quadratic Laplace approximation to the red curve, with the position of the best-fit n (for $\Omega = \Omega_0$) marked by the blue line. The top panel shows the distribution along the $\Omega = \Omega_0$ line. The exact distribution is again shown in red, and the Laplace approximation to it is shown in dashed black. The "profile likelihood" approximation, which fixes n to its best-fit value, is shown in blue. Finally, the bottom right panel shows the distribution marginalised over n, $p(\Omega)$. The true marginal is shown in red. The result of analytically marginalising over n using the Laplace approximation is shown in dashed black, and recovers the true marginal almost exactly. Finally, the conditionally maximised distribution accounting only for the *profile* term in Eq. (1.128) is shown in blue. As mentioned above, the profile-only approximation recovers a distribution that is centred at the best-fit value of Ω (marked by the dotted line).

Two qualitative results should be borne in mind in what follows. First, the Laplace approximation provides a reasonably accurate prediction for the marginal for sufficiently well-behaved distributions. Secondly, keeping only the *profile* term, $P(\boldsymbol{\Omega}, \boldsymbol{n}_*)$, recovers a distribution that has approximately the same width but is, by construction, centred on the maximum of the full (un-marginalised) distribution, $P(\boldsymbol{\Omega}, \boldsymbol{n})$.

It is worth noting that including the Laplace term in Eq. (1.128) should come at virtually no additional computational cost. Finding $\boldsymbol{n}_*(\boldsymbol{\Omega})$ requires solving for $\partial_{\boldsymbol{n}}\chi^2 = 0$, which can be done efficiently using gradient descent methods. Finding

---

[15] In the frequentist context, [87] introduced the formula in Eq. (1.128) under the name of "modified profile likelihood".

the optimal step size in these algorithms often requires evaluating the Hessian of the function being minimised, and therefore the matrix $\mathcal{F}_*$ entering the Laplace term, is already a product of the minimisation algorithm. For instance, the iteration in the case of the Newton-Raphson algorithm is given by

$$\boldsymbol{n}_{*,i+1} = \boldsymbol{n}_{*,i} - \left([\nabla_n \nabla_n^T \chi^2]^{-1} \cdot \nabla_n \chi^2\right)_i, \tag{1.130}$$

where $\nabla_n \chi^2$ is the gradient of χ^2 with respect to $\boldsymbol{n}$, and $\nabla_n \nabla_n^T \chi^2 \equiv 2\mathcal{F}$ is its Hessian matrix. In the applications we will explore here, when Eq. (1.125) cannot be solved analytically, we will make use of a modified version of the Newton-Raphson algorithm [88].

1.2.3.2 Gaussian likelihoods

Let us now apply the method described in the previous section to the case of Gaussian likelihoods. In this case we assume that the posterior distribution takes the form:

$$-2\log p(\boldsymbol{\Omega}, \boldsymbol{n}|\mathbf{d}) = (\mathbf{d} - \mathbf{t})^T \mathsf{C}^{-1}(\mathbf{d} - \mathbf{t}) + \chi^2_{p,\Omega}(\boldsymbol{\Omega}) + \chi^2_{p,n}(\boldsymbol{n}). \tag{1.131}$$

Here, $\mathbf{t}(\boldsymbol{\Omega}, \boldsymbol{n})$ is the theory vector, which depends on the model parameters, C is the covariance matrix of the data, which we assume to be model-independent, and $\chi^2_{p,\Omega}$ and $\chi^2_{p,n}$ are the parameter priors. Although the methodology described below is straightforward to generalise to the case of arbitrary priors, for simplicity we will assume that the nuisance parameters have Gaussian priors, and therefore

$$\chi^2_{p,n}(\boldsymbol{n}) = (\boldsymbol{n} - \boldsymbol{n}_p)^T \mathsf{C}_n^{-1}(\boldsymbol{n} - \boldsymbol{n}_p), \tag{1.132}$$

where C_n is the prior covariance. In the case of non-Gaussian priors, it is often possible to apply a transformation to the nuisance parameters that Gaussianizes (e.g. via normalising flows [89]) without introducing any pathologies (singularities, etc.).

In order to find $\boldsymbol{n}_*$ and $\mathcal{F}$, we need the first and second derivatives of the χ^2 with respect to $\boldsymbol{n}$. In this case, these are given by:

$$\frac{\partial \chi^2}{\partial n_i} = -2\partial_i \mathbf{t}^T \mathsf{C}^{-1}(\mathbf{d} - \mathbf{t}) + 2\sum_j \left[\mathsf{C}_n^{-1}\right]_{ij} (n_j - n_{p,j}), \tag{1.133}$$

$$\mathcal{F}_{ij} = F_{ij} + \Delta\mathcal{F}_{ij}, \tag{1.134}$$

where we have used the shorthand $\partial_i \equiv \partial/\partial n_i$, and we have defined

$$F_{ij} \equiv \partial_i \mathbf{t}^T \mathsf{C}^{-1}\, \partial_j \mathbf{t} + \left[\mathsf{C}_n^{-1}\right]_{ij}, \tag{1.135}$$

$$\Delta\mathcal{F}_{ij} \equiv \partial_i \partial_j \mathbf{t}^T\, \mathsf{C}^{-1}(\mathbf{t} - \mathbf{d}). \tag{1.136}$$

On the one hand, the first contribution to $\mathcal{F}$, F, has three interesting properties:

- It is positive-definite, and therefore invertible;
- It is independent of the data $\mathbf{d}$;
- It coincides with the Fisher matrix of the Gaussian likelihood of Eq. (1.131) with respect to the nuisance parameters.

On the other hand, when evaluated on the hypersurface $\boldsymbol{n} = \boldsymbol{n}_*(\boldsymbol{\Omega})$, $\mathbf{t}$ is close to $\mathbf{d}$, and therefore the contribution from $\Delta\mathcal{F}$ is usually smaller than F.

1.2.3.3 Linear Parameters

Let us now consider the case of linear parameters. Consider a Gaussian likelihood in the form of Eq. (1.131) where all parameters live in the theory prediction, which has the form

$$\mathbf{t} = \mathbf{t}_0 + \mathsf{T}\boldsymbol{n}, \tag{1.137}$$

where $\mathbf{t}_0$ and T are a vector and a matrix independent of $\boldsymbol{n}$, but potentially dependent on $\boldsymbol{\Omega}$. For simplicity, we will assume that the prior on $\boldsymbol{n}$ is centred at zero ($\boldsymbol{n}_p = 0$). This can always be achieved by simply redefining $\boldsymbol{n}' \equiv \boldsymbol{n} - \boldsymbol{n}_p$, and adding the contribution $\mathsf{T}\boldsymbol{n}_p$ to $\mathbf{t}_0$.

The case of linear parameters is particularly interesting, because the χ^2 is quadratic in $\boldsymbol{n}$ by construction, and the Laplace approximation is exact. Since the second derivatives of the theory vector are zero, $\Delta\mathcal{F} = 0$, and the Fisher matrix is independent of $\boldsymbol{n}$ and given by

$$\mathcal{F} = F = \mathsf{T}^T\mathsf{C}^{-1}\mathsf{T} + \mathsf{C}_n^{-1}. \tag{1.138}$$

Furthermore, the best-fit parameters can be found analytically:

$$\boldsymbol{n}_* = F^{-1}\,\mathsf{T}^T\mathsf{C}^{-1}\mathbf{r}, \tag{1.139}$$

where $\mathbf{r} \equiv \mathbf{d} - \mathbf{t}_0$ is the data rescaled by the $\boldsymbol{n}$-independent component of the theory. Using $\boldsymbol{n}_*$ to compute the χ^2, and using Eq. (1.138), we obtain the marginalised χ^2_m of Eq. (1.128) which, as we said, is exact in this case.

The first thing worth noting is that, if the matrix T is independent of $\boldsymbol{\Omega}$, then F is constant, and so is the Laplace term. Up to an irrelevant overall constant, the marginalised χ^2 is then equivalent to χ^2_*, obtained by substituting the best-fit value of $\boldsymbol{n}$. Thus, the approximate relation between marginalisation and maximisation we outlined in the previous section becomes an equivalence when the data is Gaussian with a linear model in $\boldsymbol{n}$ since, in this case, there are no volume effects. All volume effects resulting from a dependence of T on $\boldsymbol{\Omega}$ are otherwise incorporated exactly in the Laplace term. If the priors on $\boldsymbol{n}$ are sufficiently tight, the second term in Eq. (1.138) dominates, and these volume effects become negligible.

Let us now focus on the profile term. Substituting $\boldsymbol{n}_*$ from Eq. (1.139), we obtain

$$\chi^2_* = (\mathsf{W}\mathbf{r})^T\mathsf{C}^{-1}(\mathsf{W}\mathbf{r}) + \mathbf{r}^T\mathsf{C}^{-1}\mathsf{T}F^{-1}\mathsf{C}_n^{-1}F^{-1}\mathsf{T}^T\mathsf{C}^{-1}\mathbf{r}, \tag{1.140}$$

where the second term comes from the prior on $\boldsymbol{n}$, and we have defined the matrix

$$\mathsf{W} \equiv \mathsf{I} - \mathsf{T}F^{-1}\mathsf{T}^T\mathsf{C}^{-1}, \tag{1.141}$$

with I the identity.

First, consider the limit of no external prior (i.e. $\mathsf{C}_n^{-1} = 0$). In this case, we can ignore the second term in Eq. (1.140), and the Fisher matrix is $F = \mathsf{T}^T\mathsf{C}^{-1}\mathsf{T}$. We can then see that the matrix W projects $\mathbf{r}$ onto the subspace that is orthogonal to all the columns of T (with orthogonality defined using the inverse covariance of the data C^{-1} as a dot product). Marginalising over linear parameters is therefore equivalent in this limit to deprojecting all modes of the data that live in the subspace spanned by the columns of T [90]. Secondly, Eq. (1.140) can be simplified significantly into

$$\chi^2_* = \mathbf{r}^T\tilde{\mathsf{C}}^{-1}\mathbf{r}, \tag{1.142}$$

where $\tilde{\mathsf{C}}$ is a modified covariance given by

$$\boxed{\tilde{\mathsf{C}} = \mathsf{C} + \mathsf{T}\mathsf{C}_n\mathsf{T}^T}\,. \tag{1.143}$$

To obtain this beautifully simple result, one only needs to expand first term in Eq. (1.140), simplify the result, and make use of the Woodbury matrix identity [91]. We can then plug this modified covariance matrix into Eq. 1.131 to obtain:

$$\begin{aligned}-2\log P(\boldsymbol{\Omega}|\mathbf{d}) \simeq& (\mathbf{r}-\bar{\mathbf{t}})^T\tilde{\mathsf{C}}^{-1}(\mathbf{d}-\bar{\mathbf{t}}) + \chi^2_{p,\Omega}(\boldsymbol{\Omega}) \\ &+ \log\left[\det\left(\mathsf{T}^T\mathsf{C}^{-1}\mathsf{T}+\mathsf{C}_n^{-1}\right)\right] + \text{const.}.\end{aligned} \tag{1.144}$$

It is worth stressing again that this result is an exact expression for both the marginal posterior and the conditionally maximised posterior. Maximising and marginalising over $\boldsymbol{n}$ therefore result in the same Gaussian likelihood with the theory vector evaluated at $\boldsymbol{n} = 0$ (or at its prior mean if non-zero), and a modified covariance $\tilde{\mathsf{C}}$, obtained by simply assigning additional variance in quadrature to the modes of the data that align with the columns of T (with this extra variance given by the $\boldsymbol{n}$ parameter priors).

To summarise: in the case of Gaussian data, negligible parameter dependence of the covariance matrix, and a theory model that is linear in the nuisance parameters, the Laplace approximation is exact. In this case, there is a mathematical equivalence between marginalisation, χ^2 minimisation, deprojection, and simply adding in quadrature the prior uncertainty on the marginalised parameters at the data level. If the modes associated with $\boldsymbol{n}$ (i.e. the columns of T) depend on the other parameters of the model, the associated volume effects are captured exactly by the Laplace term, which is simply given by the log-determinant of Eq. (1.138). Importantly, while this approach is algorithmically the fastest, it does not produce a best-fit value of a given nuisance parameter.

1.3 Computer Differentiation

1.3.1 Finite Differences

In Chap. 1.2 we showed how the gradients of the posterior distribution will be an invaluable tool to fight off the curse of dimensionality as the number of parameters in cosmological analyses increases. However, one can imagine a situation where the cost of the gradient surpasses the efficiency improvement of the better transition kernel. Thus, careful attention has to be paid to how the gradients are computed. This is particularly true in high dimensions for which the methods we currently use to numerically compute such gradients become costly.

The derivative of a function, $f(x)$, is a second function which tracks the rate of the change of the former [92]. A function is differentiable across its domain iff [93]:

$$\boxed{\frac{df}{dx} = \lim_{\epsilon \to 0} \frac{f(x+\epsilon) - f(x)}{\epsilon}} \qquad \forall x \in \mathbb{R}. \tag{1.145}$$

If the function is multivariate, $f(\boldsymbol{x})$, the object of interest is often the gradient, defined as:

$$\nabla f(\boldsymbol{x}) = \left(\frac{df}{dx_n}, \frac{df}{dx_{n-1}}, ..., \frac{df}{dx_0}\right) \tag{1.146}$$

Besides being the fundamental definition of a derivative, Eq. 1.145 acts as a prescription to compute the derivative of a given function. However, humans rarely use Eq. 1.145 to compute derivatives, at least not explicitly. Instead, we memorise the derivative of common place algebraic expressions which we then compose to compute the derivatives of complicated functions. Thus, given a complicated function, $f(x)$, we can always express it as a composition of simpler functions $W_i(x)$ whose derivative we know, $f(x) \equiv W_n(W_{n-1}(...W_1(W_0)))$ where $W_n \equiv f$, $W_0 \equiv x$ and n is the number of compositions needed. It can then be shown from Eq. 1.145 that the derivative of $f(x)$ can be expressed in terms of known derivatives of the composites as:

$$\boxed{\frac{df}{dx} = \prod_{i=0}^{n-1} \frac{dW_{n-i}}{dW_{n-(1+i)}}}. \tag{1.147}$$

This expression is known as the chain rule.

The story is slightly different for computers. Inside computers, functions are often represented as algorithms, also referred to as programmes. In order to compute the gradient of a programme, computers have traditionally relied on numerically approximating Eq. 1.145. Since computers cannot take infinitesimal limits, this method is known as *finite* differences [94]:

$$\frac{df}{dx} = \frac{f(x+\epsilon) - f(x)}{\epsilon}\,, \tag{1.148}$$

where ϵ is a small but finite quantity which determines the error of the approximation. More accurate schemes that improve on Eq. 1.148 can be devised by comparing to the Taylor expansion of $f(x)$. However, they often require evaluating $f(x)$ at more points. Thus it is common to talk of the m-th point finite differences derivative where m is the number of evaluations of $f(x)$.

The main problem of using finite differences to compute gradients of computer programs is that the algorithm scales poorly with the number of dimensions of the programme being differentiated. This can be seen by the fact that, given a function that takes n inputs, computing the m-th point finite differences gradient takes $n \times m$ calls of the original function. Moreover, the obtained derivative will be subject to truncation errors of order $\mathrm{O}(\epsilon^m)$.

1.3.2 Auto-Differentiation

Ideally, one would compute the gradient of the program analytically and then write a second program for it. The problem with this approach is that writing analytical derivatives for modern computer programs is often too cumbersome and error prone given their complexity [16].

Would it not be nice then if computers could systematically apply the chain rule through their programmes just like their programmers do for analytical expressions? This is exactly what computer scientists have been attempting since the inception of computer science. Quite literally, the first ever PhD thesis in computer science awarded to R.E. Wengert proposed "A simple automatic derivative evaluation program"[96], kick-starting the entire field of Automatic Differentiaton (AD). As we will see, AD would go on to become an essential enabler of the neural network revolution we are currently living through.

Fundamentally, AD is a series of algorithms that allow computers to apply the chain rule to differentiate through computer programs. Just like we humans do, AD achieves this by algorithmically reducing complex computer programs to a series of compound primitive operations (i.e. $+$, $\surd$, sin, exp ...etc) whose derivatives are

---

[16] Note that, while tedious, this approach is not impossible. With heroic efforts, the BORG collaboration managed to write fully analytical gradients for complex N-body simulations [95].

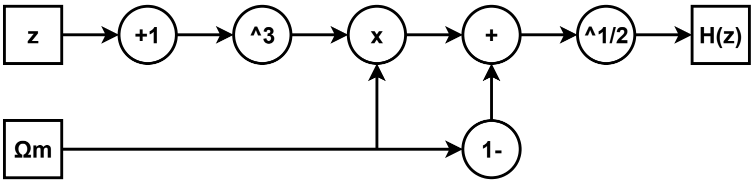

Figure 1.7: Wengert tape of Alg. 4.

known such that the derivative of the whole program can be computed. The reason why AD is hard, but also mesmerising for computer scientists, is because it requires teaching computers some form of *understanding* of the programs they execute. This is akin to how humans intuitively perceive complicated functions as compositions of simpler ones.

AD achieves this understanding by building an schematic representation of the computer program that specifies the relationships between its inputs, the primitive operations and the final output. This schematic representation is commonly refereed to as the Wengert tape, after Wengert [96].

In order to understand this taping process, let us consider a simple toy example of a computer program that will be familiar to most cosmologists: the expansion history of a Universe with matter, a cosmological constant but no radiation

$$H(z) = \sqrt{(1+z)^3\Omega_\mathrm{m} - (1-\Omega_\mathrm{m})}. \tag{1.149}$$

The associated algorithm with Eq. 1.149 is given by Alg. 4.

Algorithm 4 Expansion History

```
Require: $\Omega_\mathrm{m}$, $z$
  return $\sqrt{(1+z)^3\Omega_\mathrm{m} - (1-\Omega_\mathrm{m})}$
```

The key insight is to think of this program as a composition of sums, products and powers whose relationship is given by the Wengert tape shown in Fig. 1.7. Once we know the primitives and their relationships, we can differentiate through them to obtain the derivatives of the program as shown in Tab. 1.1. In Tab. 1.1 $\partial_i W_0$ and $\partial_i W_1$ are known as the seeds and control the variable with respect to which the derivative is taken. Hence, once the tape is recorded, setting $\partial_i W_0 = 1$ and $\partial_i W_1 = 0$ allows AD to compute $\partial_{\Omega_\mathrm{m}} H(z)$. Similarly, the seed $\partial_i W_0 = 0$ and $\partial_i W_1 = 1$ results in $\partial_z H(z)$. This allows AD to easily compute multivariate gradients with one single tape [97].

Table 1.1: Tape of Alg. 4 for a flat-ΛCDM radiation-less expansion history.

Primitives	Derivatives
$W_0 \equiv \Omega_\mathrm{m}$	$\partial_i W_0$
$W_1 \equiv z$	$\partial_i W_1$
$W_2 = 1 + W_1$	$\partial_i W_2 = \partial_i W_1$
$W_3 = W_2^3$	$\partial_i W_3 = 3W_2^2 \partial_{\Omega_\mathrm{m}} W_2 = 3W_2^2 \partial_i W_1$
$W_4 = W_0 * W_3$	$\partial_i W_4 = \partial_i W_0 W_3 + W_0 \partial_i W_3 = \partial_i W_0 W_3 + 3W_0 W_2^2 \partial_i W_1$
$W_5 = 1 - W_0$	$\partial_i W_5 = -\partial_i W_0$
$W_6 = W_4 + W_5$	$\partial_i W_6 = \partial_{\Omega_\mathrm{m}} W_4 + \partial_{\Omega_\mathrm{m}} W_5 = \partial_i W_0 W_3 + 3W_0 W_2^2 \partial_{\Omega_\mathrm{m}} W_1 - \partial_i W_0$
$H(z) \equiv W_7 = \sqrt{W_6}$	$\partial_i W_7 = \frac{\partial_i W_6}{2\sqrt{W_6}} = ((W_3 - 1)\partial_i W_0 + 3W_0 W_2^2 \partial_i W_0)/(2\sqrt{W_6})$

Analogously to when humans perform the chain rule, AD can operate in two different modes, forwards and backwards. The distinction boils down to whether we start applying the chain rule by computing the derivatives of the primitives with respect to the independent variables until reaching the original function or if instead we start by computing the derivative of the original function with respect to the primitives until reaching the independent variables. The first strategy corresponds to forwards AD while the second is backwards AD. Going back to our example, if we wanted to compute $\partial_{\Omega_\mathrm{m}} H(z)$, forwards AD would compute it as $\partial_{\Omega_\mathrm{m}} H(z) = \frac{\partial W_6}{\partial W_1} \frac{\partial W_7}{\partial W_6} \frac{\partial W_8}{\partial W_7}$ while backwards AD would instead perform $\partial_{\Omega_\mathrm{m}} H(z) = \frac{\partial W_8}{\partial W_7} \frac{\partial W_7}{\partial W_6} \frac{\partial W_6}{\partial W_1}$.

While backwards and forwards AD are algebraically identical, they are computationally very distinct. Given a function $f\colon \mathbb{R}^\mathbb{N} \to \mathbb{R}^\mathbb{M}$, forwards AD will be more efficient at computing ∇f if $N < M$. If, on the other hand, $M < N$, backwards AD is preferred [97][17]. This difference in performance originates from the fact that the number of passes the forwards AD needs to generate the Wengert tape of a program scales with the number of inputs while backwards AD scales with the number of outputs. This makes backwards AD particular useful in optimisation problems with a scalar loss function [97]. Wengert's first algorithm operated in the forwards mode and went on fairly unnoticed. Backwards AD was originally proposed by Seppo Linnainmaa in his Master's thesis written in Finnish and was only translated into English in 1976 [98]. Today, backwards AD is one of the key technologies underpinning the multi-billion dollar AI industry. Indeed, most of the popularity of AD is due to its application to neural networks where its backwards mode allows us to

---

[17] The performance of backwards and forwards AD is also impacted by the size of the operation being differentiated.

optimise models with millions of parameters. Hence why AD is also commonly known as backpropagation.

A careful reader might have realised that I have stealthily glided over how computers *actually* generate the Wengert tape of a program. This is because taping is a rather arcane process with a great deal of complexity. However, we need to develop an operational understanding in order to learn how to write computer programs compatible with AD algorithms. The two most popular approaches to write this tape are through either Operator Overload (OO) or by Source Transformation (ST) [99, 100]. On the one hand, OO extends the operators of a given programming language to include a tracing operation that tracks the operations that the inputs of a given program undergo. This makes OO fairly permissive of most modern coding practices such as control flow or variable mutation. However, tracking every single input can add significant overhead as well as memory requirements, leading to inefficiencies. On the other hand, ST extends the programming language compiler to explicitly write a tape for the program as a form of intermediate representation. This puts much more stringent demands on the control flow of the program as well as other open-ended operations it may contain. However, it is far more efficient than OO [101].

In addition to this, it is also possible to perform completely tapeless AD using dual numbers. Dual numbers are special type of numbers of the form $a + b\epsilon$ where $\epsilon^2 = 0$ but $\epsilon \neq 0$. Thus, it can be shown that given a generic function, $f(x)$, its derivative can automatically obtained by considering $f(a+\epsilon) = f(a) + \epsilon f'(a)$ where $a \in \mathbb{R}$. This can be proven by considering its Taylor expansion of the previous expression:

$$f(a+\epsilon) = \sum_{n=0}^{\infty} \frac{f^{(n)}(a+\epsilon)}{n!}(x-a-\epsilon)^n \approx \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a-\epsilon)^n = \tag{1.150}$$

$$(f(a) + f'(a)(x-a) + \frac{1}{2}f''(a)(x-a)^2 + ...) + \epsilon(f'(a) + f''(a)(x-a) + ...) = \tag{1.151}$$

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x+a)^n + \epsilon \sum_{n=0}^{\infty} \frac{f^{(n+1)}(a)}{n!}(x+a)^n = \tag{1.152}$$

$$f(a) + \epsilon f'(a)\,. \tag{1.153}$$

Therefore, AD can also be achieved by teaching computer programs how to parse dual numbers without the need of a schematic representation of the program being differentiated through. Since dual numbers are a special type of input, AD based on this technique is limited to the forward mode which might be a problem depending on the application.

1.4 Gaussian Processes

1.4.1 Motivation

In the previous sections we have focused on how better modelling of systematics will lead to an unmanageable number of free parameters for future cosmological analyses. However, there are good reasons to believe that systematics might not be the only aspect that need better modelling.

The dramatic increase in precision experienced by cosmology over the last 30 years has recently led to the discovery of potential inconsistencies in our cosmological model that previously might have been obscured by statistical errors. The most famous manifestation of this tension is the 4σ to 5σ difference in the value of the Hubble constant [102, 103], H_0, between supernova luminosity distance estimates [104, 105] and the early-Universe cosmic microwave background (CMB) probe *Planck* [13]. Somewhat less known but equally important is the fact that late-Universe probes of the large-scale structure, such as weak gravitational lensing (WL) and galaxy clustering, also prefer a 3σ to 4σ lower amplitude of the growth of structure than *Planck*. This tension is encapsulated in the parameter $S_8 = \sigma_8\sqrt{\frac{\Omega_\mathrm{m}}{0.3}}$, where σ_8 is the standard deviation of matter density fluctuations in spheres of radius $8\,h^{-1}$Mpc today and Ω_m is the matter density parameter. Hence, it is known as the S_8 tension [54, 60–62, 106, 107].

Both of these tensions are highly suggestive of either unaccounted systematic effects or new physics beyond the ΛCDM model [108]. If we assume we can trust our measurements, we would still have to identify the physics that need to be modified. Moreover, even if we successfully identified the new physics, we would still have to come up with a specific parametrisation for the phenomena [109]. This problem has single-handedly propped an entire publication industry based on considering ever so slightly different parametrisations of different beyond-ΛCDM physics. Unfortunately, there is no a candidate currently that can resolve the aforementioned tensions while leaving the successes of the ΛCDM model intact [110].

Model-agnostic Cosmology offers a path forward to continue probing potential new physics by inverting the aforementioned process. Instead of marrying a particular parametrisation of the new phenomena and then contrasting it with the available data, one can *infer* directly from the data the favoured functional forms and then map said forms to particular theories. Thus, model-agnostic Cosmology is also known as data driven Cosmology or non-parametric Cosmology. The later refers to the fact that no one parametrisation is preferred. However, these methods often involve more parameters, not less!

As one can imagine, model-agnostic Cosmology boils down to the particular method one uses to determine the functional forms from the data. Two of the most popular ways of doing so are symbolic regression [111, 112] and Gaussian processes [113]. The key difference between the two approaches is that symbolic regression considers the functional forms explicitly while Gaussian processes do so implicitly as expectations values of high dimensional distributions. In this thesis, we will focus our discussion of Gaussian processes for the sake of brevity.

1.4.2 Formalism

In the same way that probability distributions describe the statistical properties of scalars or vectors, stochastic processes describe the statistical properties of functions. Thus, a Gaussian process (GP) is a continuous generalisation of the Gaussian probability distribution. Formally, a GP is a collection of random variables (nodes), each of them sampled from a multivariate Gaussian distribution [114]. Thus a GP $g(\mathbf{x})$ where $\mathbf{x}$ is a arbitrary vector representing the position of the nodes, is fully specified by a mean function:

$$m(\boldsymbol{x}) \equiv \mathbb{E}[g(\boldsymbol{x})] \,, \tag{1.154}$$

where $\mathbb{E}[\cdots]$ is the expectation value over the ensemble of realisations of $g(\boldsymbol{x})$, and a covariance function:

$$K(\boldsymbol{x}, \boldsymbol{x'}) \equiv \mathbb{E}[(g(\boldsymbol{x}) - m(\boldsymbol{x}))(g(\boldsymbol{x'}) - m(\boldsymbol{x'}))] \,. \tag{1.155}$$

In combination, the mean and covariance functions determine the statistical properties of the random variables that thus define the family of shapes (functions) that the GP can take. This is obvious if we consider the eigenfunction decomposition of the covariance function:

$$K(\boldsymbol{x}, \boldsymbol{x}') = \sum_{i=1}^{\infty} \lambda_i \phi_i(\boldsymbol{x}) \phi_i(\boldsymbol{x}')^\dagger \,, \tag{1.156}$$

where λ_i denotes the eigenvalues and ϕ_i the eigenfunctions. These eigenfunctions describe the basis of a space of functions thus establishing the link between the covariance matrix and the family of functions it embodies.

The choice of mean function is trivial since it is always possible to map the problem to a zero-mean scenario by removing the mean of the observations. However, parametrising the covariance matrix has to be done rather carefully in order to reflect the statistical properties of the observations or the assumed underlying function. The most common such parametrisation is the square exponential covariance matrix given by the following kernel:

$$K(\boldsymbol{x}, \boldsymbol{x}') = \eta^2 \exp\left[\frac{-|\boldsymbol{x} - \boldsymbol{x}'|^2}{2l^2}\right] , \tag{1.157}$$

where η and l are two so called "hyperparameters" that control the statistical properties of the family of functions described by the GP. In this particular case, η controls the amplitude of the GP in the absence of data while l acts as a correlation length. In other words, l gives us the distance after which a point in the GP becomes effectively independent from another one.

GPs have been extensively used in astrophysics as tools to model different quantities in an agnostic way [1, 2, 115–126]. Given a likelihood $\mathcal{L}(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\sigma})$ for a set of data points $\boldsymbol{y}(\boldsymbol{x}_y) \equiv \boldsymbol{y}$ at a set of positions $\boldsymbol{x}_y$, with a set of errors $\boldsymbol{\sigma}$, and a set of random variables $\boldsymbol{x}$, a GP can be employed as a prior over all the possible families of functions used to fit the observations.

Observations can then be used to inform the GP posterior (i.e. the statistical properties of the assemble of random variables), $\mathcal{P}(g(\mathbf{x})|\mathbf{y}, \boldsymbol{\sigma})$, which determines the family of functions most consistent with the data. This makes GPs extremely versatile as agnostic models for functions we might know little about. Moreover, GPs are also easy to interpret. Being probability distributions over a basis of functions, GPs have a naturally built-in concept of statistical certainty encapsulated in the covariance matrix of their posterior distribution. In Fig. 1.8 we show an example how data can be used to constrain the family of functions described by a GP. The left-hand side panels show the GPs mean and standard deviation prior to the inference process (bottom panel) as well as the correlation matrix between the nodes (top panel). Similarly, the right-hand side panels show the equivalent quantities after the inference process. We can observe how the mean GP has been adapted to follow the trends of the data. Moreover, its standard deviation follows the sparsity of the data points, narrowing in populated regions and brocading where observations are lacking. Looking at the correlation matrix we can see that the data has induced a non-trivial structure which washes off once data becomes sparse (where the GP returns to its prior).

When the relationship between the data and the GP is linear, GPs become particularly attractive modelling tools since the posterior distribution of the GP can be found analytically. Consider a set of points $\boldsymbol{x}^*$ where we want to probe the predictions of GP, it can be shown [113] that the mean of the GP in light of observations is given by:

$$\boxed{\mathbb{E}\left[g(\boldsymbol{x}^*)\right] = K(\boldsymbol{x}_y, \boldsymbol{x}^*)^T \left[K(\boldsymbol{x_y}, \boldsymbol{x_y'}) + \boldsymbol{\sigma}\right]^{-1} \boldsymbol{y}} . \tag{1.158}$$

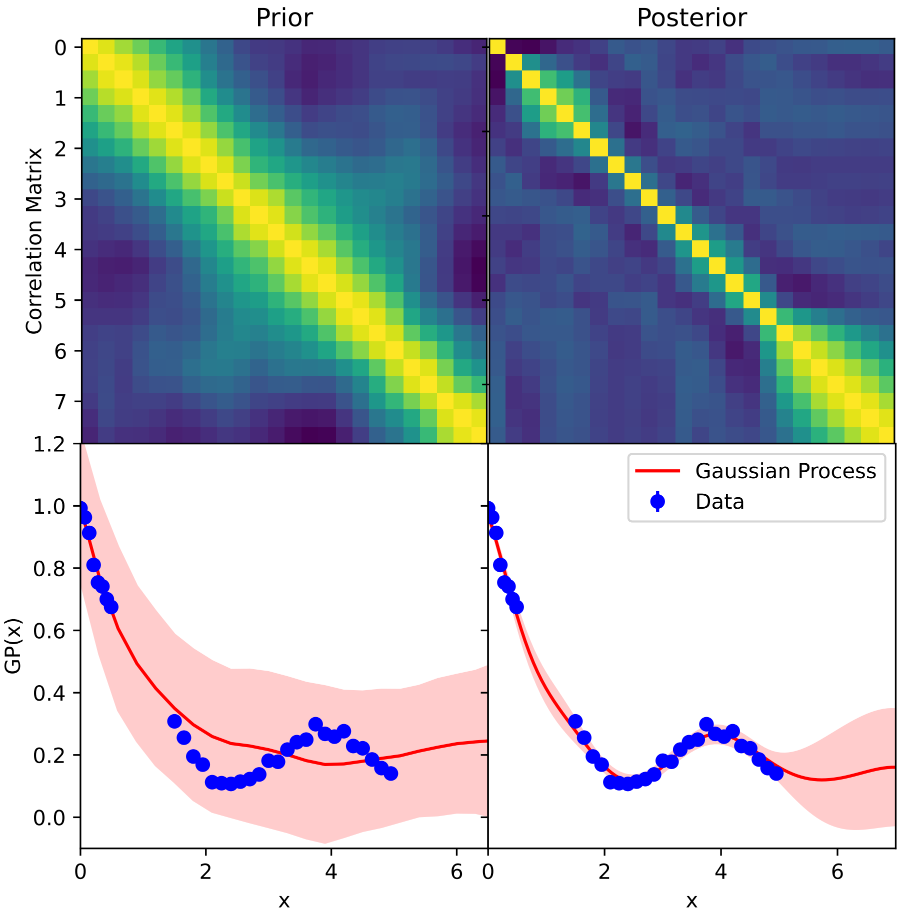

Figure 1.8: The left-hand side panels show the GPs mean and standard deviation prior to the inference process (bottom panel) as well as the correlation matrix between the nodes (top panel). Similarly, the right-hand side panels show the equivalent quantities after the inference process.

Similarly, the covariance matrix of these predictions is given by:

$$\boxed{cov\left[g(\boldsymbol{x^*})\right] = K(\boldsymbol{x}^*, \boldsymbol{x}^{*'}) - K(\boldsymbol{x}_y, \boldsymbol{x}^*)^T \left[K(\boldsymbol{x}_y, \boldsymbol{x}_y') + \boldsymbol{\sigma}\right]^{-1} K(\boldsymbol{x}_y, \boldsymbol{x}^*)}\,. \tag{1.159}$$

Most importantly, neither Eq. 1.158 nor Eq. 1.159 depend on the original random variables $\boldsymbol{x}$. Thus the final predictions only depend on the chosen kernel for the GP which implicitly depends on the GP hyperparameters.

When the relationship between the data and the GP is non-linear, it is no longer possible to find the posterior of the GP analytically. Moreover, the posterior will depend on the position of the original random variables $\boldsymbol{x}$ meaning that we have to explicitly treat them as free parameters. In such scenario, it is useful to think of GPs as a mapping from a white noise process to a correlated process. Formally this can be expressed as:

$$\boxed{g(\boldsymbol{x}) = m(\boldsymbol{x}) + \mathbb{L}(K) \cdot \boldsymbol{\nu}}\,, \tag{1.160}$$

where $\boldsymbol{\nu} \sim \mathcal{N}(0, \mathbb{I})$ and $\mathbb{L}(K)$ is the lower Cholesky triangle of the GP kernel which acts as a rotation of $\boldsymbol{\nu}$ on to the correlated space. The reason why Eq. 1.160 is a useful representation is because it is much simpler to sample a white noise process than directly from a GP.

Since in the non-linear scenario nodes become free parameters it often becomes undesirable, if not simply unfeasible, to directly sample the GP at the data points. This is specially true if the data are numerous, which is a pre-requirement to perform data-driven Cosmology. Thankfully, sampling GPs directly at the points is not only inefficient but also unnecessary. Indeed it is more advantageous to consider a small number of nodes which we can then map to the positions of the data using the Wiener filter given by the kernel of the GP:

$$g(\boldsymbol{x}_y) = \mathcal{W}_{\boldsymbol{x}\boldsymbol{x}_y}\left[g(\boldsymbol{x})\right] = \frac{K(\boldsymbol{x}_y, \boldsymbol{x})}{K(\boldsymbol{x}, \boldsymbol{x}')} \cdot g(\boldsymbol{x})\,. \tag{1.161}$$

Predictions for the data can then be found as:

$$y \sim \mathcal{N}(\mathbb{F}\left[g(\boldsymbol{x}_y)\right], \boldsymbol{\sigma})\,, \tag{1.162}$$

where we have assumed a Gaussian likelihood for the sake of argument and $\mathbb{F}$ is the non-linear mapping between the GP and the data.

However, even after using a Wiener filter to map the GP to a lower dimensional space, one might find that the number of free parameters is still too large for traditional inference methods (See Sect. 1.2.1). Here's where another important property of GPs comes into play: GPs are as differentiable as their covariance kernel [113].

Indeed, the derivative of a GP $g(\boldsymbol{x}) \sim \mathcal{N}(m(\boldsymbol{x}), K(\boldsymbol{x}, \boldsymbol{x}'))$ is another GP given by $\dot{g}(\boldsymbol{x}) \sim \mathcal{N}(\partial_{\boldsymbol{x}} m(\boldsymbol{x}), \partial_{\boldsymbol{x}} \partial_{\boldsymbol{x}'} K(\boldsymbol{x}, \boldsymbol{x}'))$. Moreover GPs, by virtue of are abased on linear algebra, are also compatible with all forms of auto-differentiation. Thus, GPs are extremely synergistic with the gradient-based inference algorithms discussed in Sect. 1.2.2.

Finally, once the posterior of the GP has been found numerically or otherwise, its values can be probed at arbitrary points using the Wiener filter given by the posterior covariance matrix of the GP and the set of probing points.

1.4.3 Interpretation

In the previous section, we described how to use GPs as agnostic priors over families of functions. In this section, we will discuss how to interpret the results obtained. However, as we will see, this is no easy task. The problem fundamentally stems from the fact that a GP is not a single parameter with a singular figure of merit (e.g. the standard deviation), but a vector of parameters. Nonetheless, we need to devise a compact and useful way of compressing (and comparing) the information we get from the GP.

As discussed in Sect. 1.4.2, the statistical properties of a GP are encapsulated in its mean and covariance function. Therefore, if one wishes to measure how constrained a GP is, the first intuition would be to turn to the covariance matrix of the GP's posterior; the multi-dimensional equivalent of the standard deviation. The problem that arises is finding a way to compress such a covariance matrix into a meaningful measurement

On the one hand, one could look at the determinant of said covariance matrix. However, the determinant mixes contributions from both the diagonal elements of the matrix; i.e. the standard deviation in each node, and from the off-diagonal elements of the matrix; i.e. the correlations between the nodes, in a non-trivial way that obfuscates its interpretation. One could think of just think of looking at the diagonal of said covariance. However, this approach runs into the opposite problem. How does one interpret two GP's with the same variance but with different degrees of correlation? In the context of CMB data compression, Bond *et al.* [127] proposed to study one over the diagonal of the inverse covariance matrix as a way of incorporating off-diagonal contributions. Alternatively, one could diagonalise the covariance matrix. However, both inverting ad diagonalising are non-linear operations which makes interpreting the resulting errors a non-trivial task.

On the other hand, one could take advantage of the so called hyperparameters of the GP. Hyperparameters dictate the values that the nodes are allowed to take and that act as a high level description of the statistical properties of the nodes. The most relevant hyperparameter would be the amplitude of the GP covariance matrix which dictates how much the GP can deviate from its mean. This measurement partially solves the issue of including the off-diagonal entries in an interpretable manner since the hyperparameter controls the amplitude of both the diagonal and off-diagonal elements of the matrix. However, it is unclear how to compare two covariance matrix amplitudes with two different correlation length values. Most importantly, this measurement of uncertainty does not directly relate to the nodes of the GP themselves, only to their allowed values. In summary, there is not a singular way of quantifying the uncertainty of a GP, especially using one single number.

Chapter 2

Model-independent constraints on Ω_{m} and $H(z)$ from the link between geometry and growth

2.1 Introduction

When observing and characterising the Universe on large scales, there are two broadly different, yet intertwined, types of observations [128]. In the first type of observation, one endeavours to constrain the expansion rate of the Universe at different times. This can be done by measuring the expansion rate itself at a particular redshift. Alternatively, one can measure the expansion rate as an integrated effect in form of different cosmological distance measures such as angular diameter distances, luminosity distances, standard sirens, etc. Fundamentally, all of these distances are derived from the radial comoving distance,

$$\chi(z) = \int_0^z \frac{c dz}{H(z)} \,. \tag{2.1}$$

In the second type of observation, one focuses on the rate at which structures undergo gravitational collapse on the expanding background. It can be shown (see Sect. 1.1.2) that under the assumptions that the bulk of matter in the recent past can be described by a pressureless fluid and that neutrino effects can be neglected, the density contrast, δ, obeys the following evolution equation of the form:

$$f' + f^2 + \left(1 + \frac{\mathrm{d}\ln aH}{\mathrm{d}\ln a}\right) f = \frac{3}{2}\Omega_{\mathrm{m}}(z) \,, \tag{2.2}$$

where the prime denotes a derivative with regards to $\ln a$, the growth rate is defined as $f \equiv \mathrm{d}\ln\delta/\mathrm{d}\ln a$ and $\Omega_{\mathrm{m}}(z)$ is the fractional energy density in matter as a function of redshift. The latter quantity depends, through the Einstein Field Equations, on $H \equiv H(z)$ so that

$$\Omega_{\mathrm{m}}(z) = \frac{\Omega_{\mathrm{m}} H_0^2}{a^3 H^2}, \tag{2.3}$$

with $\Omega_{\mathrm{m}} \equiv \Omega_{\mathrm{m}}(0)$, which we use for ease of notation. Thus, a measurement of the growth rate of structure at different redshifts (or times) can also be used to reconstruct $H(z)$ over time as well as the fractional matter density today, Ω_{m}.

The challenge of modern cosmology has been to use a wide range of different cosmological observables to constrain $H(z)$ as a function of time or redshift (and, of course, Ω_{m}) and, crucially, to pin down the underlying cosmological model which describes $H(z)$ in terms of a greatly reduced set of cosmological parameters. The front-runner is the Λ Cold Dark Matter (ΛCDM) model, whose parameters are now constrained to an unprecedented precision [13]. However, interestingly, inconsistencies, or *tensions*, have begun to emerge. Different probes are leading to different constraints on, for example, the Hubble constant, H_0 [e.g. 104, 129, 130] or the density-weighted amplitude of fluctuations, $S_8 \equiv \sigma_8 \left(\Omega_{\mathrm{m}}/0.3\right)^{0.5}$ [e.g. 54, 60, 61, 131–135], where σ_8 is the variance of δ, in spheres of radius $8\,h^{-1}$ Mpc.

In this chapter, we ask if it is possible to obtain meaningful, or precise, constraints on cosmological parameters with minimal assumptions about the cosmological model. To be more specific, we step back and try to find model-independent constraints on $H(z)$ and Ω_{m} from measurements of the expansion history itself, cosmological distances and the growth rate. This allows us to contrast the constraining power of these two very different sets of observables, to explore how combining them improves constraints and, most importantly, how much constraints improve once one assumes a cosmological model. The hope is that understanding this process will shed light on the analysis of theories that go beyond ΛCDM, but also may have bearing on the current inconsistencies in parameter constraints. We will implement our model-independent approach using Gaussian processes [113, 114]. Gaussian processes allow us to reconstruct a well-defined distribution of histories of $H(z)$. The data then allows us to constrain the parameters of this distribution and, in doing so, tells us at what redshifts $H(z)$ is well determined and at what redshifts it is determined poorly.

The literature already hosts a number of examples of the possible uses of Gaussian Processes to test certain aspects of the present cosmological paradigm in model-agnostic ways. For example, one can find tests for the model dependence of the H_0 [115–118] and S_8 [119, 120] tensions, for the non-zero curvature of space-time [121–123], as well as tests for the density and the equation of state of dark energy

[124, 125]. More closely related to the topic of this work, in Perenon *et al.* [126] Gaussian processes were used to study the statistical correlations between the expansion history, cosmological distances, and the linear growth rate without appealing to the physical relationships between the three functions. However, the clearest precedent of the methodology used here is Li *et al.* [136] who already employ Gaussian Processes to obtain model-independent constraints for Ω_m and σ_8 based on the relationship between the expansion history and the linear growth rate. The methodology developed in the present chapter expands and improves many aspects of their analysis. First, our Gaussian process extends all the way to recombination. This allows us to solve the Jeans equation (Eq. 3.3) without using fitting formulas. Moreover, we also employ cosmological distances to constrain the evolution of the expansion history. This, in combination with the extended range of the Gaussian Process, allows us to use the position of the first acoustic peak of the CMB temperature power spectrum to constrain expansion history far into the past. Finally, and most importantly, we sample our Gaussian Process simultaneously with the cosmological parameters, allowing us to observe potential correlations between the two.

The structure of this chapter is as follows: in Section 2.2 we present the methodology of this work regarding the use of Gaussian Process to compute predictions for cosmological observables. In Section 2.3 we describe the cosmological observables from which we employ data and motivate their use in the context of this work. We present our results in Section 2.4 and discuss the implications of our work in Section 2.5.

2.2 Methods

2.2.1 A Gaussian Process for $H(z)$

Connecting the expansion rate with observations when $H(z)$ is directly observed is trivial. In the case of distances, we can see from Eqs. 1.22, 1.25 and 1.27 that it is possible to generate predictions for the observables from any choice of $H(z)$ by performing the relevant integral. In the case of growth (Eq. 1.48), we are faced with the problem that most observations do not report f but the combined quantity $f\sigma_8$. Thus, it is convenient to rewrite Eq. 1.48 in terms of the latter. This can be done as follows: assuming that perturbations grow in a self-similar manner, we can define $\delta(t, \mathbf{x}) \equiv D(t)\delta_0(\mathbf{x})$, where δ_0 is the density contrast today. Then, we can rewrite Eq.

1.48 as

$$\frac{d}{da}\left(a^2 H \frac{dD}{d\ln a}\right) = \frac{3}{2}\Omega_{\rm m}(a)\, a\, H\, D\,. \tag{2.4}$$

Defining $y \equiv a^2 E \frac{dD}{d\ln a}$, with $E \equiv H/H_0$, and switching to the integration variable $s \equiv \ln(1+z)$, this equation can be written in terms of a system of coupled first-order equations:

$$\frac{dy}{ds} = -\frac{3}{2}\frac{\Omega_{\rm m}}{aE} D\,, \quad \frac{dD}{ds} = -\frac{y}{a^2 E}. \tag{2.5}$$

Then, it is possible to transform the quantities $y(s)$ and $D(s)$ into $f\sigma_8(s)$ and $\sigma_8(s)$ by applying the following transformation

$$\sigma_8(s) = \sigma_8(0)\frac{D(s)}{D(0)}\,, \quad f\sigma_8(s) = \frac{y(s)\,\sigma_8(0)}{E(s)D(0)} e^{2s}\,. \tag{2.6}$$

The thrust of this work is to keep the analysis as model-independent as possible. Yet, as we can see, it is still possible to extract information about some of the cosmological parameters. For a start, combining the information from distances and growth allows us to constrain $\Omega_{\rm m}$. But we also have, automatically, $H_0 \equiv H(z=0)$ and, as we just saw, we can calculate $\sigma_8 \equiv \sigma_8(z=0)$ (or S_8 as a derived parameter).

In this chapter, we will model the time (or redshift) dependent Hubble rate, $H(z)$, as a Gaussian Process (GP) spanning over the redshift range $0 < z < 1100$. We describe GPs extensively in Sect. 1.4. For the purposes of this chapter it is enough to remember that GPs are a generalisation of Gaussian distributions and that as such they are fully determined by a mean and covariance function:

$$GP(z) \sim \mathcal{N}(m(z), K(z,z'))\,. \tag{2.7}$$

Since the Hubble rate is generally regarded as a monotonically increasing function, it is important to define a non-zero mean for the GP. This prevents the GP from simply fitting the long range upwards trend of $H(z)$ while washing out interesting local features [137]. In this chapter, we define the mean of the GP in terms of the Planck 2018 TTTEEE [13] best fit ΛCDM expansion history, $H_{\rm P18}(z)$.

In order to prevent the choice of mean from biasing our constraints on the cosmological parameters, we define a free amplitude parameter A_0 that multiplies the mean of the GP such that

$$H_m(z) \equiv A_0\, H_{\rm P18}(z)\,. \tag{2.8}$$

In a Bayesian framework, this is equivalent to setting a prior over a family of possible mean functions where the distribtion on A_0 captures our prior degree of confidence in Planck 2018's expansion history being the true expansion history of the Universe.

In [1], we show these effects can substantially bias the constraints we obtain in the absence of A_0. Simply put, if the data is somewhat away from what one might expect from the fiducial ΛCDM background, the GP will have to soak up the large scale differences, or trends, in detriment to the small scale, local features. Moreover, in those regions, where data is sparse, the GP's constraints are highly dominated by their prior; i.e. the chosen mean. If such a mean systematically falls beneath or above the data, it can lead to spurious trends in the final results. For these reasons, we advise against common practices such as employing a constant as a mean for a GP modelling the recent expansion history or extrapolating GP results to regions where they become dominated by the choice of mean and covariance function.

The Hubble rate must also be a continuous and smooth function. Therefore, it is important that the covariance function of the GP reflect such properties [138]. Ultimately, we chose a square exponential covariance function to model the correlations between the different nodes of the GP (see Eq. 1.157). This decision was made based on the fact that the square exponential is computationally inexpensive and infinitely differentiable kernel appropriate for modelling smooth fluctuations around the mean of the GP. In addition to this, a white noise term with amplitude 10^{-3} was added to the covariance function to ensure numerical stability. This particular amount of noise was found to reliably make the kernel invertible while remaining a sub-dominant component of the noise budget. Thus, the final covariance function is given by:

$$K(\boldsymbol{x}, \boldsymbol{x}') = \eta^2 \exp\left[\frac{-|\boldsymbol{x} - \boldsymbol{x}'|^2}{2l^2}\right] + \sigma\mathbb{I}\,, \tag{2.9}$$

where η and l are known as the GP's hyperparameters and control the amplitude and the correlation length of the process respectively. Moreover, σ is the amplitude of the *jitter*. Finally, we define the GP not over the Hubble rate itself but as a relative deviation from the ΛCDM background:

$$H(z) = H_m(z)[1 + \delta H(z)]\,. \tag{2.10}$$

This allows us to define our GP as a zero mean process given by:

$$\delta H(z) = \mathbb{L}(K) \cdot \boldsymbol{\nu}\,, \tag{2.11}$$

where nodes of the GP are distributed as $\boldsymbol{\nu} \sim \mathcal{N}(0, \mathbb{I})$ and $\mathbb{L}(K)$ denotes the lower Cholesky triangle of the GP kernel. This approach has the added benefit of normalising the amplitude of the oscillations of the GP with respect to the mean making the sampling of the hyperparameters of the GP kernel far more efficient by virtue of reducing its potential range of values.

2.2.2 Likelihood

As discussed in Sect. 1.4, sampling over GPs is extremely efficient when the model for the observed data is linearly related to the GP and the likelihood is Gaussian. In this case, each node of the GP; i.e. each dimension of the multivariate Gaussian distribution, must be treated as a newly added degree of freedom in the sampling process (i.e. a new parameter). As described in Sect. 1.2, exploring such a vast parameter space is effectively unfeasible with traditional non-gradient-based sampling algorithms and requires more sophisticated samplers. In this work, we employ the No U-turn Sampler (NUTS) [77] described in Sect. 1.2.2.2. We employ the NUTS implementation of the python package `Pymc3` [139]. This inference algorithm allows us to handle hundreds or thousands of parameters during the inference process efficiently.

A concern when using gradient-based inference algorithms is how to efficiently compute the gradient of the likelihood function. As discussed in Sect. 1.3.2, in the absence of analytical derivatives auto-differentiation algorithms are the most efficient way of computing said gradients. `Pymc3` makes use of the tensor package `Theano` [140] as a back-end to build a symbolic graph (See Sect. 1.3.2 for a discussion of program graphs and Wegnert tapes) of the model used to fit the data. This graph is then used to perform AD [141] to obtain the gradient of the regression model.

Through out the likelihood we make use of simple integration schemes; e.g. Runge-Kutta, trapezoidal rule, on which THEANO can apply automatic-differentiation to make the likelihood differentiable in all parameters. In the case of distances (see Eq. 1.22), the integral over $1/H_{\mathcal{GP}}$ was performed using the trapezoidal rule with $s \equiv \ln(1+z)$ as the integration variable. In the case of the linear growth rate, the system of differential equations shown in Eq. 2.5 was solved for $y(s)$ and $D(s)$ employing a second order Runge-Kutta scheme, integrating over the redshift range $0 < z < 1100$ from the past into the present with initial conditions $D(z = 1100) = a(z = 1100)$ and $y(z = 1100) = a(z = 1100)^3 E(z = 1100)$. Note that these initial conditions are only strictly correct for a purely matter-dominated universe, which is not accurate at $z \sim 1000$. However, the impact of this assumption is negligible by comparison to current uncertainties on growth measurements. In combination, these methods allow us to obtain sub-percent accuracy through out the entire redshift range of the GP with respect to the output of cosmological CLASS [22], which performs the full numerical calculation.

In order to cover the redshift range $0 < z < 1100$ while keeping the numerical errors under control, we sampled the GP evenly in s as opposed to redshift itself. This variable concentrates most of the GP samples at low redshift where most of our

Table 2.1: Sampled parameters and their priors in the analysis. $\mathcal{N}(m, s)$ describes a Gaussian distribution with mean m and standard deviation s and $U(a, b)$ a uniform distribution with boundaries a and b.

Parameter	Prior
η	$\mathcal{N}_{1/2}(0, 0.2)$
l	$U(0.01, 6)$
Ω_m	$U(0, 1)$
σ_8	$\mathcal{N}(0.8, 0.5)$
A_0	$U(0.8, 1.2)$
M	$\mathcal{N}(-19.2, 1)$
r_s	$\mathcal{N}(150, 5)$

data lays while offering great numerical accuracy [142]. This allowed us to cover our desired redshift range with only 200 nodes; i.e. with a 200-dimensional GP.

In addition to the 200 parameters associated with the GP, we also sample over the the amplitude of GP mean, A_0, and the hyperparameters of the GP kernel, η and l, described in Eq. 1.157. We also marginalise over the cosmological parameters present in Eq. 1.48; i.e. Ω_m and σ_8, as well as the absolute magnitude of the supernovae M (See Sect. 2.3.2 for details), the scale of the sound horizon r_s.

It is important to emphasise that we perform a fully Bayesian inference over the GP, as opposed to what it is known as an empirical Bayesian analysis in which first, the marginal likelihood of the GP hyperparameters is maximised; and then, keeping their values fixed, the conditional posterior over the GP is inferred. While for large sets of data, the output of the two approaches converges to the same results, when only sparse data is available the fact that the empirical Bayesian analysis does not account for the full posterior volume of the hyperparameters can lead to an under-accounting of uncertainties.

A summary of the priors used in the fiducial analysis can be found in Tab. 2.1. As a general rule, when performing Bayesian inference one should avoid broad, uniform priors [143]. This is mainly due to the fact that they do not accurately represent the prior knowledge, put a lot of posterior mass in unlikely values and introduce hard boundaries which can be difficult to motivate. This last consideration is particularly important when using HMC since sharp edges in parameter space can lead to the sampling process becoming inefficient. Following this principle, we employed a half-Gaussian distribution with zero mean and standard deviation 0.2 as prior distribution for the hyperparameter η to down-weight extreme deviations (i.e. 20% and above, well within the observational errors on $H(z)$) from the chosen GP mean without introducing any unnecessary hard boundaries[1]. In the case of l we used a uniform distribution between $0.01 < l < 6$ for the following reasons. First, there is no reason to down-weight long correlation modes against short ones or vice versa. Second, looking at Eq. 1.157, it is possible to see that as $l \to 0$ the value of the covariance function approaches zero regardless of the value of η. This opens a vast a volume in the parameter space which can lead to an inefficient sampling. Third, the likelihood function becomes flat at such small scales since there is no information in the data to constrain those small scale modes. Fourth, the expansion rate is expected to have some degree of temporal correlations, and therefore the limit $l = 0$ must be excluded. The parameter Ω_m was sampled from a uniform distribution between 0 and 1 to enforce the physical boundaries on the allowed values for the cosmological matter density when considering flat cosmologies. In the case of σ_8, for which there is no physically-motivated upper limit, we employ a better behaved Gaussian prior centred at 0.8 and with 0.5 standard deviation. Given the degeneracy of A_0 with the GP parameters, the parameter was sampled from a uniform distribution between 0.8 and 1.2, which amply encompasses the current discrepancy in H_0 between CMB data and local measurements. The supernova absolute magnitude parameter M was sampled from a Gaussian distribution centred at value found by the SH0ES collaboration for the parameter $M = -19.2$, with a standard deviation of 1. While the mean value corresponds to a local expansion rate of $H_0 \simeq 74.0$ km/s/Mpc, the standard deviation ensures that all values in the range $50 < H_0 < 100$ km/s/Mpc fall within the 1σ region of the prior distribution. Finally, the sound horizon scale, r_s, was sampled from a Gaussian distribution centred at 150 Mpc with a standard deviation of 5 Mpc. We do this instead of computing r_s from the expansion history in combination with the BBN prior to allow for larger deviations from the fiducial cosmology.

2.3 Observables and Data sets

In order to make the most of the modelling freedom offered by GPs we consider as much data as possible. In this work, we use a combination of several different probes that together account for 91 data points for a variety of cosmological observables. A summary of the data used in this work can be found in Tab. 2.2 and Fig. 2.1.

---

[1] Note that the standard deviation of a half-Gaussian is not well defined. Thus the value we quote is what the standard deviation of a full Gaussian would be.

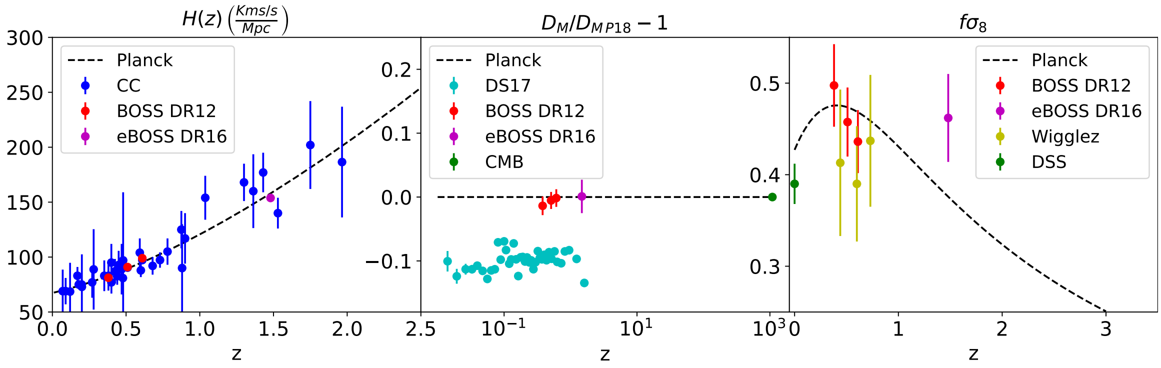

Figure 2.1: Data points from the different surveys used in this work across redshift for the three cosmological functions of interest $H(z)$, D_M and $f\sigma_8$. Radial comoving data is displayed relative to the Planck 2018 mean to show the spread of the data points.

Table 2.2: Data sets used in our analysis, listing the probe, the redshift range of the probe, the choice of observable and the size of the data vector. See Fig. 3.1 for a pictorial representation.

Data Set	Probe	Redshift	Observable			Data Points
			$H(z)$	$D_M(z)$	$f\sigma_8$	
CC's [121]	CC's	0.07 - 2.36	✓	×	×	35
Pantheon DS17[144]	SNe Ia	0.38 - 0.61	×	✓	×	40
BOSS DR12 [73]	BAO+RSD	0.38 - 0.61	✓	✓	✓	3×3
eBOSS DR16 [145]	BAO+RSD	1.48	✓	✓	✓	1×3
Wigglez [146]	RSD	0.44 - 0.73	×	×	✓	3
DSS [147]	RSD	0	×	×	✓	1
Planck 2018 [13]	CMB	1090.30	×	✓	×	1
DESI [148]	BAO+RSD	0.15 - 1.85	✓	✓	✓	3×18

2.3.1 Cosmic Chronometers

Cosmic Chronometers (CCs) are tracers of the evolution of the differential age of the Universe as a function of redshift. Since $H(z) \equiv \dot{a}/a = -(dt/dz)/(1+z)$ a measurement of dt/dz directly yields the expansion rate [149]. By measuring the age difference between two ensembles of passively evolving galaxies at different redshifts, one can determine the derivative of redshift with respect to cosmic time, dz/dt. Massive, early, passively-evolving galaxies have been found to be very good tracers in this sense [150–154] and have been used extensively over the past two decades to measure $H(z)$ up to $z \approx 2$. In this work we make use the $H(z)$ measurements from CCs summarised in Table 1 of Vagnozzi *et al.* [121].

2.3.2 Supernovae

Type Ia supernovae (SNe Ia) — titanic explosions of white dwarfs in multi-star systems [155, 156], are highly prized observations in cosmology due to their capacity to act as standard candles [157, 158].

However, SNe Ia by themselves can only inform their relative distance to one another and need to be calibrated with nearby SNe Ia of known redshift and luminosity distance D_L to obtain absolute distances. Thus, SNe Ia inform the relationship

$$\mu(z) = 5\ln_{10} D_L(z) + 25 + M\,, \tag{2.12}$$

known as the luminosity distance modulus where M is the calibrator known as the absolute magnitude of the SNe Ia. Therefore, once M has been determined SNe Ia luminosity distance measurements can be used to inform the Hubble rate.

In this work, we fit the compressed data vector of the Pantheon sample, known as DS17, composed of 40 measurements of the distance modulus (See Eq. 2.12) in the range $0.15 \leq z \leq 1.615$ [144]. The original Pantheon sample is composed of the optical light curves and redshifts for 365 spectroscopically confirmed Type Ia supernovae (SNe Ia) discovered by the Pan-STARRS1 (PS1) Medium Deep Survey combined with the subset of 279 PS1 SN Ia ($0.03 < z < 0.68$) with useful distance estimates of SN Ia from SDSS, SNLS, various low-z and HST samples to form the largest combined sample of SN Ia consisting of a total of 1048 SN Ia ranging from $0.01 < z < 2.3$ [144].

In the light of recent works in the literature questioning the accuracy of the absolute calibration of important sectors of Pantheon sample [159], we marginalise over the absolute magnitude of the supernovae as opposed to fixing its value. Due to the degeneracy between M and H_0, this is equivalent to fitting the expansion rate, $E(z) = H(z)/H_0$, as opposed to the Hubble rate, $H(z)$.

2.3.3 Baryon Acoustic Oscillations

Baryon acoustic oscillations (BAOs) enhance matter overdensities at a characteristic physical separation scale which corresponds to the size of the sound horizon at the end of the drag epoch, $r_\mathrm{s}(z_\mathrm{d})$ [160, 161]. The sound horizon is defined as the distance a pressure wave can travel from its time of emission in the very early Universe up to a given redshift. This can be expressed as

$$r_\mathrm{s}(z) = \int_z^\infty \frac{c_\mathrm{s}\,\mathrm{d}z'}{H(z')} \,, \tag{2.13}$$

where c_s denotes the speed of sound, and where $H(z)$ is the expansion rate at redshift z. The end of the drag epoch is defined as the time when photon pressure can no longer prevent gravitational instability in baryons around $z \sim 1020$ [162].

The BAO feature can be measured in the directions parallel and perpendicular to the line of sight. Perpendicular to the line of sight, the BAO feature informs the trigonometric relationship

$$\theta = \frac{r_\mathrm{s}(z_\mathrm{d})}{D_M(z)} \,, \tag{2.14}$$

where θ is the angle under which the scale of the sound horizon is observed. Parallel to the line of sight, the BAO feature informs the relationship $\Delta z = H(z) r_\mathrm{s}(z_\mathrm{d})$ which can be used to constrain the expansion history of the Universe directly.

In this work, we make use of the twelfth data release of the galaxy clustering data set of the Baryon Oscillation Spectroscopic Survey (BOSS DR12) which forms part of the Sloan Digital Sky Survey (SDSS) III. BOSS DR12 comprises 1.2 million galaxies over an area of 9329 deg^2 and volume of 18.7 Gpc^3, divided into three partially overlapping redshift slices centred at effective redshifts 0.38, 0.51, and 0.61.

We fit the Alcock-Paczynski (AP) parameters $\alpha_\parallel$ and $\alpha_\perp$ as reported by the BOSS DR12 data products

$$\frac{[H(z)]_\mathrm{fid}}{\alpha_\parallel} = \frac{H(z)[r_\mathrm{s}(z_\mathrm{d})]_\mathrm{fid}}{r_\mathrm{s}(z_\mathrm{d})}, \quad \alpha_\perp [\chi(z)]_\mathrm{fid} = \frac{\chi / r_\mathrm{s}(z_\mathrm{d})}{[r_\mathrm{s}(z_\mathrm{d})]_\mathrm{fid}}, \tag{2.15}$$

from the reconstruction of the BAO feature at the three different redshift bins where $[r_\mathrm{s}(z_\mathrm{d})]_\mathrm{fid} = 147.78$ Mpc is the scale of the sound horizon at drag epoch as given by the fiducial cosmology used for the reconstruction. $[H(z)]_\mathrm{fid}$ and $[\chi(z)]_\mathrm{fid}$ are the corresponding Hubble parameter and comoving radial distance for the fiducial cosmology, respectively.

In addition to this, we employ the anisotropic clustering of quasars in the sixteenth data release of the extended Baryon Oscillation Spectroscopic Survey (eBOSS DR16 163), which forms part of the Sloan Digital Sky Survey (SDSS) IV [164]. The eBOSS DR16 catalogue contains 343,708 quasars between $0.8 < z < 2.2$, from which BAO and RSD measurements are obtained at an effective redshift of $z_{\text{eff}} = 1.48$ [165]. We use the results from the configuration space analysis performed by measuring the two-point correlation function and decomposing it using the Legendre polynomials. Similarly to BOSS DR12, the BAO signal is measured both parallel and perpendicular to the line of sight. This allows for the measurement of the geometrical relationships $D_H(z_{\text{eff}})/[r_{\text{d}}]_{\text{fid}}$ and $D_M(z_{\text{eff}})/[r_{\text{d}}]_{\text{fid}}$ respectively, where $D_H(z) \equiv c/H(z)$ and $[r_{\text{d}}]_{\text{fid}} = 147.3$.

Finally, we make use of the *Planck* 2018 measurement of the BAO angular scale $\theta_* = \frac{D_{\text{A}}(z_*)}{r_{\text{s}}(z_*)}$, where $z_* \sim 1100$ is the redshift of the last scattering surface. We use the *Planck* measurement from the temperature and polarisation maps denoted as TTTEEE + lowE.

2.3.4 Redshift Space Distortions

Redshift space distortions (RSDs) are modifications to the observed redshift of a given object caused by its radial peculiar velocity [166]. RSDs are caused by deviations from the Hubble flow that are gravitationally induced by inhomogeneities in the gravitational potential of the surrounding matter distribution. On large, linear scales, RSDs are dominated by the infall towards overdense structures, known as the Kaiser effect [167]. Clustering two-point statistics as a function of transverse and line-of-sight separation are sensitive to the quantity $f\sigma_8$ via RSDs [168]. Alternatively, SNe Ia themselves can be used as a probe of the local velocity field [147, 169, 170], which can also be used to measure this parameter.

We use the three measurements of $f\sigma_8(z)$ from RSD from the BOSS DR12 data, obtained using the anisotropic clustering of the pre-reconstruction density field [73]. We also include the value of $f\sigma_8(z_{\text{eff}})$ measured from the BOSS DR16 quasar sample. The BAO and RSD measurements of both datasets are extracted from the same set of observations. As such, they are not statistically independent, and we account for their full covariance matrix in our analysis [73, 145]. In addition to these, we use the $f\sigma_8$ measurements reported by the WiggleZ Dark Energy Survey at the Australian Astronomical Observatory [171] at redshifts 0.44, 0.60 and 0.73.

Finally, we also use the value of $f\sigma_8(z=0)$ derived from the measured peculiar velocities of the Democratic Samples of Supernovae [147, DSS], the largest catalogue used to study bulk flows in the nearby Universe, compiled of 775 low-redshift Type Ia and II supernovae (SNe Ia & II).

2.3.5 Synthetic Stage-IV Data

In addition to the previously discussed observables and data sets, we also produce forecasts for future experiments, focusing on the Dark Energy Spectroscopic Instrument (DESI). DESI is a galaxy and quasar redshift survey currently taking data from the Mayall 4 meter telescope at Kitt Peak National Observatory. The baseline area is 14000 sq. deg. We use the forecast errors for the measurements of the Hubble rate, $H(z)$; the angular diameter distance, $D_A(z)$; and the growth rate measurement $f\sigma_8$ reported by Font-Ribera *et al.* [148] to create a set synthetic measurements for $H(z)$, $D_A(z)$, and $f\sigma_8$ at an array of 18 redshifts from 0.15 to 1.85. The synthetic measurements were generated using the best-fit *Planck* 2018 cosmology, including measurement noise following the statistical uncertainties reported by Font-Ribera *et al.* [148].

2.4 Results

2.4.1 Current constraints

2.4.1.1 Cosmological functions

Before presenting the results for the cosmological functions of interest; i.e. the expansion history $H(s)$ and linear growth rate $f\sigma_8$, we must first discuss the results on the core of our analysis, the GP on $\delta H(s)$. We found that our analysis of current data produces $\delta H(s)$ constraints compatible with 0 at all redshifts at less than 1σ deviation. This means that the mean of the GP, $H_m(s) = A_0 H_{\mathrm{P18}}(s)$, is capable of capturing the long range trend of the data allowing the GP to fit local features. The recovered bounds on δH are shown in Fig. 2.2.

Moreover, we quantified how well different combinations of data can constrain $\delta H(s)$. In other words, we measured how strong the agreement of $\delta H(s)$ with zero is, and how it is affected by the data considered and the analysis choices. However, the fact that $\delta H(s)$ has a multivariate distribution means that there is not a unique figure of merit for how well data constrains it. In this work, we focused in two numbers. On the one hand, we looked at the hyperparameter η that constrains the prior distribution

Table 2.3: One dimensional constraints on the hyperparameters η and l (first and second column) and the mean variance of $\delta H(z)$ (third column) for different data combinations.

Analysis	η	l	$\overline{\sigma}(\delta H(s))$
All data (ΛCDM)	Fixed	Fixed	0.011 ± 0.003
DESI+CMB (ΛCDM)	Fixed	Fixed	0.002 ± 0.001
All data	0.113 ± 0.075	3.22 ± 1.62	0.094 ± 0.004
No CMB	0.142 ± 0.096	3.26 ± 1.59	0.125 ± 0.016
No DSS	0.111 ± 0.074	3.22 ± 1.60	0.091 ± 0.005
Growth data	0.147 ± 0.114	3.21 ± 1.65	0.183 ± 0.003
Geometry data	0.122 ± 0.077	3.37 ± 1.49	0.098 ± 0.005
DESI + CMB	0.085 ± 0.074	3.47 ± 1.54	0.08 ± 0.004

of values that $\delta H(s)$ can take. On the other hand, we computed the mean variance of the $\delta H(s)$ samples across redshift. In other words, for each parameter of $\delta H(s)$ (i.e. $\{\delta H(s_1), \delta H(s_2), ..., \delta H(s_{200})\}$), we computed the variance of the HMC samples. Then, we took mean value of those variances, which is equivalent to averaging over redshift. We refer to this statistic as $\overline{\sigma}(\delta H(s))$. The motivation behind $\overline{\sigma}(\delta H(s))$ lies in the fact that it directly translates to average fractional constraints on $H(z)$ that can be readily interpreted.

A summary of the impact of the different data sets and analysis choices on the distribution of η, l and $\overline{\sigma}(\delta H(s))$ can be found in Tab. 2.3. Note that all the analyses quoted in this table were run until the convergence criterion $R-1 < 0.05$ was satisfied for all variables where R is the Gelman-Rubin statistic (See Sect. 1.2 for a discussion of the Gelman-Rubin statistic). For the combination of data sets employed in our analysis, we find $\eta = 0.113 \pm 0.075$ and an $\overline{\sigma}(\delta H(s)) = 0.094 \pm 0.004$ corresponding to an average $9.4 \pm 0.4\%$ constraint on $H(z)$ across redshift. Removing the CMB data point significantly worsened both constraints finding $\eta = 0.142 \pm 0.096$ and an average $12.5 \pm 1.6\%$ constraint on $H(z)$ across redshift. In order to better understand this effect we can look at the first panel in Fig. 2.2. In this figure, we can observe how removing the CMB data from the analysis significantly widens the constraints of the GP, specially beyond $z > 2.5$ (see Table 2.3). This is to be expected as the CMB data is the only point we have above $z = 2.5$. Thus, the constraints from this integrated effect are expected to become dominant in the redshift range between $2.5 < z < 1100$. However, it is important to note that the CMB is not the only contributor to the $\delta H(s)$ constraints over this redshift range since $f\sigma_8$ data also constraints $\delta H(s)$ over its whole domain through its role in solving the Jeans equation. The associated constraint is however very weak. Finally, looking at the obtained l constraints, we can observe that regardless of the data employed l remained largely unconstrained, returning its prior. Thus, our GP on the expansion history doesn't exhibit a strong preference for a particular correlation length when considering either current or future data.

For the purpose of studying the effect of different data types, we split the data points within the data sets employed in the fiducial analysis in two groups: *geometry* – exclusively containing measurements of the expansion history – and *growth* – solely containing $f\sigma_8$ measurements.

As we can see in the second panel of Fig. 2.2, growth data only is much weaker at constraining $\delta H(z)$. From Table 2.3 we see that constraints from growth data alone on $\overline{\sigma}(\delta H(s))$ are approximately twice as wide as those resulting from analysing the entire data set. These constraints are consistent with the prior on the hyperparameter η. On the other hand, the constraining power of the geometry data is only slightly weaker than that of the entire data set. Hence the $\delta H(z)$ constraints are mostly dominated by the geometry data sets as one would expect. Nonetheless, the addition of growth data increases constraining power. This is shown explicitly in the last panel of Fig. 2.2, which shows the results of using geometry data alone compared to those using the full data set. This recovers the expected behaviour: more data increases the constraining power and the contours shrinks.

We now shift the focus of our discussion to the constraints we derive from $\delta H(s)$ for the expansion history itself, $H(z)$, and the linear growth of matter anisotropies, $f\sigma_8$. Comparing the constraints for both cosmological functions from our analysis of current data with the *Planck* 2018 predictions, we find an overall good agreement, finding both functions to contain the *Planck* 2018 predictions within their 2σ confidence contours. This can be seen in Fig. 2.3. Nonetheless, two remarks can be made. First, we observed a greater than 1σ preference for a lower $f\sigma_8$ between $0 < z < 0.75$, mostly driven by the DSS data point. However, the constraining power of current $f\sigma_8$ data is too weak to make a case for the presence of new physics. Second, our model-independent analysis finds the supernova absolute magnitude parameter to be $M = -19.43 \pm 0.026$, a constraint which is in 5σ tension with the SHOES preferred value [103]. However, this is not surprising given the known tension between the data sets used to inform the GP reconstruction.

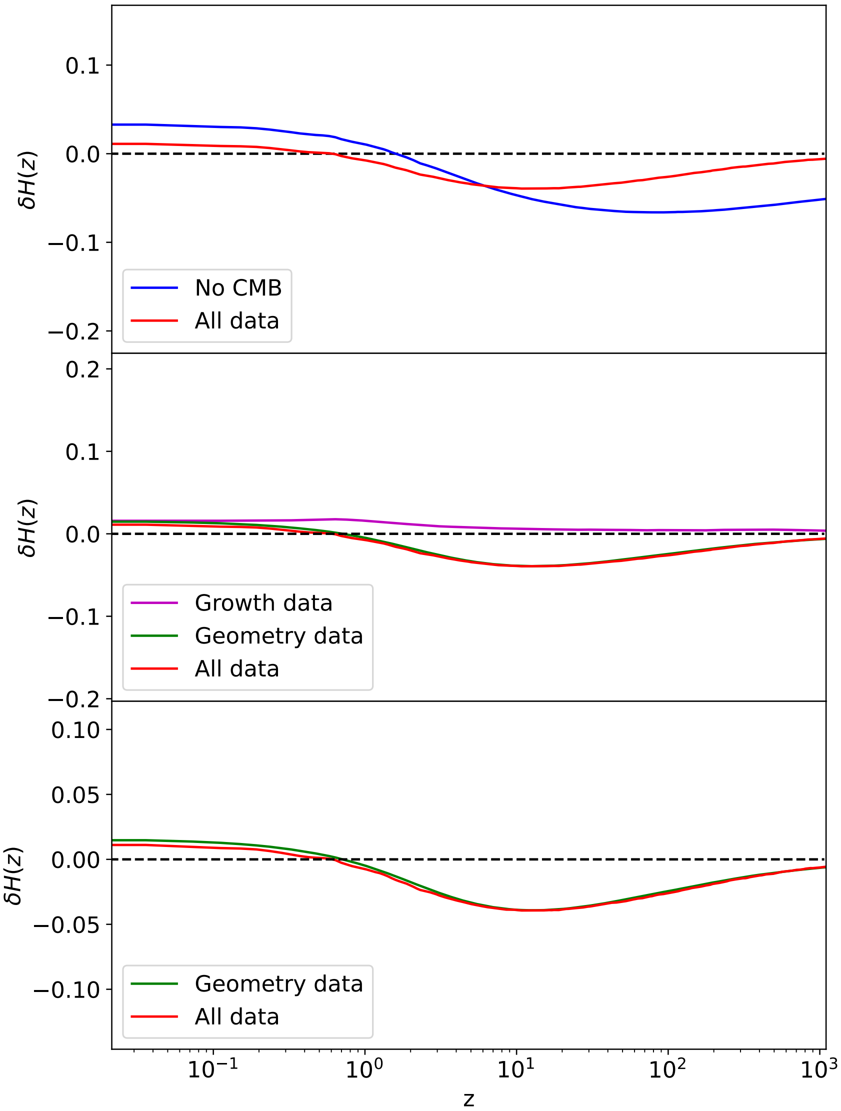

Figure 2.2: 1σ-constraints on $\delta H(z)$ broken down by type of data considered. Solid lines represent the mean of the GPs at each redshift. In red we display the constraints resulting from the analysis of all present data, in blue the effect of removing the CMB data set, in black the effect of fixing the $\Omega_{\rm m}$ and σ_8 to their best-fit (BF) value; in magenta, the constraints resulting of only considering growth data; and in green, the constraints resulting of only considering geometry data.

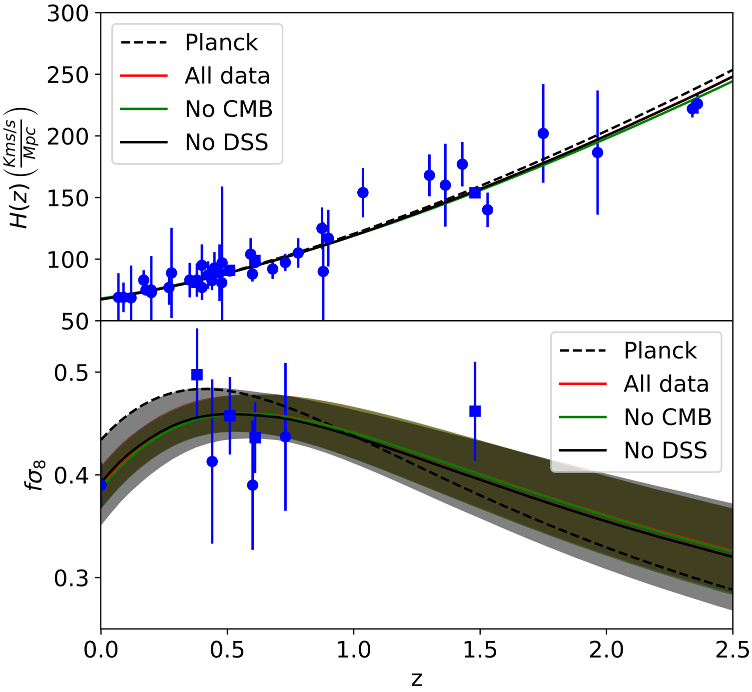

Figure 2.3: 1σ-constraints for the cosmological functions $H(z)$ and $f\sigma_8$ (top and bottom panel respectively) broken down by combination of data set. Solid lines represent the mean of the GPs at each redshift. The dashed black lines show the prediction for each cosmological function using our Planck 2018 best-fit cosmology. In red, we show the constraints resulting of the analysis of all present data; in green, the impact of removing the CMB data set; and in black, the impact of removing the DSS data set.

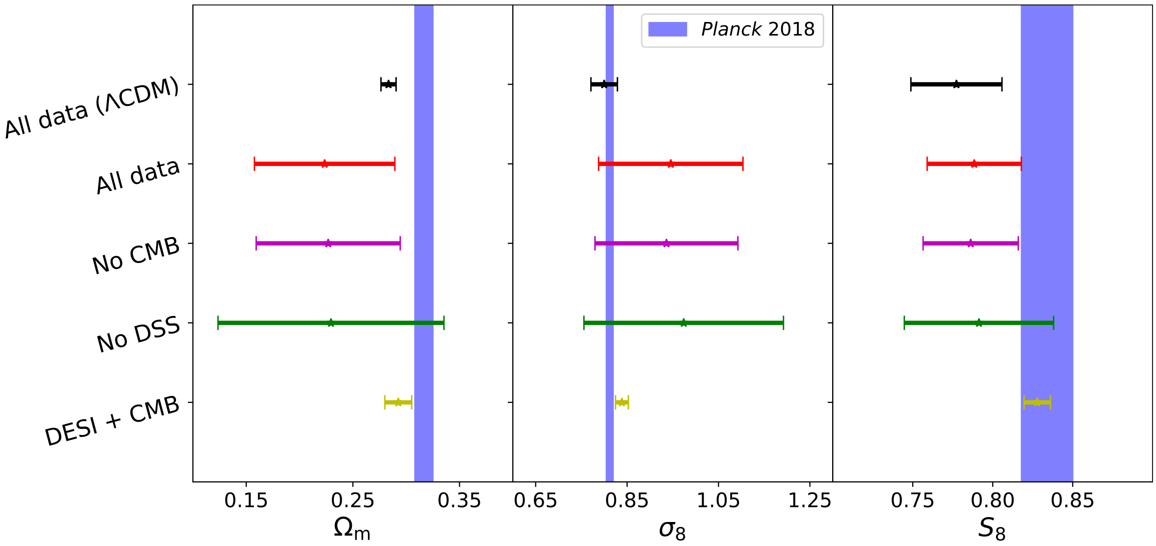

Figure 2.4: Comparison between the different 1σ constraints for the cosmological parameters S_8, σ_8 and Ω_m (first, second and third panel respectively). In each panel the *Planck* 2018 constraint is displayed in the form of a blue band for comparison.

Table 2.4: Constraints on the cosmological parameters Ω_m, σ_8, S_8, and H_0 (first to fourth columns respectively) for each of the different analyses (rows), as well as the *Planck* 2018 constraints and ΛCDM analyses of current and DESI+CMB for reference (first to third row respectively).

Analysis	Ω_m	σ_8	S_8	H_0
Planck 2018	0.317 ± 0.008	0.812 ± 0.007	0.834 ± 0.016	67.27 ± 0.006
All data (ΛCDM)	0.283 ± 0.007	0.8 ± 0.029	0.777 ± 0.028	68.601 ± 0.775
DESI+CMB (ΛCDM)	0.316 ± 0.006	0.812 ± 0.003	0.834 ± 0.008	66.992 ± 0.311
All data	0.224 ± 0.066	0.946 ± 0.158	0.788 ± 0.029	67.715 ± 0.93
No CMB	0.227 ± 0.068	0.936 ± 0.156	0.786 ± 0.03	67.94 ± 1.034
No DSS	0.229 ± 0.106	0.974 ± 0.218	0.791 ± 0.047	67.766 ± 0.96
DESI + CMB	0.293 ± 0.013	0.839 ± 0.014	0.828 ± 0.008	66.788 ± 0.371

2.4.1.2 Cosmological parameters

We now focus on the constraints on the specific cosmological parameters Ω_m and σ_8. We will also show constraints on the derived parameter $S_8 = \sigma_8\sqrt{\Omega_m/0.3}$ and on the local value of the expansion rate, given in our case by $H_0 \equiv A_0\, H_{P18}(s = 0)\,(1 + \delta H(s = 0))$. A summary of the constraints obtained by the different analyses considered in this work can be found in Table 2.4 with a graphical illustration in Fig 2.4. We primarily focus our discussion on the cosmological parameter Ω_m, the only remaining degree of freedom in Eq. 3.3.

To establish a benchmark against which to compare our method, we start by examining the constraints obtained assuming a ΛCDM model in which we vary H_0, Ω_m and σ_8. This allows us to quantify the impact of performing a model-independent analysis using GPs on the final constraining power. Looking at Fig. 2.4 and Table 2.4 we can observe that, assuming ΛCDM, $\Omega_m = 0.283 \pm 0.007$. In other words, for the setup used in this work, we can constrain Ω_m to around 2% precision if we undertake a model-dependent analysis with ΛCDM. In this case, Ω_m receives information from both background and perturbations.

In turn, our fiducial model-independent analysis yields $\Omega_m = 0.224 \pm 0.066$, inflating the uncertainty by a factor of ~ 9. Comparing this result with the *Planck* 2018, $\Omega_m = 0.317 \pm 0.008$, our Ω_m constraint is lower but statistically compatible with the *Planck* 2018 constraint and our ΛCDM analysis both at 1.5σ.

Looking at the relevance of the different individual data sets on the constraints, we observe an excellent agreement between all the different combinations considered (see Fig. 2.4 and Tab. 2.4). Removing the DSS data, one of the most precise $f\sigma_8$ measurements, significantly worsens the Ω_m constraints by nearly 60%. On the other hand, removing the CMB data set resulted in nearly identical constraints for the cosmological parameters. This is due to the fact that, in the presence of other geometry data to inform the expansion history, Ω_m constraints become dominated by growth data through the relationship between both. Thus, it is fair to ask what the impact of completely removing any type of geometry data from the analysis is. Analytically, we can see from Eq. 1.48 that in the presence of measurements of the linear growth rate and an arbitrary value of Ω_m one can always find a expansion history that solves the differential equation. This degeneracy between the expansion history and Ω_m in Eq. 1.48, prevents our methodology from obtaining any meaningful constraints on Ω_m in the absence of Geometry data that is not completely dominated by our choice of GP mean and hyperparameter priors.

Looking at the parameters σ_8 and S_8 in more detail, we found compatible constraints with the *Planck* 2018 cosmology up to 1σ. However, it is worth mentioning that our results show a mild tendency towards higher σ_8 values which, combined with the tendency towards lower Ω_m values, results in lower S_8 values. This is consistent with the underprediction of $f\sigma_8$ between $0 < z < 0.75$ we discussed in the previous section. A lower value of S_8 would also be consistent with the most recent measurements by large scale structure experiments [54, 60, 61, 135, 172], the origin of which could lie in a lower value of Ω_m [173].

2.4.2 Forecasts

Our model-independent analysis leads to far weaker constraints than assuming the ΛCDM model. Being data-driven, the performance of the method used here may improve significantly with the advent of next-generation surveys with significantly tighter uncertainties. To quantify this, we repeated our fiducial analysis pipeline on mock data generated based on the forecasted errors for the DESI mission [148] in combination with the *Planck* measurement of θ_*. The results can be found in Fig. 2.5. In this figure we can observe a 10% improvement between the expected $\delta H(z)$ constraints from DESI with respect to those of current data. The corresponding η and $\overline{\sigma}(\delta H(s))$ constraints (shown in Tab. 2.3) improve by approximately 15%: $\eta = 0.085 \pm 0.074$ and $\overline{\sigma}(\delta H(s)) = 0.080 \pm 0.004$. Looking at the cosmological parameters, DESI [148] in combination with the CMB data set results in nearly 5 times tighter Ω_m constraints and 10 times better constraints on σ_8.

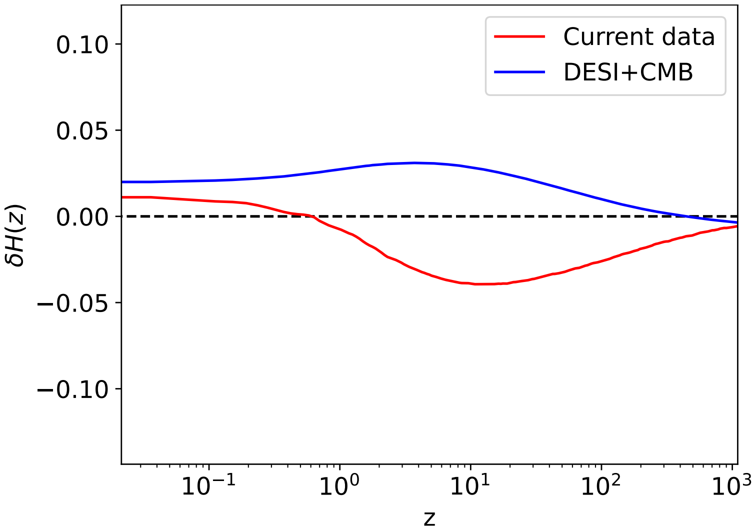

Figure 2.5: Forecast 1σ-constraints on $\delta H(z)$ when using a Gaussian Process on H(z) (the model of this work) to analyse DESI data (blue) and current data (red). Solid lines represent the mean of the GPs at each redshift.

While these constraints are still significantly worse than the *Planck* 2018 results, they are comparable if not better than using the ΛCDM model to analyse the currently available data used in this work. For example, while GP constraints on Ω_m from DESI are two times wider than those currently found assuming a ΛCDM model, the constraints on σ_8 and S_8 are 2.3 and 3.6 times tighter respectively.

If we, instead, compare how a model-independent analysis of DESI data using this methodology will pitch against a ΛCDM analysis, we find that the gap between the model-independent and the model-dependent constraints shrinks. This can be see in the fact that, with current data, the ΛCDM constraint on Ω_m is 9.2 times tighter than the one obtained with our methodology. However, in our DESI forecast, it is only twice as good. Similarly, the improvement for the parameter σ_8 reduces from five times tighter constraints to about four times. Our understanding is that, as better data allow us to better reconstruct $H(z)$, the difference between the reconstructed and model-dependent $H(z)$ shrinks, as long as the assumed model fits the data well. As a consequence, the constraints on the other parameters (Ω_m, σ_8 and S_8) become more similar. Thus, we expect that in the future, as the quality of the data keeps improving, cosmological constraints from model-independent methodologies, such the one proposed in this work, will rival model-dependent ones.

2.5 Conclusions

In this work we have developed a method to obtain constraints on $H(z)$ and Ω_m purely based on the relationship between the expansion history and the linear growth rate. In order to do so, we employ a Gaussian process to model the evolution of the expansion history from present time to recombination.

From this expansion history, we have derived predictions for the comoving distance and the linear growth rate $f\sigma_8$ making use of the physical relationships between the three quantities. Constraints for ΛCDM parameters were obtained by simultaneously fitting a suite of the latest measurements of these three cosmological functions. The data combination used for our fiducial analysis consisted of cosmic chronometers, the Pantheon supernova catalogue, BAO and RSD data, peculiar velocity data from supernovae, and the position of the first acoustic peak in the CMB power spectrum. Moreover, we also obtained forecast constraints on these cosmological functions from the future DESI data.

Current data can constrain the $H(z)$ Gaussian process at an average 9.4% throughout all redshifts. These constraints are compatible up to 1σ with the best-fit *Planck* 2018 expansion history across $0 < z < 1100$. Our constraints on the expansion rate $f\sigma_8$ lie below the corresponding *Planck* prediction by less than 2-σ in the range $0 < z < 0.75$ (a result driven by the Democratic Supernova Sample data).

Translating the Gaussian Process constraints into constraints of cosmological parameters, we find a model-independent measurement of $\Omega_\mathrm{m} = 0.224 \pm 0.066$. This result is lower than, but statistically compatible with, the Planck 2018 [13] cosmology. We also find $S_8 = 0.788 \pm 0.029$, an intermediate value, statistically compatible with both the Planck 2018 cosmology [13] and recent local measurements from weak lensing and galaxy clustering[54, 60, 61, 135, 172].

The forecast analysis performed using the methodology of this work predicts that combining the DESI measurements with the CMB BAO data used in this work will yield 15% tighter constraints on $H(z)$ across redshift as well as five times tighter constraints on Ω_m. Moreover, constraints applying our methodology to future DESI data would achieve five times tighter Ω_m and four times tighter S_8 constraints than when using currently available data. This would make future model-independent constraints on cosmological parameters comparable to current model-dependent constraints. Thus, in the future it will be possible to weigh in on the ongoing S_8 tension making use of model-independent methods. Moreover, we also show that, as the quality of the data increases as we go into the future, the gap between the constraining power of model-independent and model-dependent constraints will significantly shrink.

Future implementations of this methodology could explore several possible extensions. On the one hand, the types of data used in this work are greatly limited by lack of differentiable tools to obtain theoretical predictions for observables with greater constraining power. The development of tools such as differentiable Boltzmann codes, differentiable emulators of the non-linear matter power spectrum, or differentiable Limber integrators would allow us to fit the power spectrum data directly. These developments would enable model-agnostic analyses similar to that of García-García *et al.* [54] but with a greatly reduced number of assumptions and a more reliable measure of uncertainty in their results. On the other hand, it would also be possible to explore the use of Gaussian processes to constrain general forms of modified gravity, generalising works such as Raveri *et al.* [174], Espejo *et al.* [175], Park *et al.* [176], Raveri *et al.* [177], Pogosian *et al.* [178], and study how these theories can be informed by the relationship between background and perturbations. Gaussian Processes are an exceptional tool to constrain modified gravity since they don't require assuming a particular parametrisation of such deviations. A comprehensive list of different departures from ΛCDM that could be explored with a similar methodology to the one presented in this work can be found in Baker *et al.* [179]. Alternatively, one could consider comparing the performance of Gaussian processes against other popular tools for non-parametric cosmology such as genetic algorithms [180, 181] or neuronal networks [182]. The convenience of using Gaussian processes lies in that they naturally provide easily interpretable information on both the reconstructed function and its posterior uncertainties.

Chapter 3

The impact of the Universe's expansion rate on constraints on modified growth of structure

3.1 Introduction

As described in Sect. 1.4, the numerous tensions in modern Cosmology are suggestive of new, undiscovered physics. The growth of structure in the Universe is one of the most sensitive probe of fundamental physics [183, 184]. It is driven by gravitational collapse but is also sensitive to additional forces which may be undetectable on smaller, laboratory scales. It has been shown that measurements of the *rate* of growth of structure can be used to test gravity and constrain, as yet, elusive fifth forces [185].

To be specific, the motion of matter in the Universe can, in general, be subjected to an effective force, $\vec{F}_{\text{eff}}$ of the form:

$$\vec{F}_{\text{eff}} = -\vec{\nabla}\Phi_N - \vec{\nabla}\Phi_5. \tag{3.1}$$

Here, Φ_N is the Newtonian potential and Φ_5 is the potential for a possible long range force that co-exists with gravity on large scales. The properties of Φ_5 may depend on the state of the Universe (its expansion rate, the fractional energy densities of its different constituents) or even on local environmental properties [186, 187]. Thus Φ_5/Φ_N will, generally, be a function of space and time.

If we restrict ourselves to purely long range forces with no environmental dependence, we can define a generalised Newtonian potential, $\Phi \equiv \Phi_5 + \Phi_N$. In an expanding Universe with scale factor, a, Φ satisfies a Newton-Poisson equation on sub-horizon scales (See Sect. 1.1.2),

$$\nabla^2 \delta\Phi = 4\pi G \mu a^2 \bar{\rho} \delta \,, \tag{3.2}$$

where G is Newton's constant, $\bar{\rho}$ is the background energy density of non-relativistic matter, δ is the density contrast. We will assume μ is a function of time only although, in certain scenarios, it can be scale dependent. The relative amplitude of the new force, at any moment in time, is given by $\mu - 1$.

From the linearised Newton-Poisson, continuity and Euler equations one can derive and evolution equation for the growth rate of structure, $f \equiv d \ln \delta / d \ln a$, given by

$$f' + f^2 + \left(1 + \frac{d \ln aH}{d \ln a}\right) f = \frac{3}{2} \mu \Omega_{\mathrm{m}}(a) \,, \tag{3.3}$$

where prime is derivative with regards to $\ln a$, H is the Hubble rate and $\Omega_{\mathrm{m}}(z)$ is the fractional energy density in matter as a function of redshift [128, 179, 183]. Thus, as we can see, the evolution of f depends on μ. This means that, in theory, one can use measurements of the growth rate to constrain the presence of fifth forces.

The situation is, of course, more complex. The evolution of the growth rate depends on the evolution of H and $\Omega_{\mathrm{m}}(a)$. The latter quantity depends, through the Einstein Field Equations, on $H(a)$ so that

$$\Omega_{\mathrm{m}}(a) = \frac{\Omega_{\mathrm{m}}(0) H_0^2}{a^3 H^2} \,. \tag{3.4}$$

Thus, measurements of the growth rate can be used to place constraints on the time evolution of μ and H, and on the fractional matter density today, $\Omega_{\mathrm{m}}(0)$ (for ease of notation, we will now refer to it as Ω_{m} with no argument). But this means that constraints on these various quantities are intertwined and, unless we have independent methods for pinning down H and Ω_{m}, they will hamper our ability to determine μ.

This degeneracy between μ and the expansion history (encapsulated in H, for example) was discussed in Simpson and Peacock [188]. There, it was shown that there is a degeneracy between $\gamma \equiv \partial \ln f / \partial \ln \Omega_{\mathrm{m}}$ and the equation of state of the dark energy component, $w \equiv P_{\mathrm{DE}}/\rho_{\mathrm{DE}}$, where ρ_{DE} (P_{DE}) is the energy density (pressure) of the substance responsible for the accelerated expansion of the Universe at late time (the dark energy). In Baker *et al.* [179], explicit expressions for the degeneracy between μ and w were found using the linear response approach.

Most attempts at constraining $\mu(z)$ have assumed a Universe in which the accelerated expansion at late time is driven by a cosmological constant: the Λ Cold Dark

Matter (ΛCDM) model [13, 107, 189]. A further assumption is that $\mu(z)$ can be modelled in terms of a simple function with one (or at most two) parameters [13]. In a few cases, a more general form for $\mu(z)$ has been assumed with a few independent values at different redshifts (for a notable example see Joudaki *et al.* [107]). Alternatively, model specific time dependencies for $\mu(z)$ have been assumed to arise from theoretical arguments, either from the Effective Field Theory of dark energy [175, 190] or from choices for the underlying model of gravity (such as shift symmetric scalar tensor gravity and its extensions [191]). Most of these attempts at constraining $\mu(z)$ have side-stepped the issue of the degeneracy described above although we highlight Raveri *et al.* [177] in attempting to obtain model-independent constraints.

In this chapter, we will explore how current constraints on μ are affected by our assumptions about the expansion rate of the Universe. In particular, we will see how more or less restrictive assumptions about the parametric form of $H(z)$ impact the uncertainty with which we can determine $\mu(z)$. In the limit in which we do not assume a parameterised form for $H(z)$ we will show that a fundamental degeneracy between $\Omega_{\rm m}$ and $\mu(z)$ manifests itself and, in that regime, we must resign ourselves to constraining the combination $\Omega_{\rm m}\mu(z)$.

The structure of this chapter is as follows. In Section 3.2 we present the main method of this work, the use of a Gaussian Process as a model-independent parameterisation of $\mu(z)$. In Section 3.3 we describe the cosmological observables and the associated data sets which we will use to find the constraints in this chapter. In Section 3.4.1 we present our constraints on $\mu(z)$ and how they depend on what we assume as a model for the background evolution; we will focus on ΛCDM and its extension wCDM, in which we assume an (possibly time varying) arbitrary equation of state, w. In Section 3.4.2 we completely free the background evolution and model $H(z)$ as a Gaussian Process. This gives rise to a strong degeneracy between $\Omega_{\rm m}$ and $\mu(z)$ and we can only constraint $\tilde{\mu}(z) = \Omega_{\rm m}\mu(z)$. In Section 3.5 we discuss both our finding about the role of Gaussian Processes in cosmological analysis and the constraints we have found on $\mu(z)$.

3.2 Methods

3.2.1 A Gaussian Process for $\mu(z)$

The goal of this work is to quantify the uncertainty in our knowledge of $\mu(z)$. The quality of this constraint will depend on both the quality of the data and the assumptions we make about the underlying cosmology through the expansion rate. We want to assume that we have no prior knowledge of the time dependence of $\mu(z)$, apart from the fact that it is relatively smooth. Thus, we choose to model $\mu(z)$ as a Gaussian Process (GP) (See Sect. 1.4).

As opposed to the expansion history, $\mu(z)$ has, a priori, a trivial mean, $\bar{\mu}(z) = 1$, corresponding to the GR value. Thus we only have to concern ourselves with the choice of kernel function. Following the same reasoning as for $H(z)$, we choose a combination of a square exponential kernel (See Eq. 1.157) with a white noise term for numerical stability as the covariance function of the GP:

$$K(\boldsymbol{x}, \boldsymbol{x}') = \eta^2 \exp\left[\frac{-|\boldsymbol{x} - \boldsymbol{x}'|^2}{2l^2}\right] + \sigma\mathbb{I}\,, \tag{3.5}$$

where η is the amplitude of the oscillations around the mean, l is the correlation length between the GP realisations and σ is the noise amplitude chose at 10^{-5}. Thus the final process for $\mu(z)$ is given by:

$$\mu(z) = 1 + \mathbb{L}(K) \cdot \boldsymbol{\nu}\,, \tag{3.6}$$

where the nodes of the GP are distributed as $\boldsymbol{\nu} \sim \mathcal{N}(0, \mathbb{I})$ and $\mathbb{L}(K)$ denotes the lower Cholesky triangle of the GP kernel.

3.2.2 Likelihood

Given a likelihood $\mathcal{L}(\mathbf{y}|\mathbf{x}, \boldsymbol{\sigma})$ for a set of data points $\mathbf{y}$, with a set of errors $\boldsymbol{\sigma}$, and a set of random variables $\mathbf{x}$, a GP can be employed as a prior over all the possible families of functions used to fit the observations. Observations can then be used to inform the GP posterior (i.e. the statistical properties of the assemble of random variables), $\mathcal{P}(g(\mathbf{x})|\mathbf{y}, \boldsymbol{\sigma})$, which determines the family of functions most consistent with the data.

Since we do not have direct measurements of $\mu(z)$, we have to infer it from measurements of the growth rate. However, as one can see from Eq. 3.3, $f\sigma_8$ also depends on $H(z)$ and $\Omega_{\rm m}$. Thus, we must jointly determine $\mu(z)$, $H(z)$ and $\Omega_{\rm m}$ in terms of measurements of $f\sigma_8$ and $H(z)$, or derived quantities such as the comoving ($D_M(z)$), luminosity ($D_L(z)$) or the angular diameter ($D_A(z)$) distances, which relate to $H(z)$ via

$$(1+z)D_A = \frac{D_L}{1+z} = D_M \quad = \quad \int_0^z \frac{dz'}{H(z')} \quad . \tag{3.7}$$

In summary, as we can see from Eqs. 3.3 and 3.7, computing predictions for our observables will involve a non-linear, non-local mapping between the quantities we are interested in ($\mu(z)$, $\Omega_{\rm m}$, $H(z)$...) and the data. For example, a measurement of $f\sigma_8$ at a particular redshift, z, constrains the *history* of μ up until that redshift and not only the value of μ *at* that redshift.

As in Sect. 2.2, the relationship between the GP and the data is once again non-linear meaning that each GP node has to be sampled as an independent parameter (See Sect. 1.4 for details). This means that our model will contain of the order of $\mathcal{O}(10^2)$ parameters. In order to explore this large parameter space, we resort to the gradient-based inference algorithm NUTS described in Sect. 1.2.2.2. We write our likelihood using the `Python` package `PyMC3` [139]. `PyMC3` uses the auto-differentiation (AD) [141, 192] library `Theano` [140] to draw a graph (i.e. a symbolic representation) of statistical models written on it (See Sect. 1.3.2 for a discussion of program graphs and Wegnert tapes). This then allows `Theano` to compute the gradients of the likelihood function required by NUTS cheaply and accurately.

Table 3.1 contains the prior distributions assumed for each for our models. In general, the priors are chosen broad enough to prevent biasing our results. In particular, the priors on the hyperparameters of the GP on $\mu(z)$ (η_μ and l_μ), common in all the studied cases. As discussed in Ruiz-Zapatero *et al.* [1], when using gradient based methods it is best-practice to use smooth priors unless there's physical limit on the values that the parameter can take (e.g. $\Omega_{\rm m} \in [0, 1]$).

Thus, the prior of the amplitude of the GP, η_μ, is a half normal distribution $\mathcal{N}_{1/2}(0, 0.5)$; i.e. centred at 0 with 0.5 standard deviation. On the other hand, the correlation length l_μ has an uniform prior $U(0.01, 6)$. The reason for a uniform prior (i.e. not smooth) is two fold. On the one hand, when sampling l_μ, it is extremely important to avoid small values in order to avoid volume effects (See Eq. 1.157). On the other hand, we do not want to down/up-weight a particular correlation scale for the nodes of GP.

Moving on to the cosmological parameters, only $\Omega_{\rm m}$ has an uniform prior $U(0, 1)$ to enforce the physical limits on the values of the parameter. All the others have normal distributions whose details can be found in Table 3.1. For the cases with a *Planck* 2018 prior, we use the values quoted in Planck Collaboration *et al.* [13]. In particular, for the ΛCDM $_{P18} + \mu_{GP}$ case (second column), we used the TTTEEE+lowE+lensing+BAO ΛCDM constraints (last column of Tab. 2 in Planck Collaboration *et al.* [13]), while for the wCDM $_{P18} + \mu_{GP}$ case (fourth column), we used the TTTEEE+lowE+lensing+BAO+SNe wCDM constraints (first column of Tab. 6 in Planck Collaboration *et al.* [13]). Note that in the wCDM case the constraints also include SNe data which are not present in the ΛCDM constraints. This is because TTTEEE+lowE+lensing+BAO data cannot constrain wCDM models by itself. Note that for both the ΛCDM and the wCDM models we fix $\Omega_R = 9.245 \times 10^{-5}$. Ω_Λ is then derived using $1 = \Omega_\mathrm{m} + \Omega_R + \Omega_\Lambda$.

We must also consider a number of nuisance parameters needed to model the specific data sets chosen for this work. For instance, in order to relate the luminosity curves of the Pantheon data set to luminosity distances one needs to know the value of the absolute magnitude of the supernovae, M. In this work we choose the agnostic way and marginalise over M, assuming a normal prior $\mathcal{N}(-19.2, 1)$, which encompasses both Riess *et al.* [104]'s and Planck Collaboration *et al.* [13]'s H_0 values. On the other hand, we make extensive use of measurements of both parallel and perpendicular BAO measurements. In order to relate these measurements to $H(z)$ and $D_M(z)$ one needs to know the value of the sound horizon at either drag (r_d) or recombination (r^*) epochs. In order to obtain r_d and r^* we use a modified version of the Eisenstein and Hu fitting formula [26, 193] given by

$$r_d \approx \frac{45.5337 \ln\left(7.20376/\Omega_\mathrm{m}\right)}{\sqrt{1 + 9.98592\omega_b^{0.801347}}}\,\mathrm{Mpc}, \tag{3.8}$$

where $\Omega_\mathrm{m} = \Omega_\mathrm{m}(H_0/100)^2$ and $\omega_b = \Omega_b(H_0/100)^2$. Then, noting that the ratio between r_d and r^* can be approximated as a function exclusively of Ω_b, we derive the fitting formula

$$\left(\frac{r_d}{r^*}\right)(\Omega_b) \approx 1.11346 - 2.7985\Omega_b + 16.5111\Omega_b^2\,. \tag{3.9}$$

Hence, combining Eqs. 3.8 and 3.9 we can obtain a prediction for $r^\star$. This approach is capable of reproducing the CLASS ΛCDM predictions for r_d and r^* to an average of 1.5% precision within the considered $\Omega_\mathrm{m} \in [0.1, 0.6]$ and $\Omega_b \in [0.03, 0.07]$. Since the wCDM model we consider doesn't include early dark energy, we can also use Eqs. 3.8 and 3.9 to predict the values of r_d and r^* in such case. On the other hand, when a second GP is used to model $H(z)$, Ω_m is absorbed into the GP on $\mu(z)$ to form $\tilde{\mu}(z)$. This disallows us from following the same approach to obtain r_d and r^* as when assuming a ΛCDM or wCDM model. In this scenario we sample r_d directly as a parameter from $\mathcal{N}(145, 5)$. Then, to get r^* we use Eq. 3.9 as a function of r_d and Ω_b using the same Ω_b as in the ΛCDM and wCDM case.

Finally, in order to interpret the results of the analysis, we will focus on two metrics. At the most basic level, we will study $\mu(z)$ itself and our constraints on its full redshift dependence. We will pay particular attention to $\mu(z = 0)$ since it gives us information on the strength of the fifth force today and can easily be related to other, laboratory or astronomical constraints [184]. In a more abstract level, we will

Table 3.1: Priors used for the different parameters of the models considered in this work. The first column shows the complete list of parameters. U stands for a uniform distribution; $\mathcal{N}(a,b)$ and $\mathcal{N}_{1/2}(a,b)$, for a normal and half-normal distribution, respectively, centred at a and with standard deviation b. Empty entries represent parameters not sampled by the model.

	$H^{\rm P18}_{\Lambda\rm CDM}+\mu_{gp}$	$H_{\Lambda\rm CDM}+\mu_{gp}$	$H^{\rm P18}_{\rm wCDM}+\mu_{gp}$	$H_{\rm wCDM}+\mu_{gp}$	$H_{gp}+\tilde{\mu}_{gp}$
A_0	-	-	-	-	$\mathcal{N}(1.0, 0.2)$
$\eta_{\rm H}$	-	-	-	-	$\mathcal{N}_{1/2}(0, 0.2)$
$l_{\rm H}$	-	-	-	-	$U(0.01, 6)$
η_μ	$\mathcal{N}_{1/2}(0, 0.5)$	$\mathcal{N}_{1/2}(0, 0.5)$	$\mathcal{N}_{1/2}(0, 0.5)$	$\mathcal{N}_{1/2}(0, 0.5)$	$\mathcal{N}_{1/2}(0, 0.5)$
l_μ	$U(0.01, 6)$	$U(0.01, 6)$	$U(0.01, 6)$	$U(0.01, 6)$	$U(0.01, 6)$
$\Omega_{\rm m}$	$\mathcal{N}(0.316, 0.008)$	$U(0, 1)$	$\mathcal{N}(0.307, 0.011)$	$U(0, 1)$	-
$\Omega_{\rm b}$	-	$U(0.03, 0.07)$	-	$U(0.03, 0.07)$	$U(0.03, 0.07)$
H_0	$\mathcal{N}(67.27, 0.6)$	$\mathcal{N}(70, 5)$	$\mathcal{N}(68.31, 0.82)$	$\mathcal{N}(70, 5)$	-
σ_8	$\mathcal{N}(0.811, 0.007)$	$\mathcal{N}(0.8, 0.5)$	$\mathcal{N}(0.82, 0.01)$	$\mathcal{N}(0.8, 0.5)$	$\mathcal{N}(0.8, 0.5)$
w_0	-	-	$\mathcal{N}(-1, 0.5)$	$\mathcal{N}(-0.95, 0.08)$	-
w_a	-	-	$\mathcal{N}(0, 0.5)$	$\mathcal{N}(-0.29, 0.3)$	-
M	-	$\mathcal{N}(-19.2, 1)$	-	$\mathcal{N}(-19.2, 1)$	$\mathcal{N}(-19.2, 1)$
r_d	-	Derived	-	Derived	$\mathcal{N}(150, 5)$
r^*	-	Derived	-	Derived	Derived

Table 3.2: Data sets used in our analysis, listing the probe, the redshift range of the probe, the choice of observable and and the size of the data vector.

Data Set	Probe	Redshift	Observable			Data Points
			$H(z)$	$D_M(z)$	$f\sigma_8$	
CC's [121]	CC	0.07 - 1.965	✓	×	×	33
Pantheon DS17[144]	SNe Ia	0.38 - 0.61	×	✓	×	40
BOSS DR12 [73]	BAO+RSD	0.38 - 0.61	✓	✓	✓	3 × 3
eBOSS DR16 [145]	BAO+RSD	1.48	✓	✓	✓	1 × 3
Wigglez [146]	RSD	0.44 - 0.73	×	×	✓	3
Vipers [194]	RSD	0.60 - 0.86	×	×	✓	2
6dF [195]	RSD	0.067	×	×	✓	1
FastSound [196]	RSD	1.4	×	×	✓	1
DSS [147]	RSD	0	×	×	✓	1
Planck 2018 [13]	CMB	1090.30	×	✓	×	1
DESI [148]	BAO+RSD	0.15 - 1.85	✓	✓	✓	3 × 18

also look at the constraints on the hyperparameter η that describes the amplitude of the covariance matrix of the GP (i.e. the allowed deviation of the nodes from their mean).

3.3 Observables and Data sets

As previously stated, the quality of our data is just as important as our assumptions on $H(z)$ to determine our ability to constrain $\mu(z)$. In this section, we will discuss the data used in this work, as well as how we forecast what future data will be capable of. Let us begin by discussing the currently available data. We employ the same ensemble of data used in Sect. 2.3, as well as additional measurements of $f\sigma_8$. These can be seen in Fig. 3.1 and in the summary Tab. 3.2. The observables and data sets we consider are:

Cosmic Chronometers (CCs) are tracers of dt/dz where t is cosmic time. Since $H(z) \equiv \dot{a}/a = -(dt/dz)/(1+z)$, a measurement of dt/dz directly yields the expansion rate [149]. Here, we use the $H(z)$ measurements from CCs summarised in Table 1 of Vagnozzi *et al.* [121].

Type Ia supernovae (SNe Ia) are explosions of white dwarfs [155, 156], which can be used as standard candles [157, 158]. SNe Ia obey the relationship $m(z) = 5\log D_L(z) + 25 + M$, where $m(z)$ is known as the distance modulus and M is the absolute (apparent) magnitude of the SNe Ia. Knowing M, one can use SNe Ia to reconstruct $D_L(z)$. Here we use a compressed version of the Pantheon sample, known as DS17, composed of 40 measurements of the distance modulus in the range $0.15 \leq z \leq 1.615$ [144]. We marginalise over the absolute magnitude of the supernovae as opposed to fixing its value [159]; this is equivalent to fitting the expansion rate, $E(z) = H(z)/H_0$.

Baryon acoustic oscillations (BAOs) are set by the size of the sound horizon at the end of the drag epoch ($z \sim 1020$), [160–162] $r_\mathrm{s}(z) = \int_z^\infty [c_\mathrm{s}/H(z')]\,\mathrm{d}z'$, where c_s denotes the speed of sound. The BAO feature can be measured in the directions parallel and perpendicular to the line of sight to determine $H(z)$ and $D_M(z)$ respectively. Here we use the twelfth data release of the Baryon Oscillation Spectroscopic Survey (BOSS DR12) which forms part of the Sloan Digital Sky Survey (SDSS) III. In addition to this, we employ the sixteenth data release of the extended Baryon Oscillation Spectroscopic Survey (eBOSS DR16 163), which forms part of the Sloan Digital Sky Survey (SDSS) IV [164]. Finally, we make use of the *Planck* 2018 measurement of the BAO angular scale at $z_* \sim 1100$. We use the *Planck* measurement from

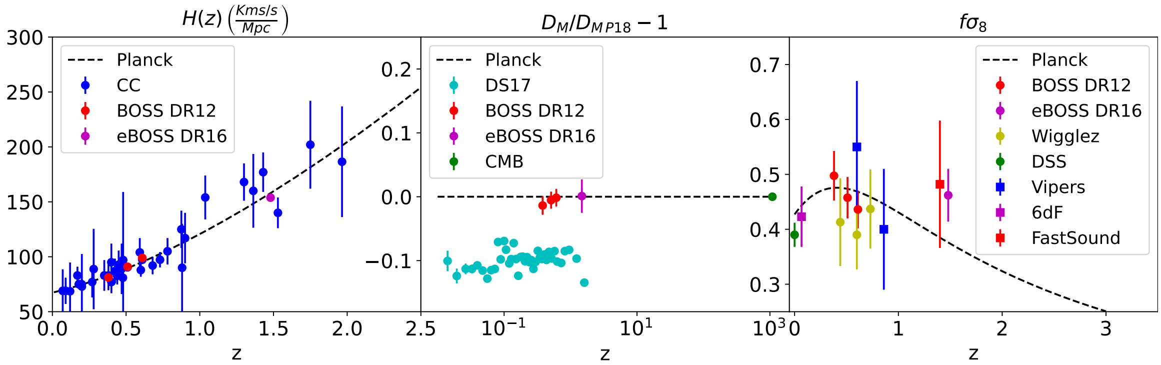

Figure 3.1: Data points from the different surveys used in this work across redshift for the three cosmological functions of interest $H(z)$, D_M and $f\sigma_8$. Radial comoving data is displayed relative to the Planck 2018 mean to show the spread of the data points. Note the addition of extra $f\sigma_8$ data with respect to Fig. 1 2.1.

the temperature, polarisation and lensing maps combined with BAO measurements denoted as TTTEEE+LowE+Lensing+BAO.

Redshift space distortions (RSDs) are modifications to the observed redshift of a given object caused by its radial peculiar velocity [166]. These leave a characteristic anisotropic imprint in the correlation function of galaxies that can be used to measure the growth of structure. Here, we use the three measurements of $f\sigma_8(z)$ from the BOSS DR12 data [73], and one value from the BOSS DR16 quasar sample. We include full covariance matrix between the BAO and RSD measurements from these data sets [73, 145]. We also include the $f\sigma_8$ measurements reported by the WiggleZ Dark Energy Survey [171]. Despite not being RSDs based, we also include the $f\sigma_8(z=0)$ derived from the measured peculiar velocities of the Democratic Samples of Supernovae [147]. In addition to these, we consider three additional RSD based $f\sigma_8$ measurements not included in Sect: 2.3. Namely, the $f\sigma_8$ measurements from the VIMOS Public Extragalactic Redshift Survey (VIPERS), the 6dF Galaxy Survey and the Subaru FMOS galaxy redshift survey (FastSound).

Finally, we are interested in how future surveys will allow us to improve on current measurements. In order to do so, we generate synthetic data based on the forecast errors for The Dark Energy Spectroscopic Instrument (DESI). DESI is currently taking data from the Mayall 4 meter telescope at Kitt Peak National Observatory to construct a galaxy and quasar redshift survey. We use the Font-Ribera *et al.* [148] forecast errors for the observables – $H(z)$, $D_A(z)$, and $f\sigma_8$ – over 18 redshift bins from 0.15 to 1.85. Then, we use the fiducial values of these quantities for the best-fit *Planck* 2018 TTTEEE+LowE+Lensing+BAO ΛCDM cosmology ($\Omega_{\rm M}^{\rm P18}=0.315$, $\Omega_{\Lambda}^{\rm P18}=0.685$, $\Omega_{\rm b}^{\rm P18}=0.049$, $H_0^{\rm P18}=67.36$ and $\sigma_8^{\rm P18}=0.811$) to generate a synthetic data set. In the following sections, we will use this synthetic data to forecast how well a stage IV survey will do in constraining $\mu(z)$ relative to existing data.

3.4 Results

3.4.1 Model-dependent constraints

Having discussed our modelling of $\mu(z)$ and the data we will use to constrain it, we are now at a position to start obtaining constraints for $\mu(z)$. In this section, we will focus on constraints which assume a particular model for the background expansion rate $H(z)$, while modelling $\mu(z)$ as a GP. Please note that we restrict ourselves to models without curvature (i.e. $\Omega_k = 0$). This is motivated by the results of Baker *et al.* [179] where it was shown that the equation of state for the energy component responsible for the accelerated expansion of the Universe would be degenerate with $\mu(z)$.

We start by considering a fiducial expansion rate – the expansion rate given by the *Planck* 18 [13] ΛCDM TTTEEE+LowE+Lensing+BAO posteriors. In this set up, we only make use of our $f\sigma_8$ measurements to constrain our model since we are already using *Planck* 2018's posterior as a constraint on the expansion history. The parameters varied in this set up with their respective priors can be found in the first column of Tab. 3.1. This will give us a best case scenario and will allow us to identify a benchmark against which all other constraints can be compared.

We then relax this assumption, removing the *Planck* prior and freeing up the ΛCDM parameters where

$$H(z) = H_0\sqrt{\Omega_\mathrm{m}(1+z)^3 + \Omega_\mathrm{r}(1+z)^4 + \Omega_\Lambda}\,, \tag{3.10}$$

and Ω_m, Ω_R and Ω_Λ are the cosmological matter, radiation and dark energy densities, respectively, today. We then use the measurements of $H(z)$, $D_M(z)$ and $f\sigma_8$ to constrain these parameters at the same time as we constrain $\mu(z)$. The details of this model can be found in the second column of Tab. 3.1.

In the next study case, to further loosen our assumptions, we chose a background rate of expansion using a general model of dark energy with an equation of state $w(a) = w_0 + w_a(1-a)$ (wCDM). In such model the expression for the expansion rate becomes

$$H(z) = H_0\sqrt{\Omega_\mathrm{m}(1+z)^3 + \Omega_\mathrm{r}(1+z)^4 + \Omega_\Lambda(1+z)^{\nu(z)}}, \tag{3.11}$$

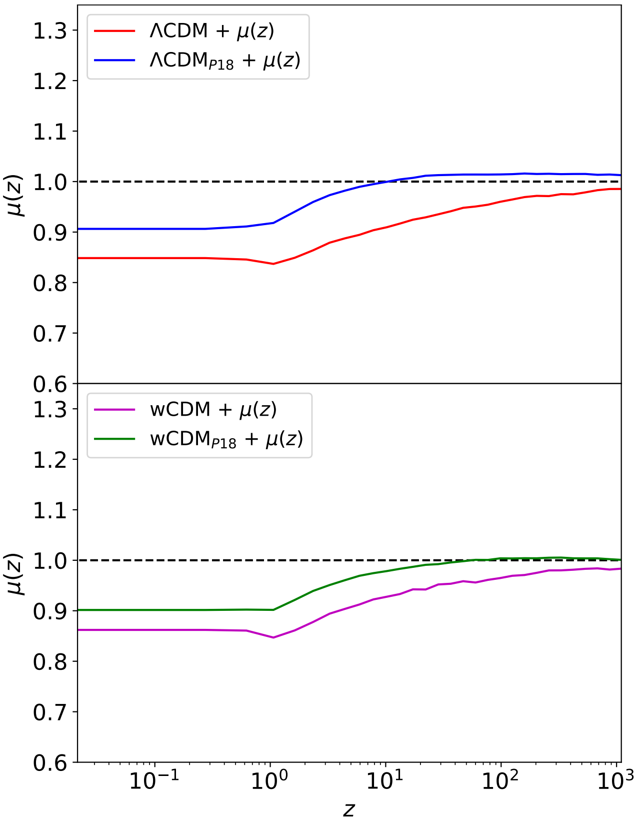

Figure 3.2: Obtained model-dependent constraints on $\mu(z)$. Top panel shows the constraints obtained assuming a ΛCDM model for $H(z)$ both when when using *Planck* 2018's ΛCDM posterior as a prior and when using current late time data to inform it (blue and red respectively). Bottom panel shows the equivalent wCDM constraints (green and purple respectively)

where

$$\nu(z) = \frac{3(1 + w_0 + z(1 + w_0 + w_a))}{1 + z} \,. \tag{3.12}$$

Similarly to the ΛCDM model, we consider two cases. In the first case, we use a fiducial wCDM expansion rate – given by Planck's wCDM TTTEEE+lowE+lensing+BAO+SNe posteriors, where we only use $f\sigma_8$ measurements to constrain our model. We include SNe measurements since the TTTEEE+lowE+lensing+BAO combination considered so far is not able to place tight constraints on the equation of state on its own. The details for this model can be found in the third column of Tab. 3.1. In the second case, we free the expansion rate parameters (including w_0 and w_a) and use our whole suite of measurements to inform our constraints. The details of this model can found in the fourth column of Tab. 3.1.

We find that regardless the model assumptions made (ΛCDM or wCDM), $\mu(z)$ is in excellent statistical agreement with the GR value $\mu(z) = 1$ at all redshifts up to

1σ. We find the same consistency with GR when using the *Planck* 18 prior on the cosmological parameters (including w_0 and w_a in the wCDM case) and when freeing them. Fig. 3.2 shows the constraints obtained on $\mu(z)$ in both cases, with the constraints obtained assuming ΛCDM on shown in the top panel and those assuming wCDM in the bottom panel. In both panels, we compare the contours obtained using the *Planck* 2018 posterior as a prior in combination with our $f\sigma_8$ measurements, and by using our whole suite of measurements to inform our constraints. We can see that imposing the *Planck* 2018 prior significantly reduces the uncertainty on $\mu(z)$ at all redshifts. More quantitatively (see Tab. 3.3), the uncertainty on $\mu_0 \equiv \mu(z=0)$ decreases by roughly $\sim 35\%$ for both a ΛCDM or wCDM cosmology and, remarkably, the uncertainty in μ_0 remains unchanged when using the more complex wCDM background model. Thus, we can conclude that the combination of cosmic chronometers, BAO and SNe data are sufficiently precise to pin down the equation of state for the purpose of constraining μ_0.

Finally, regardless of the model assumed or how the expansion history was constrained, we observe that the uncertainty in $\mu(z)$ blows up around $z \sim 2$. This is because this is the point around which our data on $f\sigma_8$ ends. The reason why the GP on μ carries some constraining power beyond the redshift of our last $f\sigma_8$ data point ($z = 1.48$) is two-fold. On the one hand, the GP has a given correlation length. In other words, the GP doesn't immediately decohere and suddenly return to its prior after the data stops. Second, in order to solve for $f\sigma_8$ we have to solve a differential equation on $\mu(z)$, which involves integrating over the later quantity. This means that points in $f\sigma_8$ do no precisely match to a singular point in $\mu(z)$.

It is interesting to understand this result in light of the discussion in Baker *et al.* [179]. There, it was shown that, while a measurement of $f\sigma_8$ at one redshift would lead to a severe degeneracy between μ and w, measurements at multiple redshifts combined with distance measurements could, in principle, break this degeneracy and decorrelate constraints between the two parameters. However, as we have seen in Fig. 3.2, current data is capable of breaking this degeneracy. This can be better appreciated in Fig. 3.3. In this figure, we show the 1D and 2D distributions for the parameters w_0, w_a and μ_0. We superpose the contours obtained when using the *Planck* 2018 prior (blue) and when only using current data to constrain the wCDM parameters (red). As we can observe, the current data contours show a degeneracy between w_0-w_a which is not present when using the *Planck* 2018 prior. However, neither w_0 nor w_a are degenerated with μ_0 in any case.

Table 3.3: Model-dependent constraints on Ω_m, σ_8 and μ_0, reporting the mean value and the 1σ errors.

	Ω_m	σ_8	μ_0
ΛCDM	0.302 ± 0.007	0.789 ± 0.027	-
wCDM	0.292 ± 0.013	0.801 ± 0.034	-
$\mu(z) + \Lambda\mathrm{CDM}_{P18}$	0.314 ± 0.007	0.811 ± 0.006	0.904 ± 0.123
$\mu(z) + w\mathrm{CDM}_{P18}$	0.306 ± 0.008	0.821 ± 0.014	0.899 ± 0.123
$\mu(z) + \Lambda$CDM	0.302 ± 0.007	0.878 ± 0.127	0.850 ± 0.191
$\mu(z) + w$CDM	0.29 ± 0.016	0.887 ± 0.127	0.862 ± 0.190

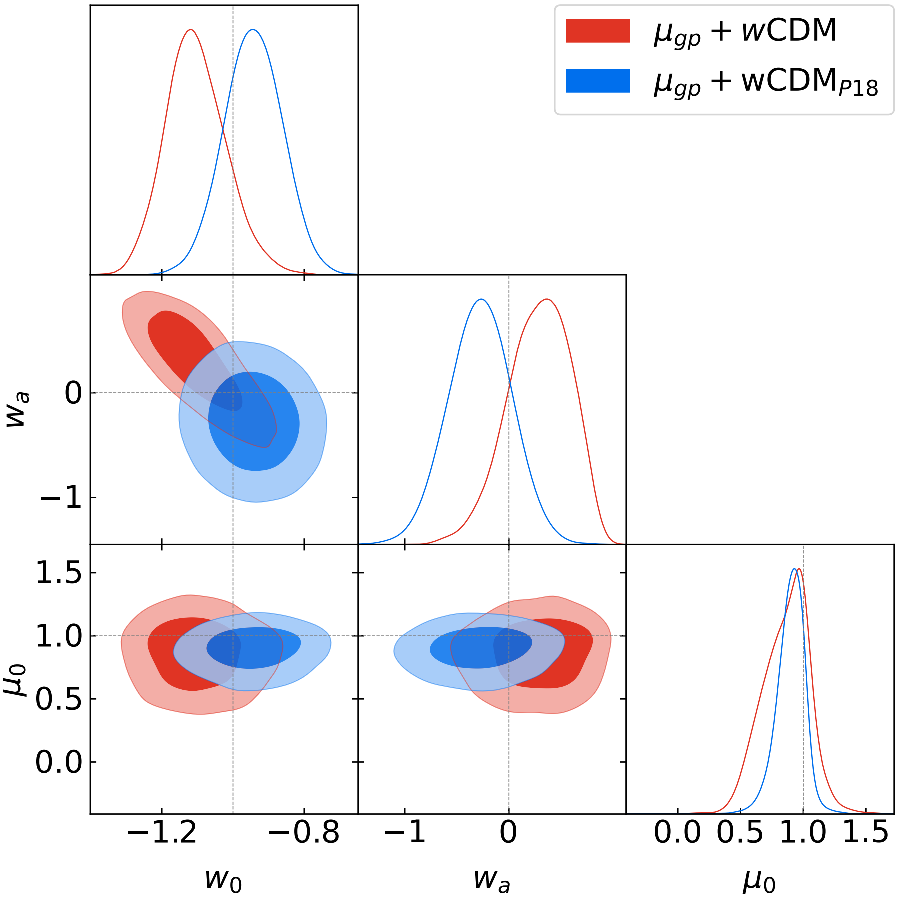

Figure 3.3: Constraints for the cosmological parameters w_0, w_a and μ_0. Diagonal panels show 1D distributions. Off-diagonal panels show 2D distributions. In each panel we superpose the contours obtained when assuming *Planck* 2018's wCDM posterior as a prior (blue) and when marginalising over a wCDM background (red) given current late time data.

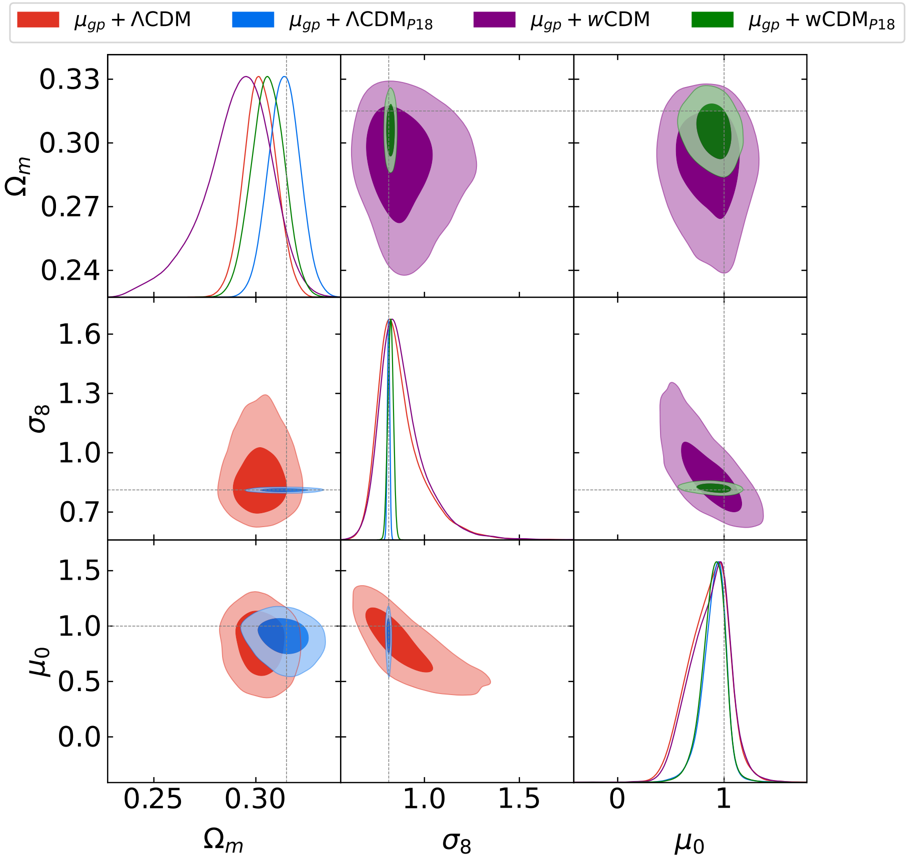

Figure 3.4: Constraints for the cosmological parameters Ω_m, σ_8 and μ_0. Diagonal panels show 1D distributions. Off-diagonal panels show 2D distributions. Bottom triangle shows the constraints obtained when assuming a ΛCDM background both when imposing the *Planck* 2018's ΛCDM posterior as a prior (blue) and when using current late time data to inform it (red). The top triangle shows the equivalent constraints when a wCDM background was assumed instead (green and purple respectively).

We further note that the uncertainty in $\mu(z)$ increases as we look at higher redshifts, but not excessively so. Two factors are at play here. First, since the data are a non-local function of $\mu(z)$ (i.e. $\mu(z)$ needs to be integrated to solve for f in Eq. 3.3), they allow us to place constraints on higher redshift values of $\mu(z)$. In addition to this, we are marginalising over the hyperparameters of the Gaussian Process. This means that the data at lower redshifts can put a constraint on the amplitude and correlation length of the GP's kernel. This effectively limits the variance of the GP even in regions with no data.

We have seen that assuming a wCDM for $H(z)$ as opposed to ΛCDM model does not degrade our constraints on $\mu(z)$. It is then interesting to explore the relationship between $\mu(z)$ and other cosmological parameters of our models, particularly Ω_{m} and σ_8. Fig. 3.4 shows the 1D and 2D contours for the parameters Ω_{m}, σ_8 and μ_0 obtained when assuming the ΛCDM and wCDM models to parameterise $H(z)$. In each panel, we superpose the results obtained when assuming *Planck* 2018's posterior as a prior for the expansion rate as opposed to letting background data inform the constraints. We show the associated numerical constraints in Tab. 3.3. We also display the constraints obtained by fitting a ΛCDM and wCDM model while keeping $\mu(z) = 1$ (i.e. GR) for context.

Looking at Eq. 3.3 one would expect a great degeneracy between Ω_{m} and $\mu(z)$. However, if we look at the bottom left corner panel (ΛCDM) and top right panel (wCDM) of Fig. 3.4 we can see how information about the background breaks this degeneracy. Therefore, it is not clear that a better constraint on one will lead to an improvement on the other.

We show our constraints on Ω_{m} for the different models in Fig. 3.5, including constraints for the ΛCDM and wCDM models when keeping $\mu(z) = 1$ (i.e. GR) for reference. Regardless of whether we assume a ΛCDM or wCDM model for $H(z)$ we obtain a slightly lower value for Ω_{m} than the one obtained by *Planck* 2018 (and the one obtained using *Planck* 2018's posterior as a prior). Nonetheless, once the size of the error bars is taken into account, the constraints are in reasonable statistical agreement (less than 1.5 σ tension). Moreover, assuming wCDM systematically results in a lower yet statistically compatible constraint of Ω_{m} than assuming a ΛCDM model. Finally, it is interesting to note that introducing $\mu(z)$ barely degrades the constraint on Ω_{m} if the background is ΛCDM. On the contrary, for a wCDM model, introducing $\mu(z)$ leads to a $\sim 20\%$ larger error bar on Ω_{m}. This is caused by the fact that freeing the equation of state reduces the ability of the background to constrain Ω_{m} and, thus, the wCDM Ω_{m} constraint increasingly depend on the growth data to inform its value.

Moving to σ_8, current growth data cannot break the degeneracy between σ_8 and μ_0. This can be seen in the middle panel of the bottom row and the right panel of the middle row. Therefore, when assuming a model for $H(z)$, the bottleneck in constraining μ_0 is how well we know σ_8. This explains why our constraints on $\mu(z)$ drastically improve when imposing the *Planck* 2018 prior since it imposes a much tighter constraint on σ_8, breaking the degeneracy with $\mu(z)$.

3.4.2 Model-independent constraints

We now proceed to further relax our assumptions about the background expansion rate by promoting $H(z)$ to a GP. We do so by following the methodology developed

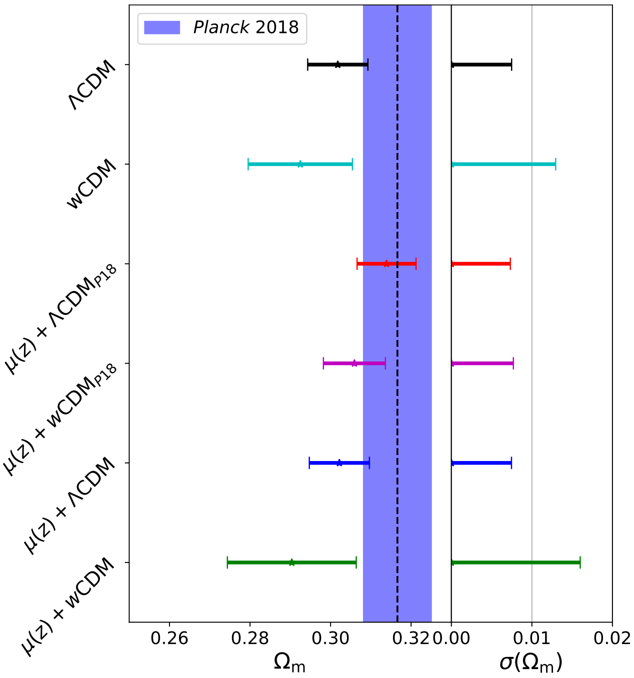

Figure 3.5: Constraints obtained for Ω_m for each model considered in this work. Side panel shows the uncertainty of each constraint.

in Ruiz-Zapatero *et al.* [1]. More specifically, we model $H(z)$ as

$$H(z) = A_0 H^{\mathrm{P18}}(z)(1 + \delta H_{gp}) \,, \tag{3.13}$$

where A_0 is a free parameter, $H^{\mathrm{P18}}(z)$ is the Hubble rate for our ΛCDM*Planck* 18 best-fit fiducial cosmology (see Sect. 3.2), and δH_{gp} is a relative deviation that we model as a Gaussian Process. This is a Bayesian approach to GP's in which one marginalises simultaneously over the GP itself and its mean. In Ruiz-Zapatero *et al.* [1] we showed that this approach shields our cosmological constraints from potentials biases introduced by our choice of mean function. More recently, Hwang *et al.* [197] showed that unphysical oscillations can appear in the reconstructed functions if one does not marginalise over a possible family of mean functions for the GP. However, it is worth noting that the degeneracies between the GP and the $A0$ parameter make exploring the parameter space significantly more expensive.

Therefore, our inference process now involves two GPs. This allows us to measure the degeneracy between modifications of the expansion history and the Poisson equation in the prediction of $f\sigma_8$ without having to assume a particular model.

However, becoming fully model-independent comes at the cost of no-longer being able to constrain Ω_m with measurements of background quantities. This is because

$H(z)$ is no longer a function of cosmological parameters. Thus, we have no independent way of constraining Ω_m apart from the relationship between $H(z)$ and $f\sigma_8(z)$. Revisiting Eq. 3.3, we can also see that we are now faced with an unbreakable degeneracy between Ω_m and $\mu(z)$. In order to deal with this degeneracy, in this section we consider the new, combined parameter

$$\tilde{\mu}(z) = \frac{\Omega_\mathrm{m}}{\Omega_M^\mathrm{P18}}\mu(z)\,, \tag{3.14}$$

where Ω_M^P18 is the *Planck* 18 TTTEEE+LowE+Lensing+BAO, ΛCDM best fit value of Ω_m.

In order to solve Eqs. 3.3 and 3.7 when considering two GPs, we employ the same combination of numerical methods as in Ruiz-Zapatero *et al.* [1] (where we also modelled $H(z)$ as a GP), albeit with some modification. In Ruiz-Zapatero *et al.* [1] we assigned a node of the GP to each node of the numerical grid used to solve the growth equation and the comoving distance integral. This approach becomes very computationally expensive when we introduce a second GP. In order to make our model more computationally efficient, we decouple the number of nodes in the numerical integration schemes from the number of nodes used for each GP, linearly interpolating where necessary. This allows us to significantly reduce the number of parameters of the model while preserving the necessary numerical accuracy. Reducing the number of nodes in the GP's means that the degeneracy between the remaining nodes is reduced. This latter aspect is particularly helpful when using HMC which is most efficient when the parameters are as uncorrelated as possible. The end result of reducing the number of parameters and the degeneracy between them is a substantial speed-up in the time needed for the sampler to converge.

We show the obtained model-independent constraints for $\tilde{\mu}(z)$ in Fig. 3.6. On the one hand, in the top panel of the figure, we can observe that the model-independent constraints on $\tilde{\mu}(z)$ are only marginally worse than the model-dependent constraints on $\mu(z)$ ($5\% - 10\%$ depending on whether we consider the ΛCDM or wCDM model). This means that, even when completely relaxing our assumptions about $H(z)$, current data have enough constraining power to break the degeneracy between $H(z)$ and $\tilde{\mu}(z)$. This can be further seen in the correlation matrix between the GP's nodes of $\tilde{\mu}(z)$ and $H(z)$. Fig. 3.7 shows that, although $\mu(z)$ and $H(z)$ nodes have a great degree of auto-correlation (as expected for a GP), the correlation coefficients between both quantities are never larger than 5%. This can be seen as a generalisation of the lack of correlation we observed between the background parameters and μ_0 in Sect. 3.4.1.

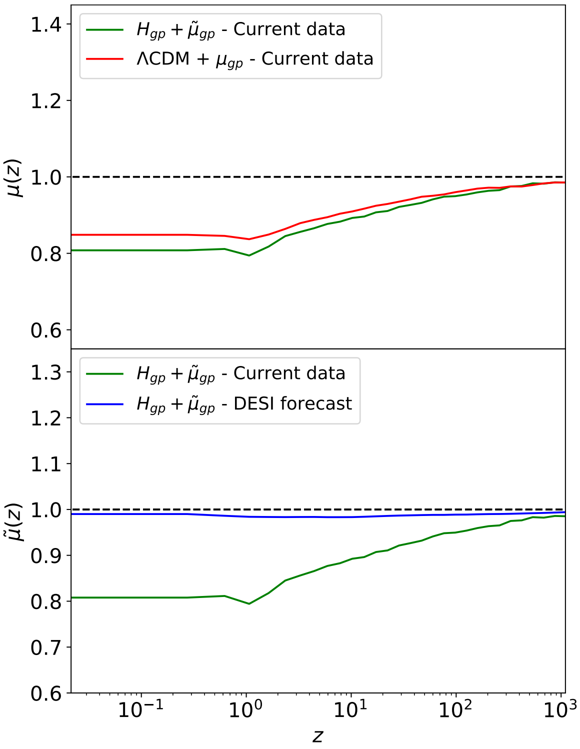

Figure 3.6: Top panel: constraints on $\mu(z)$ for current data when assuming the ΛCDM model to model background expansion of the Universe (red) and when using a second GP (green). Note that when using a second GP the quantity being constrained is $\tilde{\mu}(z)$ as opposed to $\mu(z)$. Bottom panel: constraints obtained on $\tilde{\mu}(z)$ when using a second GP to model $H(z)$ for both current data (green) and mock DESI data (blue).

Moreover, we can see that $H(z)$'s low redshift nodes are much less correlated with high redshift nodes than those of $\tilde{\mu(z)}$.

It is important to bear in mind that these are constraints on $\tilde{\mu}(z)$, not on $\mu(z)$. Converting constraints on $\tilde{\mu}(z)$ into constraints on $\mu(z)$ requires a measurement of Ω_{m}. However, in the process of freeing $H(z)$ we have lost all of our knowledge of Ω_{m}. Thus, an external, model-independent measurement of Ω_{m} would be needed to transform $\tilde{\mu}(z)$ constraints into $\mu(z)$ constraints. The constraints on $\tilde{\mu}(z)$ should therefore be understood as the most optimistic model-independent constraint on $\mu(z)$ possible given current data; i.e. the case for which we have a perfect model-independent measurement of Ω_{m}.

Finally, we find $\sigma_8 = 0.886 \pm 0.138$ when a second GP is used to model $H(z)$. This means that not assuming a ΛCDM or wCDM model for the expansion history degrades our σ_8 constraint by around $\sim 10\%$. Nonetheless, the degree of correlation between σ_8 and $\tilde{\mu}_0$ remains virtually identical to that of model-dependent analyses. Thus, model-independent constraints on $\tilde{\mu}(z)$ will also benefit greatly from ways of

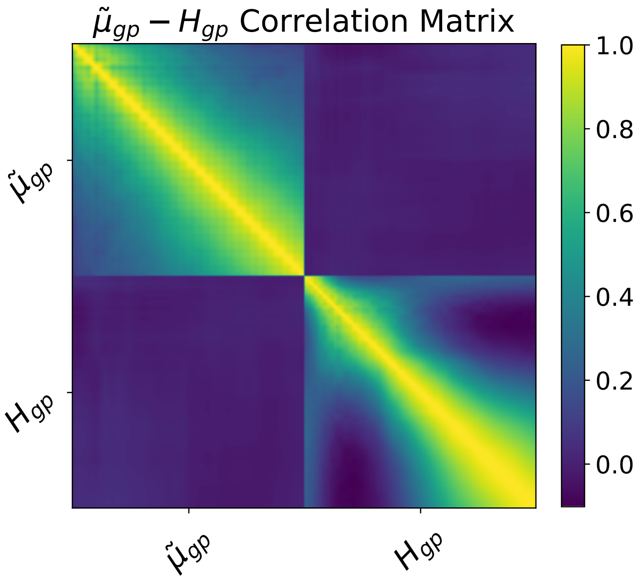

Figure 3.7: Correlation coefficients between the nodes of the GP on $H(z)$ and the GP on $\mu(z)$. This plot can be seen as a generalisation of Fig. 3.3, showing that the expansion rate and the modifications of the linear growth are already independent with the constrain level of current data.

tightening their constraint on σ_8, just as we saw in the model-dependent case. We will discuss this further in the next section when considering our analysis of mock DESI data.

3.5 Conclusions

In this chapter, we have assessed the importance of our current knowledge of the expansion rate history on our ability to constrain $\mu(z)$ in a model-independent manner. As was argued in Simpson and Peacock [188] and Baker *et al.* [179], the assumptions that go into the modelling the Hubble rate as a function of redshift, $H(z)$ will impact constraints on $\mu(z)$ from the growth rate of structure. It was shown that the more conservative (or looser) the model for $H(z)$, the weaker the constraints on $\mu(z)$ should be.

We have found that, however, current constraints on the expansion rate from cosmic chronometers, supernovae and BAO data are sufficiently tight that the assumptions made about the underlying background model are not important when constraining $\mu(z)$. To show this, we have used a completely general form for $\mu(z)$ (a Gaussian Process), and quantified whether assuming a simple equation of state for Dark Energy ($w = -1$), or a more general equation of state of the form $w = w_0 + w_a(1 - a)$

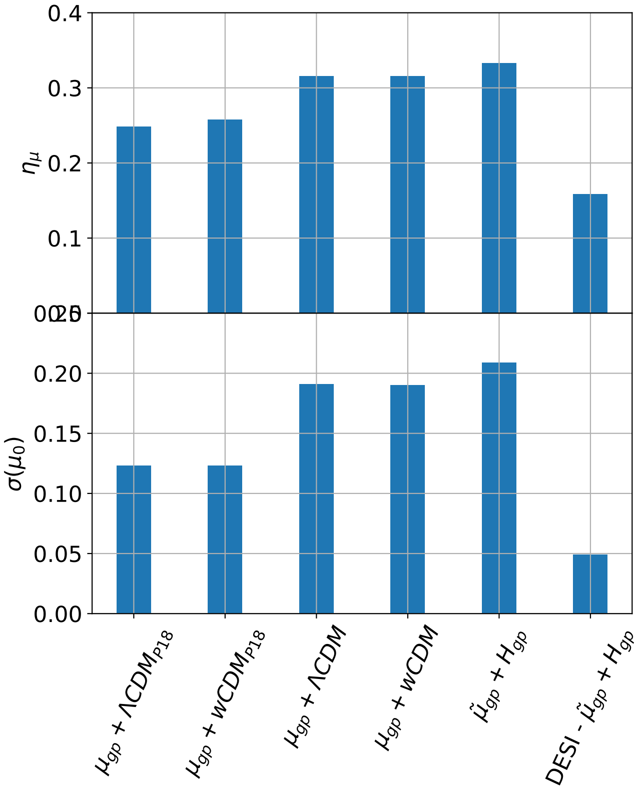

Figure 3.8: Comparison of different measures of uncertainty in $\mu(z)$ between the different models considered in this work. The top panel shows the mean value of the amplitude of the covariance matrix of $\mu(z)$ for each model. The bottom panel shows the uncertainty in $\mu(z = 0) \equiv \mu_0$. Note that when a second GP is considered to model $H(z)$ (i.e. last two entries) $\tilde{\mu}(z)$ is shown as opposed to $\mu(z)$.

affects the final constraints on $\mu(z)$. We also considered a completely general form for $H(z)$ which we also modelled as a Gaussian Process. In this case, we are faced with a fundamental degeneracy between $\mu(z)$ and Ω_m and thus, we present our results in terms of $\tilde{\mu}(z) = \Omega_\mathrm{m}\mu(z)/\Omega_M^{P18}$ where we recall that Ω_M^{P18} is the best fit value of Ω_m for the *Planck* 2018 TTTEEE+LowE+Lensing+BAO analysis of the ΛCDM model.

As discussed in Section 3.2, we summarise our results on the constraints on $\mu(z)$ using two statistics. On the one hand, we look at the uncertainty in $\mu_0 \equiv \mu(z = 0)$ as it directly relates to the strength of any possible fifth force today. On the other hand, we consider the mean value of the amplitude of the Gaussian process covariance matrix, η_μ, which is an abstract measurement of the uncertainty of the Gaussian process through its whole domain.

We present the corresponding results in Fig. 3.8. Reassuringly, we find that the two statistics offer us the same picture: the less assumptions we make on the expansion history, the more uncertainty on $\mu(z)$. However, it is extremely important to stress that the loss in constraining power is marginal. Comparing assuming a ΛCDM vs wCDM model, we find that it makes effectively no difference and there is no degradation in our constraints on $\mu(z)$. Even when a second GP is used to model $H(z)$ the constraint is only a few percentages larger.

Focusing on μ_0, we find that $\sigma(\mu_0) \simeq 0.12$ if we assume *Planck* 2018's posterior as a prior, for either the ΛCDM or wCDM model. This uncertainty increases to $\sigma(\mu_0) \simeq 0.19$ if instead of imposing *Planck* 2018's posterior as a prior we use our collection of late time $H(z)$, $D_M(z)$ and $f\sigma_8$ measurements to inform our constraints. The main difference between assuming *Planck* 2018 posteriors and using late-time data to inform our models is that the former provides us with a much tighter constraint on σ_8, the main bottle-neck when constraining $\mu(z)$ in a model-dependent fashion. Looking at the model-independent constraint, we find that $\sigma(\tilde{\mu}(z)) \simeq 0.21$.

If we instead look at the constraints on η_μ, we find the exact same trend as in μ_0. While one would expect the two statistics to agree, μ_0 only probes the GP at $z = 0$ while η_μ contains information about the whole GP domain. We find that for our best-case scenario, in which we assume *Planck* 2018's ΛCDM background, $\eta_\mu = 0.25$. Letting late time data inform a ΛCDM model instead returns $\eta_\mu = 0.32$. Furthermore, if we assume a wCDM model, we find $\eta_\mu = 0.26$ when using the *Planck*'s posteriors to pin it and $\eta_\mu = 0.32$ when letting late time data inform it. Finally, we find $\eta_\mu = 0.33$ in the model-independent case.

The fact that constraints on μ are (relatively) insensitive to our parameterisation of $H(z)$ is not unexpected. This is because current background data is powerful enough to constrain $H(z)$ independently of the assumptions made. In the analysis of Ruiz-Zapatero *et al.* [1], we found that constraints on Ω_m from the growth rate were not strongly dependent on our modelling choices of the Gaussian process on $H(z)$.

There have been other attempts at constraining $\mu(z)$. In Planck Collaboration *et al.* [13] an uncertainty of $\sigma(\mu_0) \simeq 0.25$ was found under the assumption that μ evolves as $\mu(z) - 1 \propto [1 - \Omega_\mathrm{m}(z)]$. However, different assumptions about the specific time dependence of μ (e.g. $\mu(z) \propto a^n$) lead to constraints that are strongly dependent on the choice of n[189], with μ_0 in the range $\sigma(\mu_0) \in (0.04, 1.5)$. Assuming that the modified Poisson equations arises from scalar-tensor theories, one can use the tools of Effective Field Theory [177] or simply assume specific classes of models [191] to obtain $\sigma(\mu_0) \simeq 0.25$. As we can see, our methodology returns stronger constraints with $\sigma(\mu_0) \in (0.12, 0.19)$ depending on the strength of the assumptions made on $H(z)$. We note however that it can be misleading to directly compare $\sigma(\mu_0)$ as they can be heavily dependent on the underlying model and choice of data sets one is using.

It is instructive to see how much our constraints will improve with future data. As an example, we choose the specifications for the DESI data set, described in Sect. 3.3, and combine it with the *Planck* 2018 CMB BAO measurement to pin down the GP on $H(z)$ at high redshift. Our analysis of DESI mock data shows that we will obtain constraints on $\tilde{\mu}(z)$ (i.e. with a model-independent $H(z)$) which are twice as tight as with current data when assuming either a ΛCDM or wCDM background. This is in spite of the DESI constraint on σ_8 being about six times wider than *Planck* 2018's. The reason behind this improvement in constraining power boils down to the fact that DESI alone will offer nearly twice as many measurements on $f\sigma_8$ as the number considered in this work over a larger redshift window. Moreover, DESI $f\sigma_8$ measurement will have significantly smaller errors bars than currently available ones. It is particular important to focus on the smaller size of said error bars relative to the expected dynamic range of $f\sigma_8$ in the redshift window probed. This will greatly help break the degeneracy between the amplitude of $f\sigma_8$ (given by σ_8) and its shape (given by $\mu(z)$ in the presence of background data to pin down Ω_m) present in current data.

Finally, there are several avenues through which the results and methodology presented here could be further explored. One can ask the question: how well do we need to measure Ω_m to obtain a competitive model-independent constraint on $\mu(z)$ with current data. Using propagation of errors; $\sigma(\mu)/\mu = \sqrt{(\sigma(\tilde{\mu})/\tilde{\mu})^2 + (\sigma(\Omega_\mathrm{m})/\Omega_\mathrm{m})^2}$, we find that model-independent measurement of Ω_m to 10% precision would be enough to match model-independent constraints on $\mu(z)$ to model-dependent constraints with current data. Similarly, a percentage model-independent measurement of Ω_m would allow us to constrain $\mu(z)$ to virtually the same precision as $\tilde{\mu}(z)$. This measurement of Ω_m would need to be independent from the model assumed for the background expansion and for the parameterisation of the Poisson equation. Future works could attempt to obtain an alternative model-independent constraint on Ω_m to break the $\tilde{\mu}(z) - \Omega_\mathrm{m}$ degeneracy found in this methodology.

Chapter 4

LimberJack: Auto-differentiable methods for angular power spectra analyses

4.1 Introduction

As discussed in Sect. 1.1.5, Cosmology is currently experiencing an unprecedented increase in the quantity and quality of data. The Dark Energy Spectroscopic Instrument (DESI) [198] has already recorded more galaxy spectra in its first two years of operations than previously done in the whole history of humanity. Moreover, next-generation surveys such as Legacy Survey of Space and Time (LSST) [199], Euclid [200], the Nancy Grace Roman space telescope [201], the Simons Observatory [202], CMB-HD [203] or CMB-S4 [204] promise to further accelerate this trend in the next decade.

In order to match the quality of the data, physicists are starting to incorporate into their theoretical predictions more of the physical, observational and instrumental effects which, until now, could be overlooked. In practice, this translates into a dramatic increase in the number of parameters that future analyses will have to consider. This combination of large data sets with complex models will (and in many cases already does) overwhelm the inference methods we currently use to constrain the values of these parameters.

In a Bayesian framework (See Sect. 1.2), the statistical distribution of a set of parameters, $\boldsymbol{\theta}$, given some data, $\boldsymbol{d}$, is given by Bayes theorem:

$$\mathcal{P}(\boldsymbol{\theta}|\boldsymbol{d}) \propto \mathcal{L}(\boldsymbol{d}|\boldsymbol{\theta})\Pi(\boldsymbol{\theta})\,, \tag{4.1}$$

where $P(\boldsymbol{\theta}|\boldsymbol{d})$ is the posterior distribution of the parameters given the data, $\mathcal{L}(\boldsymbol{d}|\boldsymbol{\theta})$ is the likelihood of the data for a set of parameters and $\Pi(\boldsymbol{\theta})$ are the prior beliefs on the distribution of the parameters. Current cosmological analyses explore their parameter spaces by mapping out $P(\boldsymbol{\theta}|\boldsymbol{d})$ according to some stochastic process by which the direction of exploration is chosen. Despite the success of this methodology, relying on stochastic methods becomes inefficient at high dimensions. This effect is known as the *curse of dimensionality* (See Sect. 1.2.1).

One of the most effective ways of overcoming the curse of dimensionality is using the gradient of the posterior; $\nabla P(\boldsymbol{\theta}|\boldsymbol{d}) \propto \nabla(\mathcal{L}(\boldsymbol{d}|\boldsymbol{\theta})\,\Pi(\boldsymbol{\theta}))$, to guide exploration towards regions of interest in parameter space. Algorithms that use the gradient of the likelihood to explore parameter space are known as gradient-based samplers [76–79]. Unfortunately, traditional methods of finding numerical gradients, such as the finite-difference method, scale poorly with the number of dimensions. Moreover, they can be prone to numerical instabilities. This can render the use of gradient information counter-productive.

Thankfully, in the last decades a series of algorithms known as auto-differentiation (AD) [94, 141, 192] have grown in popularity. Given a generic computer program that maps a series of inputs to a series of outputs (i.e. a function inside a computer), AD is a family of algorithms designed to produce a symbolic representation of such computer program such that the chain rule can be systematically applied to produce a second program for the gradient of the original function. Unlike finite differences, AD is not subject to truncation errors and its computational cost scales much more favourably with the dimensionality of the original function [94]. Thus, the goal of this chapter is to make the theoretical predictions described in Sect. 1.1 compatible with the AD methods described in Sect. 1.3.2 to enable the gradient-based inference algorithms described in Sect. 1.2.2.

We present `LimberJack.jl`, an auto-differentiable angular power spectra analysis code fully written in `Julia`. `LimberJack.jl` is designed after the Core Cosmology Library [CCL 205] developed by the LSST Dark Energy Science Collaboration, aiming to fulfil the similar scientific goals. `LimberJack.jl` allows the user to easily compute ΛCDM model predictions for the angular power spectra of weak lensing, CMB lensing and clustering surveys using the Limber approximation [31, 32]. Most importantly, `LimberJack.jl` can also provide the user with accurate gradients of these predictions in a computationally efficient way, due to its compatibility with `Julia`'s AD libraries `ForwardDiff.jl` and `ReverseDiff.jl`. While `LimberJack.jl` is currently more limited than `CCL`, its modular design means that it can easily be extended by the community. This will allow the community to add new features that will be necessary for the analysis of future data which are currently not present in `LimberJack.jl`;

such as more precise prescriptions of the non-linear corrections to the matter power spectrum or going beyond the Limber approximation [206].

A number of auto-differentiable cosmological codes have been presented in the literature. The most notable precedents are `JAX-COSMO` [207] and `CosmoPower-JAX` [208]. `JAX-COSMO`'s functionalities and goals are very similar to those of `LimberJack.jl` both basing their design on `CCL`. `CosmoPower-JAX` is a neural-network emulating framework for the matter power spectrum originally developed in `TensorFlow` [209] and later ported to `JAX` to be paired with `JAX-COSMO`. The main difference between `JAX-COSMO`, `CosmoPower-JAX` and `LimberJack.jl` is the AD ecosystem they make use of. Both `JAX-COSMO` and `CosmoPower-JAX` are written in `JAX`, a scripting programming language which interfaces with a lower-level language, `XLA`, a just-in-time (JIT) compiled language with AD capabilities. `JAX`'s main strengths are its powerful parallelisation schemes on both CPUs and GPUs, its performant AD methods and the fact that it shares API with the ubiquitous `Python` library `NumPy`. On the other hand, `LimberJack.jl` is fully written in `Julia`, a general-purpose, JIT-compiled, programming language with native AD capabilities. `Julia`'s main advantage is the lack of a lower-level programming language meaning that `Julia` can directly generate its machine code. This is not the case in other popular AD environments such as `TensorFlow` and `JAX` that interface with `C++` and `XLA` respectively to generate their machine code. This transparency makes `Julia` an excellent language to develop complex libraries with customised methods. A recent practical example where this feature has played a key role is the development of the first AD Boltzmann solver, `Bolt.jl` [210], a challenging task for `JAX` and `TensorFlow` but which, nonetheless, was possible in `Julia`.

The structure of this chapter is as follows: in Sect. 4.2 we describe the theoretical predictions computed by `LimberJack.jl`, how `LimberJack.jl` computes these predictions in an auto-differentiable way and how AD can be used to speed up statistical inference. In Sect. 4.3, we use `LimberJack.jl` to reproduce the DES Y1 3x2-pt analysis and to perform a Gaussian process reconstruction of the growth factor across redshift. Finally, in Sect. 4.4, we summarise our work and interpret our results.

4.2 Methods

4.2.1 LimberJack.jl

The key feature of `LimberJack.jl` which sets it apart from similar, more extensively tested codes (such as `CCL` Chisari *et al.* [205]) are its AD capabilities. AD methods can be classified into two groups, forwards and backwards. In order to understand the difference, let us consider a complicated function of an independent variable, $f(x)$, one can represent the function as a composition of simpler functions, $f(x) = w_n(w_{n-1}(...w_1(w_0))$, where $w_0 = x$. Thus, in order to obtain the gradient of $f(x)$, one can either find the gradient of the simpler functions with respect the independent variable or, alternatively, find the gradient of the original function with respect the simpler functions. The first of these different strategies to accumulate terms in the chain rule corresponds to forwards AD, while the second to backwards AD. Whether to use forwards or backwards AD will greatly depend on the nature of the problem. Generally[1], given a function $f\colon \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^{\mathbb{M}}$, forwards AD will be more efficient at computing ∇f if $N < M$. If, on the other hand, $M < N$, backwards AD is preferred.

`LimberJack.jl` is currently capable of forwards and backwards AD through the `Julia` libraries `ForwardDiff.jl` and `ReverseDiff.jl` respectively. However, only forwards AD is efficiently implemented. Both `ForwardDiff.jl` and `ReverseDiff.jl` perform AD by pushing special types of numbers through the original computer program. On the one hand, `ForwardDiff.jl` uses dual numbers. Dual numbers are expressions of the form $a + \epsilon b$ such that $\epsilon^2 = 0$ but $\epsilon \neq 0$. It can then be shown that given a generic function $f(a + b\epsilon) = f(a) + b\epsilon f'(a)$ (see Sect. 4 of Hoffmann [211] for proof) the gradient of the original function can be easily obtained. This means that `ForwardDiff.jl` imposes little to no constraints on the program it differentiates through[2]. On the other hand, `ReverseDiff.jl` uses taping numbers, a special type of numbers that records all the operations the number undergoes, to generate a trace of the basic operations that compose a generic computer program [100]. Given this record, `ReverseDiff.jl` can then generate an expression for the gradient of the program. Thus `ReverseDiff.jl` requires two passes through the original program. First, a forward pass generates the program's operation trace. Then, a backwards pass computes the partial derivatives and accumulates their values as the input is back-propagated. Because of the greater complexity of the algorithm, `ReverseDiff.jl` imposes strong demands on the computer programs it acts upon. Commonplace computations such as control flow or variable mutation are examples of operations that must be handled carefully.

The implementation of the expressions discussed in Sect. 1.1.4 within `LimberJack.jl` will are thus constrained by the demands of the AD methods used to obtain their

---

[1] The performance of backwards and forwards AD is also impacted by the size of the operation being differentiated.

[2] Since `Julia` is a typed language, some considerations have to be made about potential type instabilities introduced by the use of dual numbers.

gradients. The computation of Eq. 1.69 fundamentally involves two different types of quantities, the radial kernels and the matter power spectrum. On the one hand computing the radial kernels boils down to evaluating the expansion history of the Universe, $H(z)$, and its integral over time, i.e. the comoving radial distance, $\chi(z)$. Calculating the expansion history, as given by the ΛCDM model, amounts to evaluating $H(z) = H_0\sqrt{\Omega_{\rm m}(1+z)^3 + \Omega_r(1+z)^4 + (1-\Omega_{\rm m}-\Omega_r)}$ on a grid of redshifts where $\Omega_{\rm m}$ and Ω_r are the cosmological matter and radiation density respectively. The radial comoving distance is then obtained integrating over the grid using Simpson numerical integration which is generically compatible with AD methods. In the left and centre columns of Fig. 4.1 we show a comparison between the `LimberJack.jl` and `CCL` predictions for the expansion history and comoving distance between $0 < z < 1100$ (top panels). Each panel contains a subpanel where the relative difference between the `CCL` and `LimberJack.jl` predictions is shown. We find that the relative difference between the predictions of `CCL` and `LimberJack.jl` are smaller than 10^{-4} at all redshifts. Moreover, we also show a comparison for the derivative of said quantities with respect to $\Omega_{\rm m}$ when computed by `LimberJack.jl` using AD and finite differences for the same redshift range (bottom panels). Similarly, each panel contains a subpanel where the relative difference between the AD and numerical gradient is shown. We find that the relative difference between the two methods is smaller than 10^{-4} at all redshifts. When interpreting comparisons between finite differences and AD, it is worth noting that finite differences gradients are not inherently correct. Indeed, as stated in Sect. 4.1, AD does not suffer from the numerical errors that finite differences do. Thus, differences between AD and finite differences often occur due the numerical noise in finite differences.

On the other hand, computing the matter power spectrum in an AD-compatible way is a much more difficult problem. The first and most challenging obstacle is obtaining the linear matter power spectrum. As described in Sect. 1.1.3, computing the matter power spectrum involves solving a coupled system of linear differential equations with time- and scale-dependent coefficients. This task has proven a major challenge for most AD environments [207] as it requires tweaking the lower-level programming languages. This is however not a problem for the `Julia` AD ecosystem which has a (currently under development) full Boltzmann-Einstein solver, `Bolt.jl` by Li and Sullivan [210]. While still in its early days, `Bolt.jl` can provide users with full numerical solutions for the linear power spectrum and its derivatives with respect to the ΛCDM parameters. `LimberJack.jl` can then interface with `Bolt.jl` to obtain said predictions.

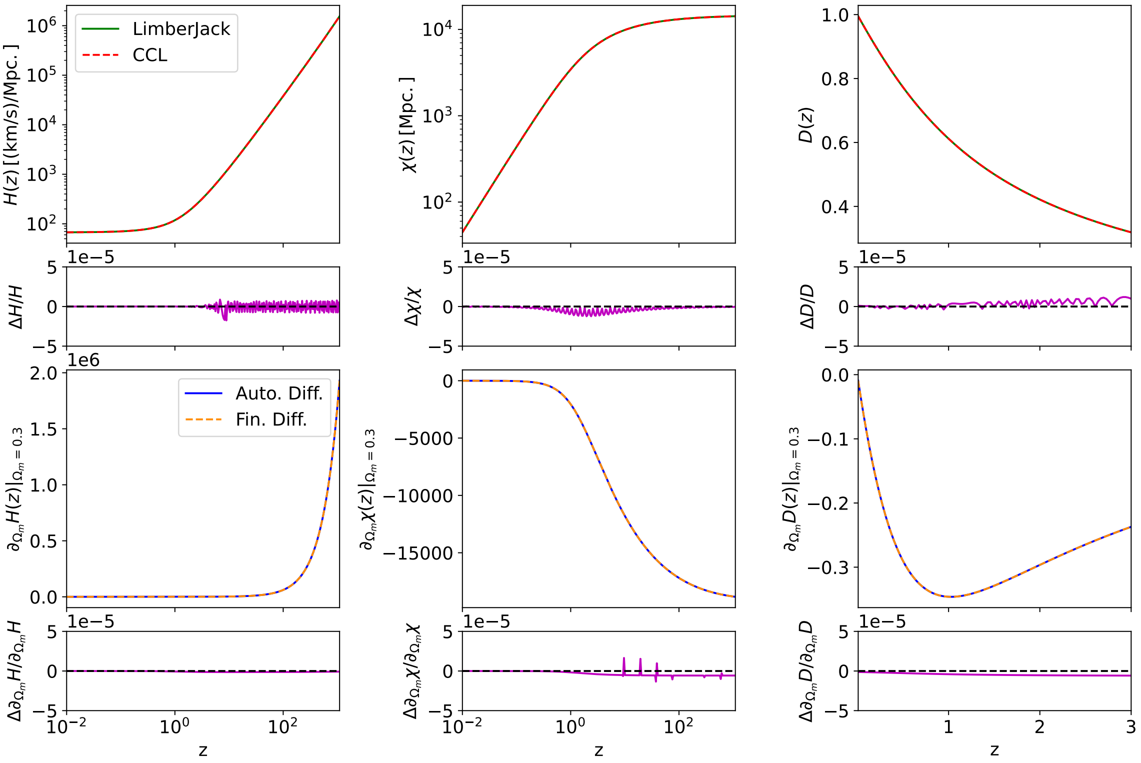

Figure 4.1: **Left column**: Top panel shows a comparison between the `LimberJack.jl` (solid green) and the `CCL` (dashed red) computation of the expansion history between $0 < z < 1100$. Bottom panel shows a comparison between the derivative of the expansion history with respect to $\Omega_{\rm m}$ computed using AD (solid blue) and finite differences (dashed orange) for the same redshift range. **Middle column**: Top panel shows a comparison between the `LimberJack.jl` (solid green) and the `CCL` (dashed red) computation of the comoving distance between $0 < z < 1100$. Bottom panel shows a comparison between the derivative of the comoving distance with respect to $\Omega_{\rm m}$ computed using AD (solid blue) and finite differences (dashed orange) for the same redshift range. **Right column**: Top panel shows a comparison between the `LimberJack.jl` (solid green) and the `CCL` (dashed red) computation of the linear growth factor $0 < z < 3$. Bottom panel shows a comparison between the derivative of the linear growth factor with respect to $\Omega_{\rm m}$ computed using AD (solid blue) and finite differences (dashed orange) for the same redshift range.

However, solving the full Boltzmann-Einstein equations is very computationally expensive (being the bottleneck of most cosmological analyses) even in fast programming languages such as `C++` (CLASS), `Fortran` (CAMB) or `Julia`. Therefore, it is common to look for ways to bypass solving the Boltzmann-Einstein equations in cosmological analyses where speed is important. When the evolution of the matter power spectrum can be assumed to be scale independent, $P(z, k)$ can be computed by constructing a fitting formula that approximates the true value of the matter power spectrum at $z = 0$, $P_0(k)$, at an array of scales and then evolving $P_0(k)$ into the past using the linear growth factor. The first commonly applied fitting formula for $P_0(k)$ was the Bardeen-Bond-Kaiser-Szalay [BBKS 24] formula which modifies the Harrison-Zeldovich power spectrum [160, 212, 213] by a transfer function:

$$P_0(k) \propto \left(\frac{ck}{H_0}\right)^{n_s} T^2(k)\,, \tag{4.2}$$

where $T(k)$ is the transfer function. The most popular fitting formula is that of [25, 26, E&H from now on], which follows the same strategy as the BBKS formula of approximating the transfer function but with a more complex expression that includes baryonic effects such as baryonic acoustic oscillations and small-scale power suppression. Thanks to these inclusions the E&H formula is accurate enough to return unbiased cosmological constraints for the 3x2-pt DES Y1 analysis as shown in Campagne *et al.* [207]. However, it is important to keep in mind that the E&H formula will not be accurate enough to analyse future data set as well as some current ones.

In recent years, a new family of fitting formulae, known as emulators, have grown in popularity. Emulators are computer models (such as neural networks or Gaussian Processes) whose weights are optimised to reproduce a target function over a certain domain. The main advantage of emulators over traditional fitting formulae is that they automate the majority of the trial and error process of building an accurate fitting formula. Moreover, they tend to require less knowledge of the physical problem to be constructed. However, emulators require vasts amounts of training data and the resulting models tend to be far larger than traditional fitting formulae. Nonetheless, emulators have found great success in astrophysics in recent years, offering extremely accurate yet affordable approximations to different functions [214, 215], including the matter power spectrum [209, 216].

The great advantage of these fitting formulae is that they effectively amount to algebraic expressions through which AD algorithms can easily differentiate as shown in Piras and Spurio Mancini [208] and Bonici *et al.* [215]. Combined with their speed, fitting formulae can often be the preferred way to obtain estimates for the primordial power spectrum in gradient-based analyses. For these reasons, `LimberJack.jl` is equipped with both a native `Julia` implementation of the E&H formula and the recently developed Gaussian process based emulator `EmuPk` [216] for the linear matter power spectrum.

The next challenge is to compute the growth factor, $D(z)$, to find the evolution of the linear power spectrum. As discussed in Sect. 1.1.4, the growth factor is obtained by solving the Jeans equation (see Eq. 2.4) which is an inhomogeneous ordinary differential equation. Differentiating through the solutions of differential equations can be done by writing a recursive numerical scheme to solve the differential equation. These schemes amount to a series of linear operations that update mutating variables. While mutating variables can pose challenges for certain backwards modes of AD, the numerical schemes can otherwise be differentiated through to yield gradients for the solution of a differential equation. Thus, `LimberJack.jl` solves the Jeans equation using a second order Runge-Kutta solver. Returning to Fig. 4.1, in right column of this figure, we show a comparison between the `LimberJack.jl` and `CCL` predictions for the linear growth factor (top panels). We also shown a comparison between the derivatives of `LimberJack.jl`'s $D(z)$ with respect to $\Omega_{\rm m}$ when computed using AD and finite difference (bottom panels) for the redshift range $0 < z < 3$. Again, each panel contains a subpanel where the relative difference between the two compared quantities is shown. Concerning the growth factor itself, we find that the relative difference between the `CCL` and `LimberJack.jl` predictions are smaller than 10^{-4} for all the redshift window. Concerning the derivatives of the growth factor, the relative differences between AD and finite differences are again smaller than 10^{-4} for all the redshift window.

To compute the necessary non-linear corrections at small scales, `LimberJack.jl` is equipped with an auto-differentiable implementation of `Halofit` as given by Smith *et al.* [27] with revisions of Takahashi *et al.* [28]. The biggest obstacle to accomplish this is differentiating through the root finding process that occurs within `Halofit`. Similarly to solving differential equations, root finding can be differentiated through by writing a recursive numerical scheme of linear operations with mutable variables [217]. The `Halofit` implementation within `LimberJack.jl` uses the secant method for its root finding through which most AD methods can differentiate. Putting all of these together `LimberJack.jl` can perform a fully auto-differentiable computation of the non-linear matter power spectrum at any redshift or scale. In Fig. 4.2 we show a comparison between the predictions for the non-linear matter power spectrum from `LimberJack.jl` and `CCL` for $z = [0.0, 0.5, 1.0, 2.0]$ and $-2 < \log(k) < 4$ (top panels). Both the `CCL` and `LimberJack.jl` predictions are obtained using the E&H linear power spectrum. Then `Halofit` was used to obtain the non-linear corrections. Moreover, we also show a comparison between the derivative of the `LimberJack.jl` power spectra with respect to Ω_m when computed using AD and finite differences. Finally, each panel contains a subpanel where the relative difference between the two compared quantities is shown. Concerning the power spectra, we find that the relative difference between the `CCL` and `LimberJack.jl` predictions are smaller than 10^{-3} for all the redshift window. However, we observe that the `Halofit` implementation of `LimberJack.jl` systematically over-predicts the non-linear matter power spectrum at higher redshifts. As we will see this bias will manifest when considering angular power spectra involving the CMB lensing tracer later on but it nonetheless doesn't prevent us from achieving our accuracy goals. In the future, we aim to implement more accurate prescriptions of the non-linear matter power spectrum such as as `HMCode` [218] or `baccoemu` [219]. Concerning the derivatives of the same power spectra, the relative differences between AD and finite differences are again smaller than 10^{-3} for all the redshift windows.

The only step left to compute angular power spectra is to bring the radial kernels and the matter power spectrum together and perform the Limber integral (see Eq. 1.69). Similarly to the computation of comoving distances, this is done within `LimberJack.jl` by evaluating all the quantities within the integrand at a regular array of logarithmic scales, $log_{10}(k)$, and then performing the numerical integral for the desired multi-poles, ℓ, using Simpson numerical integration. One small challenge in doing so is finding the corresponding redshift associated with the comoving distance given by the scale and the multipole to evaluate the radial kernels. This inconvenience is a result of `LimberJack.jl` defining the radial kernels as functions of redshift instead of comoving distance. Normally, finding the redshift at a given comoving distance would involve a costly root finding process. However, `LimberJack.jl` handles this by building an interpolator between redshift and comoving distance and then inverting the interpolator to establish a straight forward comoving distance to redshift map for a given set of cosmological values.

In Fig. 4.3 we show a comparison between the predictions of `LimberJack.jl` and `CCL` for the angular power spectra of different types of tracers for $10 < \ell < 1000$. We consider the auto- and cross-correlations of galaxy clustering, cosmic shear and CMB convergence. In all cases the E&H formula was paired with `Halofit` to obtain the matter power spectrum. We also assume a Gaussian redshift distribution centred at

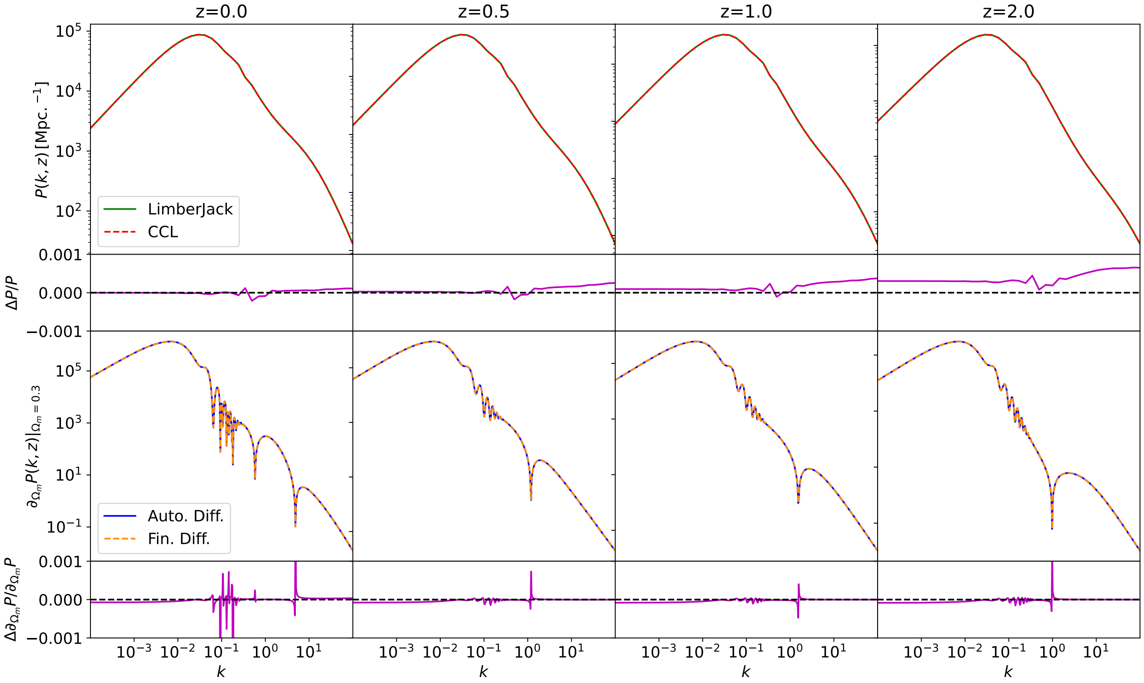

Figure 4.2: Top panels show a comparison between the `LimberJack.jl` (solid green) and the `CCL` (dashed red) computation of the non-linear matter power spectrum for $z = [0.0, 0.5, 1.0, 2.0]$ and $-4 < \log(k) < 2$. Both `LimberJack.jl` and `CCL` used the E&H formula to compute the linear matter power spectrum. Then, `Halofit` was used to obtain the non-linear corrections in both cases. Bottom panels shows a comparison between the derivative of the `LimberJack.jl` matter power spectra with respect to Ω_m computed using AD (solid blue) and finite differences (dashed orange). Each panel contains a subpanel where the relative difference between the two compared quantities is shown.

$z = 0.5$ with a standard deviation $\sigma_z = 0.05$, sampled at 1000 evenly-spaced intervals in the range $0 < z < 2$. We observe that the discrepancy between `LimberJack.jl` and `CCL` is smaller than 10^{-3} for all angular power spectra except for the auto-correlation of CMB lensing where the discrepancy is larger but nonetheless stays below 10^{-2}. This is due to the discrepancy in the evolution of the `Halofit` non-linear matter power spectrum between `LimberJack.jl` and `CCL` shown in Fig. 4.2. Similarly, in Fig. 4.4, we show a comparison for the derivatives of the same quantities with respect to $\Omega_{\rm m}$ when computed using AD and finite differences. The derivative of the clustering tracer proved to be extremely sensitive to the resolution of the numerical integration scheme used to normalise the galaxy distributions leading the observed oscillatory behaviour. Thus, even if the AD derivative is exact, the fact that we are differentiating through a numerical scheme introduces numerical noise. Despite this, we nonetheless observe a sub-percentage-level agreement between the two approaches in all cases. It also worth remembering that disagreements between AD and finite differences don't necessarily imply a mistake in the AD. Indeed finite differences are more likely to differ from the underlying true derivative given that they are subject to truncation errors.

4.3 Results

4.3.1 DES Y1 3x2-pt

In order to validate the constraints obtained using `LimberJack.jl`, we first replicate two different analyses based of DES Y1 3x2-pt [51, 220] data. The Dark Energy Survey is a photometric, 5-year survey, that has observed 5000 deg^2 of the sky using five different filter bands ($grizY$). The observations were made with the 4m Blanco Telescope, provided with the 570-Mpix Dark Energy Camera (DECam), from the Cerro Tololo Inter-American Observatory (CTIO), in Chile. In particular, we use the public `METACALIBRATION` source catalog[3], which is divided in four redshift bins covering the range $z \lesssim 1.6$ [221]. A full description of the methods used to estimate these power spectra, and their associated covariance matrix, from the DES-Y1 data is provided by the authors.

The two aforementioned analyses were done using the well-established Bayesian inference framework `Cobaya` which employs a Metropolis Hastings (MH) sampler. In the first analysis, both `LimberJack.jl` and `Cobaya` used the E&H formula to obtain

---

[3] `https://desdr-server.ncsa.illinois.edu/despublic/y1a1_files/`

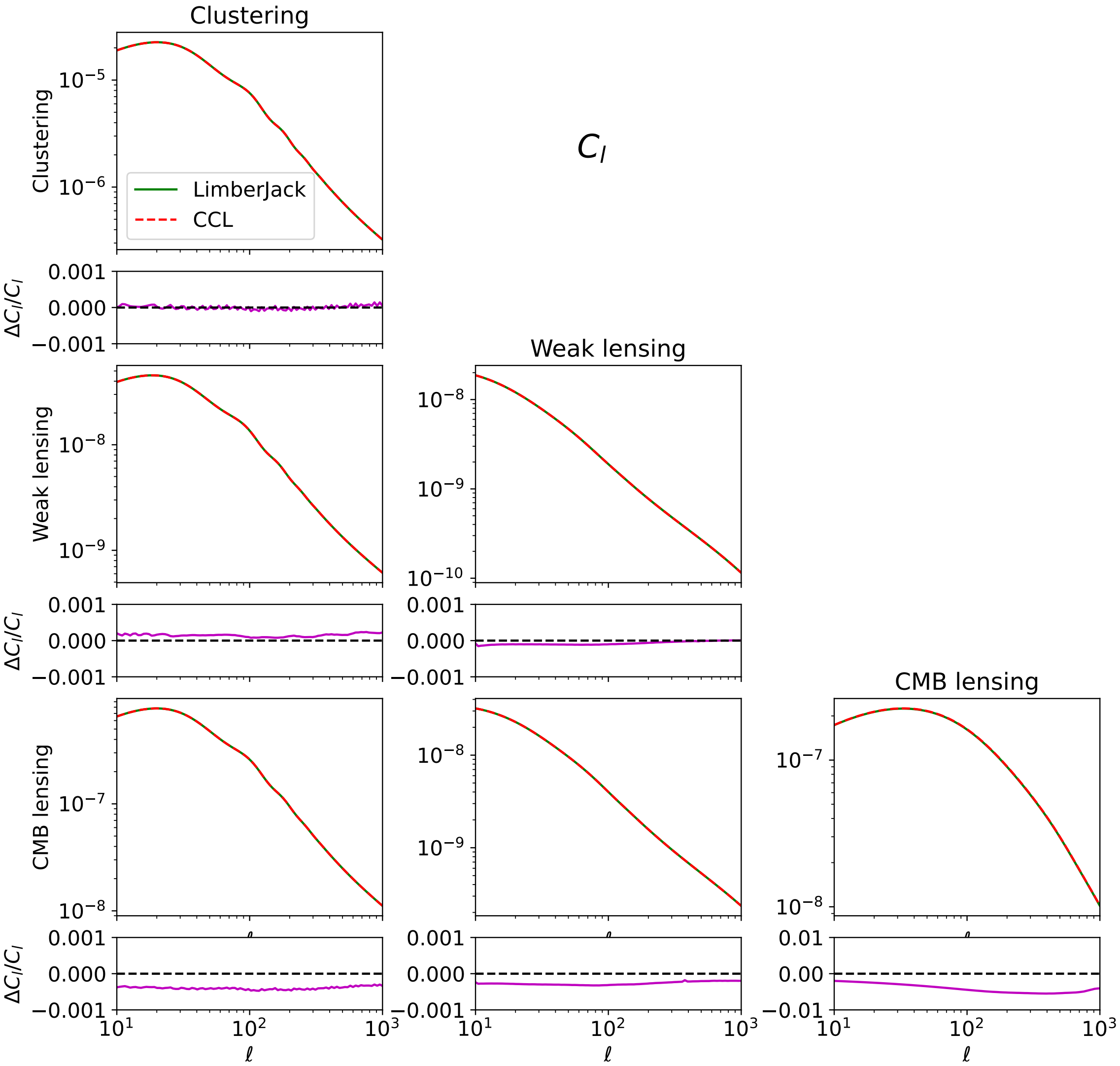

Figure 4.3: Comparison between the `LimberJack.jl` (solid green) and the `CCL` (dashed red) computation of auto- and cross-correlation angular power spectra of galaxy clustering, weak lensing and CMB lensing tracers for the range of multipoles $1 < \log(\ell) < 3$. Both `LimberJack.jl` and `CCL` used the E&H formula to compute the linear matter power spectrum. Then, `Halofit` was used to obtain the non-linear corrections in both cases.

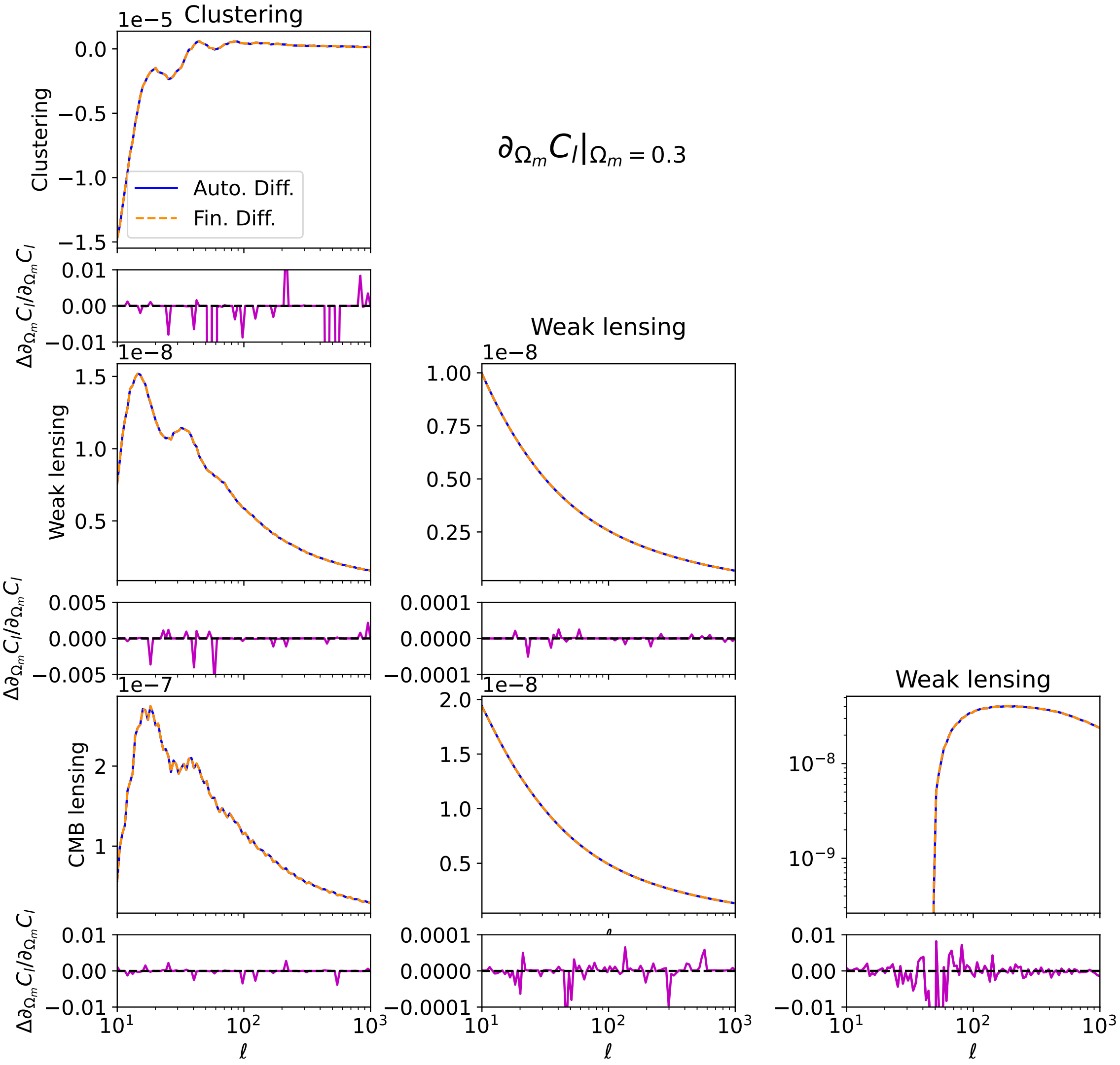

Figure 4.4: Comparison between the derivatives of the auto- and cross-correlation angular power spectra of galaxy clustering, weak lensing and CMB lensing tracers with respect to Ω_m computed using AD (solid blue) and finite differences (dashed orange) for the range of multipoles $1 < \log(\ell) < 3$. In order to compute the gradients of the angular power spectra the E&H formula was used to compute the linear matter power spectrum. Then, `Halofit` was used to obtain the non-linear corrections in both cases.

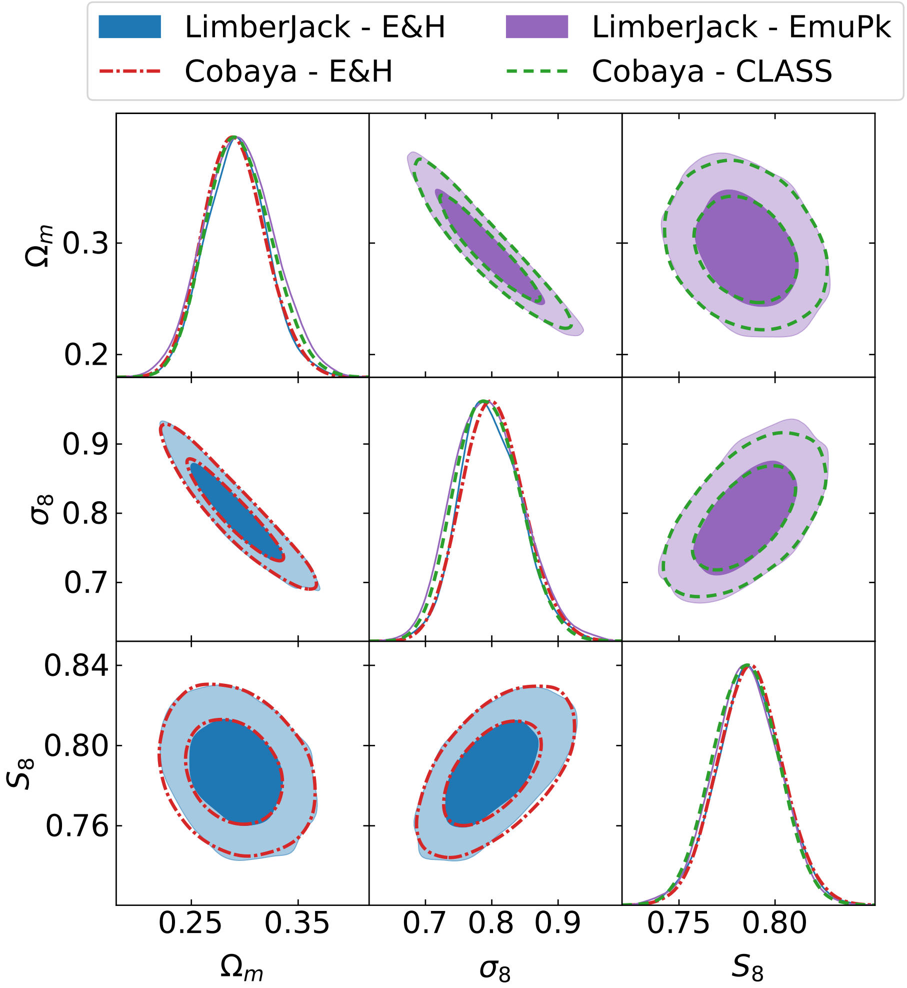

Figure 4.5: Marginalised posteriors distributions from the analysis of DES Y1 3x2-pt data for the cosmological parameters Ω_m, σ_8 and S_8. Lower panel compares the `LimberJack.jl` (solid blue) and `CCL` (dashed red) analyses when using the E&H formula. Upper triangle compares the `LimberJack.jl` (solid purple) and `CCL` (dashed green) analyses when using `EmuPk` and `CLASS` respectively. `Cobaya` contours were obtained using the MH sampler while `LimberJack.jl` contours were obtained using the `NUTS` sampler.

the matter power spectrum. In the second analysis, `Cobaya` used the `CLASS` code to obtain the matter power spectrum while `LimberJack.jl` used the `EmuPk` emulator [216], which was trained on `CLASS`. We use the García-García *et al.* [54] (CGG21 from now on) angular power spectra and covariance matrix of the DES Y1 3x2-pt analysis based of the DES Y1 cosmological catalogue Abbott *et al.* [51] (See Sects. 3.1, 4.1 and 42. of CGG21). The same scale cuts as in CGG21 (see Tab. 4 of CGG21) were applied to the data vector.

With this aim, we make use of the `Julia` library for statistical inference `Turing.jl` [222]. `Turing.jl` is a probabilistic programming language (PPL) that allows the user to create statistical models by explicitly writing the relationship between the sampled parameters and the theoretical predictions for the observed data. `Turing.jl` uses this information to draw a likelihood density function compatible with the `Julia` AD infrastructure as long as the internal computations of the model are also compatible with AD. The user can then condition this function on the observations and sample it using a series of sampling back-ends. In this work we make use of the NUTS sampler as implemented in the `Julia` library `AbstractHMC.jl` within `Turing` (See Sect. 1.2.2.2 for details on the algorithm). Moreover, we choose the AD library `ForwardDiff.jl` to provide NUTS with the gradient of the likelihood. Note that using forwards AD for statistical inference is sub-optimal since the likelihood is a function that maps a high-dimensional space to a single scalar (See Sect. 1.3.2 for more details). In the future, we aim to implement efficient backwards AD in `LimberJack.jl`.

The same priors were used in the two cases for both the `Cobaya` and `LimberJack` analyses. A summary of the priors can be found in Tab. 4.1. We show the resulting posteriors in Fig. 4.5. We observe an excellent agreement between the `LimberJack.jl` and `Cobaya` pipelines regardless of whether the E&H formula or `CLASS`/`EmuPk` is used to obtain the primordial matter power spectrum.

We compare the performance of the two samplers by looking at the number of effective samples [i.e. number of statistically independent samples in the Markov chain 223] per number of likelihood calls. This metric is independent of the hardware used to run the analysis and the time taken by `LimberJack.jl` and `Cobaya` to evaluate the likelihood. Thus, this metric allows us to look at the improvement purely brought about by using gradient-based inference methods. Analysing the Markov chains of the different samplers, we find that `NUTS` is approximately 1.5 times more efficient than `Cobaya` at generating effective samples. However, in order to fairly compare the performance of the two samplers we must take into account that `NUTS` computes the gradient of the likelihood at every step. Given the 25 parameters of the DES

Table 4.1: Prior distributions for the DES Y1 3x2-pt analysis parameters based on the prior range of `EmuPk` [216]. In this table $U(\mathcal{N}, a, b)$ represents a truncated Gaussian distribution with boundaries a and b.

DES Y1 3x2-pt Analyses Priors			
Parameter	Prior	Parameter	Prior
Cosmology		**Shear calibration bias**	
Ω_m	$U(0.2, 0.6)$	m^i	$\mathcal{N}0.012, 0.023)$
Ω_b	$U(0.028, 0.065)$	**Galaxy bias**	
h	$U(\mathcal{N}(70, 5), 0.55, 0.91)$	b_g^i	$\mathcal{N}0.8, 3.0)$
n_s	$U(0.87, 1.07)$	**Intrinsic Alignments**	
σ_8	$U(0.6, 0.9)$	$A_{\mathrm{IA},0}$	$U(-5.0, 5.0)$
		η_IA	$U(-5.0, 5.0)$
Lens redshift calibration		**Sources redshift calibration**	
Δz_1	$\mathcal{N}(0.0, 0.007)$	Δz_1	$\mathcal{N}(-0.001, 0.016)$
Δz_2	$\mathcal{N}(0.0, 0.007)$	Δz_2	$\mathcal{N}(-0.019, 0.013)$
Δz_3	$\mathcal{N}(0.0, 0.006)$	Δz_3	$\mathcal{N}(0.009, 0.011)$
Δz_4	$\mathcal{N}(0.0, 0.01)$	Δz_4	$\mathcal{N}(-0.018, 0.022)$
Δz_5	$\mathcal{N}(0.0, 0.01)$		

Y1 3x2-pt analysis, evaluating the gradient of the likelihood using `ForwardDiff.jl` is roughly 5 to 6 times more expensive than evaluating the likelihood itself which is an order of magnitude faster than if we had used finite differences. In order to compensate for this extra cost the efficiency improvement of NUTS over MH should be equal or larger than the relative cost of computing the gradient of the likelihood. Therefore the efficiency improvement of `NUTS` is not enough to compensate for the cost of computing the gradient in this particular application.

There exist two avenues to tilt the balance in favour of `NUTS`. On the one hand, increasing the efficiency of the sampler. On the other hand, reducing the cost of the likelihood gradient. However, the current bottleneck is the efficiency of the sampler as even more specialised AD methods (such as `Zygote.jl`[4]) would take at least twice the time to compute the likelihood to obtain its gradient. Future works could explore initialising the NUTS mass matrix using the posterior covariance estimate from variational inference[5] algorithms such as `PathFinder.jl` [78]. Therefore, in order to

---

[4] https://github.com/FluxML/Zygote.jl/tree/master

[5] Variational inference algorithms iteratively build an approximation of their target distribution based on a finite number of samples by combining different distributions of their variational family of distributions. Thus, they are often understood as a form of kernel estimation.

showcase the strength of the methods developed on this work we need an application for which the efficiency of gradient-based samplers truly outpaces that of traditional inference methods.

4.3.2 Growth Factor Reconstruction

In the previous section, we showed the reliability of `LimberJack.jl` by reproducing the official DES Y1 3x2-pt analysis. In this section, we will showcase how `LimberJack.jl` can be used to perform statistical inference on models outside of the reach of traditional inference methods. In order to do so we perform a model-independent reconstruction of the growth factor using a Gaussian Process (GP) with more than a hundred parameters (See Sect. 1.4 for a detailed description of GPs).

In this work, we use a GP composed of 101 nodes equally spaced through the redshift window $0 \leq z \leq 3$. The mean of the GP is given by the best-fit ΛCDM Planck 2018 [13, P18 from now on] prediction for $D(z)$. While this choice of mean will bias the reconstruction towards the P18 prediction outside of the data domain, its impact in the data dominated regions is small [see e.g. 1, 2, for similar examples of this GP behaviour in the literature.]. Regarding the covariance matrix, we choose a square exponential covariance function, defined as:

$$K(\boldsymbol{x}, \boldsymbol{x}') = \eta^2 \exp\left[\frac{-|\boldsymbol{x} - \boldsymbol{x}'|^2}{2l^2}\right] + \sigma\mathbb{I}\,, \tag{4.3}$$

where η is the amplitude of the oscillations around the mean, l is the correlation length between the GP realisations and σ is the noise amplitude chose at 10^{-5}. This decision was made based on the fact that the square exponential is a computationally inexpensive and infinitely differentiable kernel, appropriate to model smooth fluctuations around the mean of the GP.

Therefore, the growth factor is given by:

$$D(\boldsymbol{z}) = D_{\mathrm{P18}}(\boldsymbol{z}) + \mathbb{L}(K) \cdot \boldsymbol{v}\,, \tag{4.4}$$

where $\boldsymbol{v}$ is a vector of the GP nodes sampled from a unit variance, diagonal, multivariate normal distribution, $\boldsymbol{z}$ denotes the redshift array at which said nodes sit, $D_{\mathrm{P18}}(\boldsymbol{z})$ is the P18 growth factor evaluated at $\boldsymbol{z}$ and $\mathbb{L}(K)$ is lower-triangular matrix obtained from the Cholesky decomposition of the covariance matrix. A way of interpreting Eq. (4.4) is as a rotation given by L_C on a vector of white noise $\boldsymbol{v}$ that imposes the correlations of the GP kernel.

This model is similar to the one considered in CGG21, but using a GP to reconstruct the growth factor instead of splines. The reasoning behind the choice of GPs over splines for this work is threefold. First, GPs offer a well-defined measure of the uncertainty in their predictions which makes assessing the statistical significance of their results straightforward. Second, GPs are not subject to the strict assumptions that can bias spline reconstructions such as the choice of linear or cubic interpolations. However, it is important to bear in mind that GPs are far from assumption free as their structure is constrained by the properties of the chosen covariance kernel. Nonetheless, as we will show, these assumptions can indeed be neglected when the data is constraining enough. Third, GPs are as differentiable as their covariance matrix kernel [113]. The derivative of a GP $g(\boldsymbol{x}) \sim \mathcal{N}(m(\boldsymbol{x}), K(\boldsymbol{x}, \boldsymbol{x}'))$ is another GP given by $\dot{g}(\boldsymbol{x}) \sim \mathcal{N}(\partial_{\boldsymbol{x}} m(\boldsymbol{x}), \partial_{\boldsymbol{x}} \partial_{\boldsymbol{x}'} K(\boldsymbol{x}, \boldsymbol{x}'))$. This is a very desirable feature that will allow us to use growth rate measurements (i.e. measurements of the gradient of the growth factor) to further constrain the reconstructed growth. As will be shown, the growth rate measurements will highly restrict the evolution of the growth.

In order to constrain this model, we use the south data collection described in CGG21 composed of the 3x2-pt DES Y1 data, the auto-correlation of eBOSS DR16 quasars and the cross-correlation of CMB lensing data with eBOSS DR16 quasars, DES Y1 clustering and DES Y1 weak lensing data. Therefore, our analysis combines a total of 7 two-point statistics which we will refer to as "7×2-pt" hereafter. These particular cross- and auto-correlations correspond to the physical overlap of the different surveys in the sky shown in Fig. 4.7. We explicitly list the considered auto- and cross-correlations in Tab. 4.2. In this table, we can see that the DES Y1 galaxy clustering (GC) data is divided into 5 redshift tomographic bins. The DES Y1 weak lensing (WL) data is divided into 4 redshift tomographic bins. Similarly, the eBOSS DR16 quasar data is divided into 2 redshift tomographic bins. We used the Planck 2018 lensing convergence map. We process all these data following the CGG21 angular power spectra analysis described in Sects. 3.1, 3.2, 3.3, 4.1 and 4.2. Thus we consider a total of 42 different angular power spectra which amount to 665 different data points. The associated covariance matrix of these data was computed using `Cosmoteka` [6] [59] and it is shown in Fig. 4.6. The scale cuts considered for each angular power spectrum are listed in the triangle of Tab. 4.2. For a detailed description of these data, we refer the reader to Sect. 3 of CGG21. Besides the aforementioned data, we also consider a collection of redshift-space distortion (RSD) measurements by the BOSS DR12 [73], eBOSS DR16 [145], Wigglez [146], 6dF [195]

---

[6] https://github.com/xC-ell/xCell

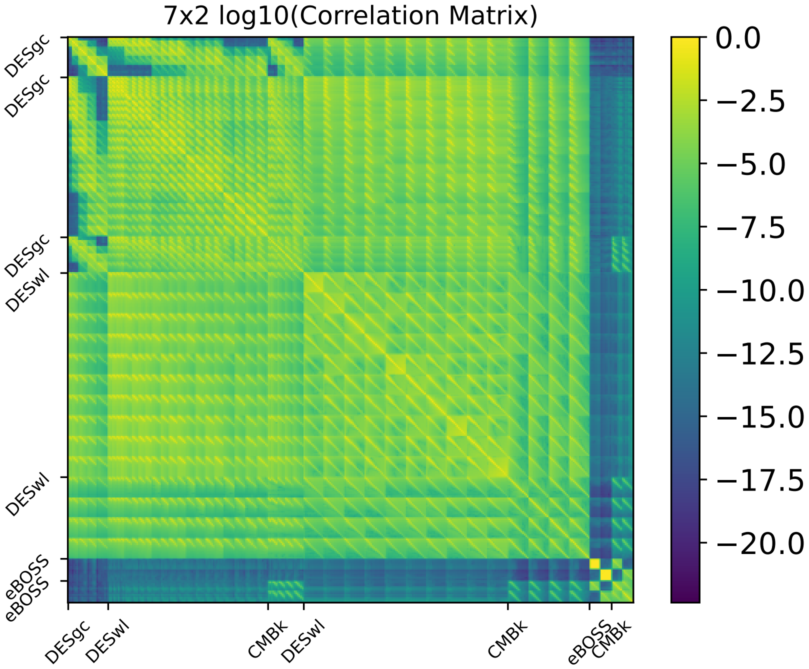

Figure 4.6: Logarithm of the entries of the correlation matrix of the 7×2-pt data set described in Tab. 4.2.

and VIPERS [194] surveys. This analysis constitutes the first combination of all these different data, dramatically improving the constraints by constraining the evolution of the growth factor. We summarise these data in Tab. 4.3 and plot it in Fig. 4.8. For a full description of these data, we refer the reader to Ruiz-Zapatero *et al.* [1].

These data allow us to constraint the growth factor and its derivative across redshift. In order to do so we relate $f\sigma_8(z)$ with the reconstructed $D(z)$ by

$$f\sigma_8(z) = -\sigma_{8\,P18}(1+z)\frac{\partial D(z)}{\partial z}\,, \tag{4.5}$$

where $\sigma_{8\,P18}$ is the σ_8 of the ΛCDM P18 cosmology and $D(z)$ is given by Eq. 4.4. The main difference in Eq. 4.5 with respect to the ΛCDM model is that σ_8 is kept fixed. This is because in our ΛCDM analysis we defined $D(z=0) = 1.0$ such that the amplitude of the growth factor varies by σ_8. Unlike in ΛCDM, in our GP model the amplitude of the growth factor is not fixed, but varies with each GP realisation. Therefore, the role of σ_8 would be completely degenerate with the amplitude of the $z = 0$ GP node and hence is kept fixed.

Once we have drawn an expression for $f\sigma_8(z)$ in our GP model, the next question is how to evaluate it. While this might seem straightforward, the gradient of the

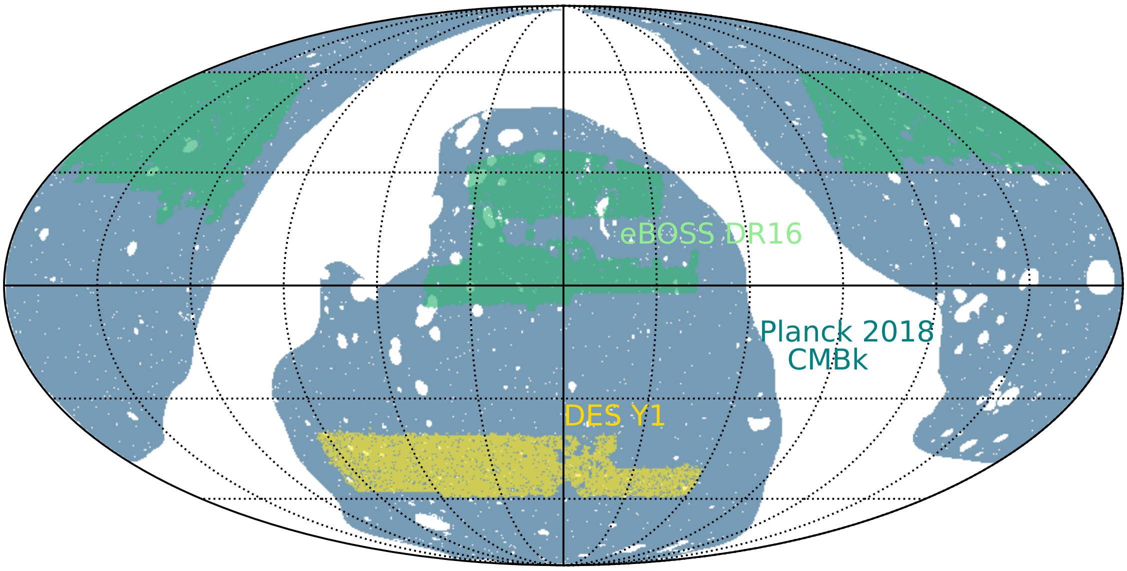

Figure 4.7: Footprint of the surveys forming the 7×2-pt data vector described in in Tab. 4.2.

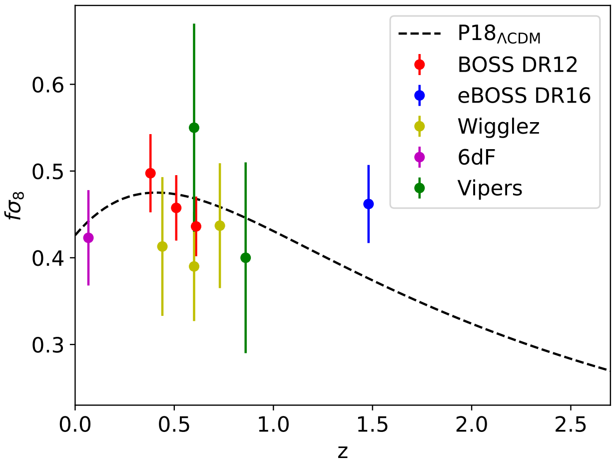

Figure 4.8: $f\sigma_8$ data points from the different surveys used in this work across redshift. Numerical values can be found in Tab. 4.3

Table 4.2: Auto- and cross-correlations that form the 7×2-pt data vector. The first row below the title shows the names of the data sets, namely DES Y1 galaxy clustering (GC) and weak lensing (WL), eBOSS DR16 quasars (eBOSS) and the Planck 2018 CMB lensing convergence map (Planck CMBκ). The row below shows the tomographic bins of each survey. The triangle table shows the scale cuts in angular scales (ℓ) applied to each angular power spectrum. In total, the 7×2-pt data vector added to 665 data points.

7×2-pt Data Vector											
DES Y1 GC					DES Y1 WL				eBOSS		Planck CMBκ
0	1	2	3	4	0	1	2	3	0	1	0
0−145					0−145	0−145	0−145	0−145			30−145
	0−225				0−225	0−225	0−225	0−225			30−225
		0−298			0−298	0−298	0−298	0−298			30−298
			0−371		0−371	0−371	0−371	0−371			30−371
				0−435	0−435	0−435	0−435	0−435			30−435
					30−2000	30−2000	30−2000	30−2000			30−2000
						30−2000	30−2000	30−2000			30−2000
							30−2000	30−2000			30−2000
								30−2000			30−2000
									79−590		79−590
										79−764	79−764

Table 4.3: Lists the RSD data used in this analysis.

Data set	Redshifts	Data Points
BOSS DR12 [73]	0.38 - 0.61	3
eBOSS DR16 [145]	1.48	1
Wigglez [146]	0.44 - 0.73	3
VIPERS [194]	0.60 - 0.86	2
6dF [195]	0.067	1

growth factor is extremely sensitive to numerical errors which can lead to non-physical predictions for $f\sigma_8(z)$. One solution to this problem would be to compute the gradient analytically using the properties of GPs as discussed above. However, in practice evaluating the derivative of the GP analytically involves treating said derivative as a second GP with a cross-covariance matrix with the original GP. Computing and inverting the cross-correlation matrix between the original GP and its derivative at every step of the Markov chain for the number of nodes used in this analysis proved too computationally expensive.

The solution found was to use the noise properties of the GP to interpolate the GP to a finer grid such that a finite difference can be taken to obtain the gradient to enough precision. Effectively, this is done by applying a Wiener filter to the original GP where the kernel used for the filtering is the covariance matrix of the GP. In this scenario, the Wiener filter is just given by $W(\boldsymbol{z}^*, \boldsymbol{z}) = K(\boldsymbol{z}^*, \boldsymbol{z})K^{-1}(\boldsymbol{z}, \boldsymbol{z})$ where $\boldsymbol{z}^*$ is the finer nodes array of redshifts. The resampled growth factor is then given by $D(\boldsymbol{z}^*) = D_{\mathrm{P18}}(\boldsymbol{z}^*) + W(\boldsymbol{z}^*, \boldsymbol{z})D(\boldsymbol{z})$. While this approach also involves inverting a matrix, this is just the covariance matrix of the GP which is a much cheaper operation.

Having discussed our model it is time to discuss how theory and observations are brought together to constrain the model. We built a Gaussian likelihood assuming the RSD and angular power spectra measurements are completely uncorrelated such that the final likelihood is the product of the likelihoods of the individual data types. This is a fair assumption given that the RSD data is mostly located in the northern hemisphere whereas the angular power power spectra are located further South, resulting in a small overlap between the surveys. We then derived constraints for the GP and cosmological parameters by applying Bayes theorem using the priors displayed in Tab. 4.5.

Note that the GP hyperparameters l and η were kept fixed. On the one hand, fixing the hyperparameters avoids the volume effects on the posteriors expected from including these parameters. On the other hand, fixing the hyperparameters to a particular set of values will introduce certain biases in the final reconstruction of the growth factor specially in the regions where data is sparse. For example, fixing the correlation length is expected to induce spurious oscillations in the reconstructed function outside of the data range. Nonetheless the decision to keep both parameters fixed was made as freeing them introduces severe non-Gaussian degeneracies in the posteriors which pose a challenge even to gradient-based methods such as HMC or NUTS. This is due to their hierarchical relationship with the GP nodes. Future studies that wish to undertake a completely model independent reconstruction will need to explore inference methods specifically tailored to deal with such non-Gaussianity such as Riemannian HMC [224].

The number of parameters listed in Tab. 4.5 adds up to 128, a dimensionality which requires gradient-based samplers. For this reason, we employed a Gibbs sampling set up where GP parameters were sampled by one NUTS sampler, keeping all other parameters fixed, and then the rest of the parameters (cosmology and nuisances) were sampled by their own NUTS sampler keeping the GP parameters fixed to their last sample. Moreover, the chains were started at the DES Y1 ΛCDM best-fit cosmology with the GP nodes starting from zero. The Gibbs scheme combined with the starting point proved to greatly increase the efficiency of the NUTS adapting phase in this high dimensional space.

In total, four analyses were run. First, we ran two ΛCDM analyses, one of the 7×2-pt data and another of 7×2-pt + RSD data. In these analyses, we used the ΛCDM model to predict the growth factor as opposed to doing a GP reconstruction. Then the same two analyses were ran but performing the GP reconstruction. In addition to these, we also rerun the CGG21 analysis of 7×2-pt data using the priors listed in Tab. 4.5 in order to be able to compare our results against the `Cobaya` pipeline.

We also study the sampling efficiency in each of these four cases. A summary of our analysis can be found in Tab. 4.4. Starting with the ΛCDM, we find that the effective sample size per calls of the likelihood of the NUTS algorithm with the 7×2-pt data is once again approximately 1.5 times larger than when using the MH algorithm. Remarkably, we found that computing the gradient of the likelihood in this analysis, using AD, was roughly 5 times more expensive than computing the likelihood itself, the same relative cost as the DES Y1 likelihood gradient. Adding RSD data to the ΛCDM analysis results in virtually the same effective sample size per calls of the likelihood when using the NUTS sampler. This is to be expected given that fact that the posteriors remain Gaussian and the number of dimensions is unchanged. Looking at the GP analyses, the addition of 101 extra new parameters

Table 4.4: Effective sample size per calls of the likelihood for MH and NUTS samplers when analysing 7×2-pt and 7×2-pt +RSD data using both ΛCDM model as well as the GP reconstruction.

Analysis	ESS/ Lkl. calls		$\frac{t[\nabla\mathcal{L}]}{t[\mathcal{L}]}$
	NUTS	MH	
7×2-pt $_{\Lambda\text{CDM}}$	0.0105	0.0073	4.95
7×2-pt +RSD$_{\Lambda\text{CDM}}$	0.0980	-	5.12
7×2-pt $_{\text{GP}}$	0.0030	-	28.45
7×2-pt +RSD$_{\text{GP}}$	0.0071	-	28.60

renders the MH algorithm computationally unfeasible. Thus, we cannot directly compare the effective sample sizes per calls of the likelihood of the two samplers. However, we can still draw comparisons with previous analyses. We find that the efficiency of the NUTS sampler when performing a GP reconstruction based on the 7×2-pt data is around half the efficiency of the ΛCDM analysis of same data using the MH algorithm. Moreover, if we add RSD data to the analysis we find that the efficiency of the sampler becomes virtually identical to that of the ΛCDM analysis of the 7×2-pt data using the MH algorithm. This is due to the RSD data constraining the derivative of the growth and, therefore, putting stronger constraints on the GP nodes. Taking into account the cost of the likelihood gradient, we find that adding an extra 101 free parameters increases the relative cost of the gradient of the likelihood to a factor of ~ 30 when using AD. However, using AD to obtain the gradient of the likelihood remains an order of magnitude cheaper than using finite differences despite the extra 101 free parameters. This showcases the favourable scaling of AD methods with the dimensionality of the problem. Hypothetically, an even more favourable scaling could be achievable by implementing an efficient backwards AD algorithm to obtain the gradient. In combination these two achievements managed to produce converged constraints for our GP analyses in O(10^2) CPU hours. For reference, our ΛCDM analysis of 7×2-pt data using the MH sampler took 24 CPU hours to converge. Thus our methods make analyses with a O(100) parameters computationally feasible while leaving ample room for speed ups.

4.3.2.1 ΛCDM Results

We start by analysing our data using a traditional ΛCDM model where the growth factor is obtained from the ΛCDM parameters by solving the Jeans equation for the matter anisotropies. This allows us to establish a frame of reference against which to

Table 4.5: Prior distributions for parameters considered for the Growth Factor Reconstruction. We sample the cosmological parameters keeping σ_8 fixed to avoid degeneracies with the Gaussian process. For the same reason the Gaussian process hyperparameters are also kept fixed.

Growth Factor Reconstruction Priors			
Parameter	Prior	Parameter	Prior
Cosmology		**DES Y1 - Sources redshift cal.**	
$\Omega_{\rm m}$	$U(0.2, 0.6)$	$\Delta z^1_{\rm wl}$	$N(-0.001, 0.016)$
$\Omega_{\rm b}$	$U(0.028, 0.065)$	$\Delta z^2_{\rm wl}$	$N(-0.0019, 0.013)$
h	$U(N(70, 5), 0.55, 0.91)$	$\Delta z^3_{\rm wl}$	$N(0.009, 0.011)$
$n_{\rm s}$	$U(0.87, 1.07)$	$\Delta z^4_{\rm wl}$	$N(-0.018, 0.022)$
σ_8	0.81		
DES Y1 - Lens redshift cal.		**Intrinsic Alignments**	
$\Delta z^1_{\rm gc}$	$N(0.0, 0.007)$	$A_{\rm IA,0}$	$U(-5, 5)$
$\Delta z^2_{\rm gc}$	$N(0.0, 0.007)$	$\eta_{\rm IA}$	$U(-5, 5)$
$\Delta z^3_{\rm gc}$	$N(0.0, 0.006)$	**eBOSS - Galaxy bias**	
$\Delta z^4_{\rm gc}$	$N(0.0, 0.01)$	$b^i_{\rm QSO}$	$U(0.8, 5.0)$
DES Y1 - Shear calibration bias		**Gaussian Process**	
m^i	$N(0.012, 0.023)$	η	0.2
DES Y1 - Galaxy bias		l	0.3
$b^i_{\rm gc}$	$U(0.8, 3.0)$	v^i	$U(N(0, 1), -2, 2)$

compare the results of the GP reconstruction. Moreover, it also allows us to ground our analysis by comparing it with previous analyses of the same combination of data in the literature, namely CGG21.

In Fig 4.9, we show the obtained $\Omega_{\rm m}$, σ_8 and S_8 posteriors when performing a ΛCDM analysis of the 7×2-pt data set with and without including additional RSD data (green and red contours respectively). We also show the contours obtained by the CGG21 pipeline when analysing the 7×2-pt data using black dashed lines. Our ΛCDM analysis of the 7×2-pt data found $\Omega_{\rm m} = 0.287 \pm 0.027$, $\sigma_8 = 0.809 \pm 0.045$ and $S_8 = 0.789 \pm 0.016$. The reanalysis of CGG21 using the priors shown in Tab. 4.5 found $\Omega_{\rm m} = 0.296 \pm 0.028$, $\sigma_8 = 0.794\ 0.043$ and $S_8 = 0.786 \pm 0.015$. Thus, we can observe that the constraints produced by the `LimberJack.jl` pipeline presented in this work are completely consistent with the results of the `Cobaya` pipeline of CGG21. Combining the 7×2-pt with the RSD data we found $\Omega_{\rm m} = 0.277 \pm 0.021$, $\sigma_8 = 0.827 \pm 0.034$ and $S_8 = 0.793 \pm 0.015$. The addition of RSD data improved the constraints on $\Omega_{\rm m}$ and σ_8 by 20%. However, the constraints on S_8 were largely unaffected. Regardless of whether or not RSD data are included, our results are in 2 σ disagreement with the P18 results which found $S_8 = 0.832 \pm 0.013$ [See Tab. 1 of 13]. The full numerical posteriors of the `LimberJack.jl` ΛCDM analyses of 7×2-pt and 7×2-pt plus RSD data can be found in the first two columns of Tab. 4.6.

4.3.2.2 GP Results

We are now in a position to start discussing the GP reconstruction of the growth factor. We start by considering how introducing the GP affects the parameter constraints previously discussed in Sect. 4.3.2.1. In order to establish a comparison we derive constraints for $\sigma_8 \equiv \sigma_8(z=0)$ and $S_8 \equiv S_8(z=0)$ from the GP reconstruction of the growth factor. From Eq. 4.5, we can see that within the GP model:

$$\sigma_8 = \sigma_8^{\rm P18} D(z=0)\,, \tag{4.6}$$

$$S_8 = \sigma_8^{\rm P18} D(z=0)\sqrt{\Omega_{\rm m}/0.3}\,, \tag{4.7}$$

where $D(z)$ is given by 4.4.

In Fig. 4.10, we show four different set of posteriors for the parameters $\Omega_{\rm m}$, σ_8 and S_8. In the lower triangle, we show the contours obtained when using the ΛCDM model (red) and the GP model (blue) to analyse the 7×2-pt data. Similarly, in the upper triangle we show the contours obtained when using the ΛCDM model (red) and the GP model (blue) to analyse the 7×2-pt data in addition to RSD data. The associated numerical constraints for the GP reconstructions are $\Omega_{\rm m} = 0.289 \pm 0.026$,

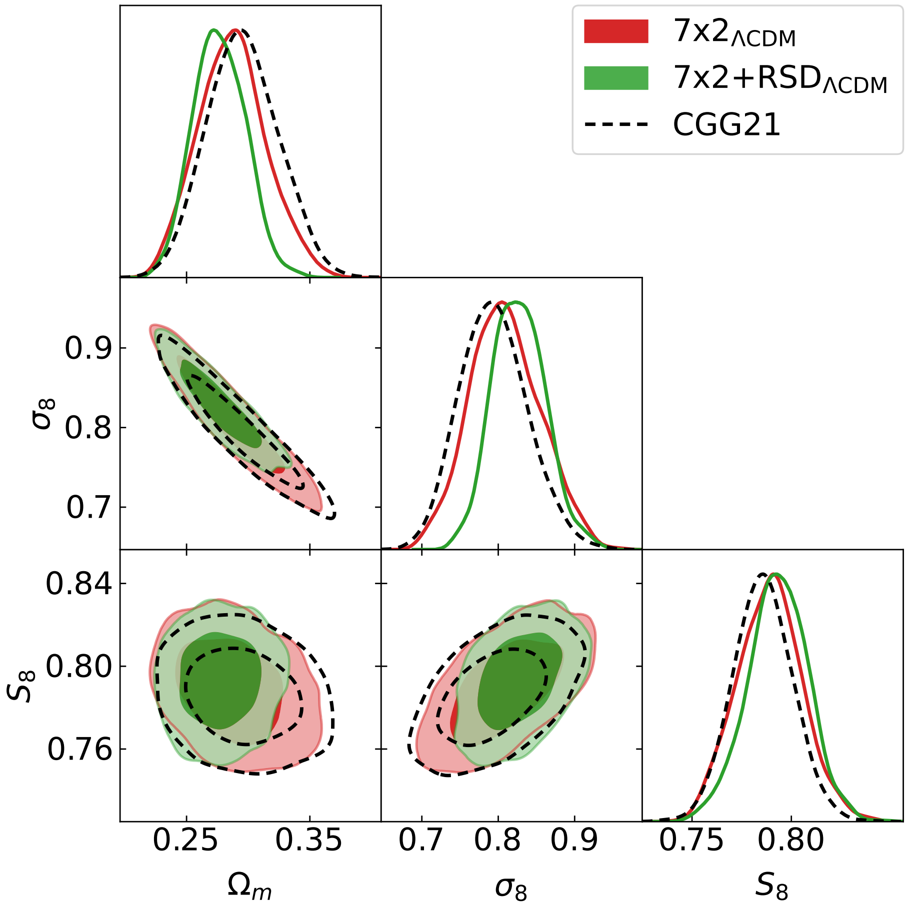

Figure 4.9: 2D and 1D marginal distributions for the parameters Ω_m, σ_8 and S_8 of the ΛCDM analysis of 7×2-pt data (red) and 7×2-pt plus RSD data (green). Black dashed contours show the reanalysis of the CGG21 using Tab. 4.5's priors.

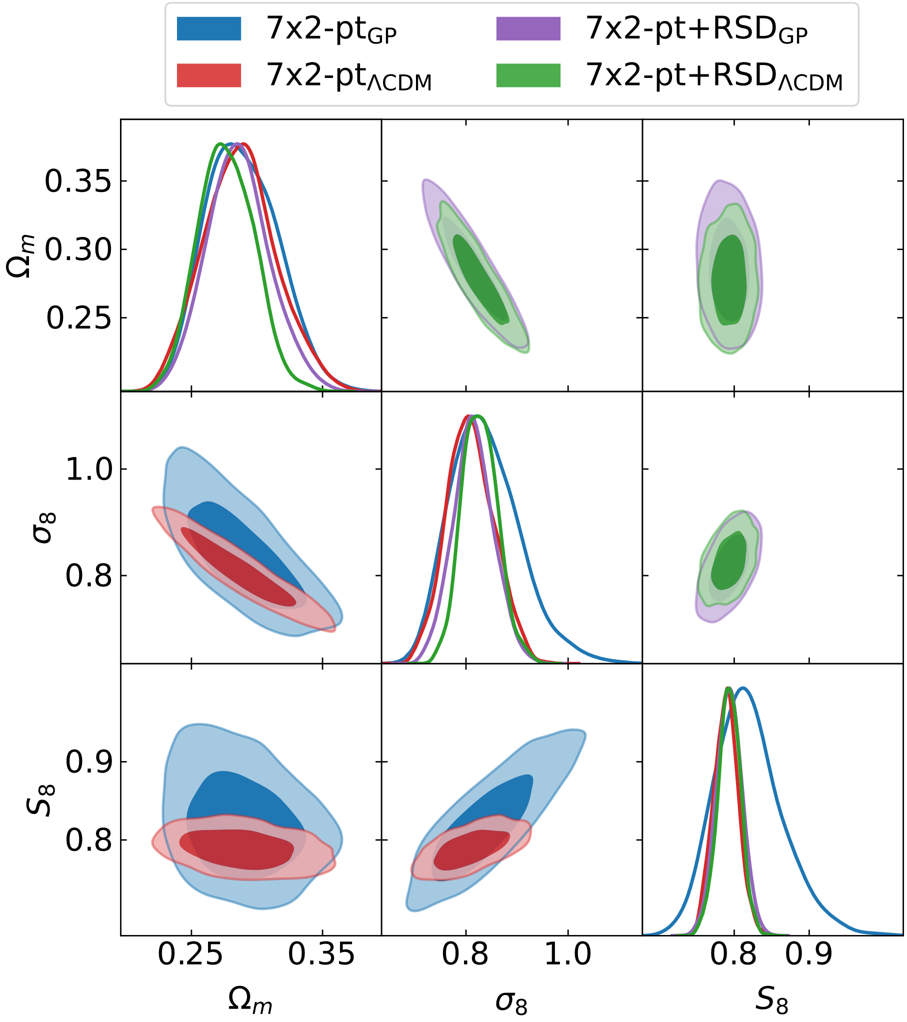

Figure 4.10: Lower triangle: 2D marginal distributions for the parameters Ω_{m}, σ_8 and S_8 of the ΛCDM (red) and GP reconstruction (blue) analyses of 7×2-pt data. Upper triangle: 2D distributions for the parameters Ω_{m}, σ_8 and S_8 of the ΛCDM (green) and GP reconstruction (purple) analyses of 7×2-pt data combined with RSD data. 1D Marginals for all the previous analyses are shown along the diagonal.

$\sigma_8 = 0.839 \pm 0.067$ and $S_8 = 0.839 \pm 0.067$ when 7×2-pt alone is consider and $\Omega_\mathrm{m} = 0.286 \pm 0.023$, $\sigma_8 = 0.815 \pm 0.039$ and $S_8 = 0.793 \pm 0.017$ when RSD data is included. The full posteriors can be found in the last two columns of Tab. 4.6. Performing the GP reconstruction of the growth factor greatly reduces the constraining power of the 7×2-pt data. We observe ~30% wider constraints when using the GP model to analyse the data on average across the three parameters. It is important to note that the overwhelming majority of the impact occurs in the parameters explicitly related to the growth factor such as σ_8 and S_8 and the linear bias parameters of the different probes. Other parameters such as h or n_s remain virtually unchanged. The S_8 constraint is also centred at a significantly higher value than when performing a ΛCDM analysis. This due to the lack of data at $z = 0$ to constrain the GP, which is reflected in the large error bar in the constraint. When RSDs are included the degradation in constraining power when performing the GP reconstruction is much smaller, with constraints only being ~10% wider. Moreover, the S_8 constraint becomes centred at exactly the same value as when performing the ΛCDM analysis of the same data. Due the larger error bar, we observe that nonetheless the tension with the P18 ΛCDM value drops to 1.7 sigma.

In order to understand how the inclusion of RSD data leads to these changes, we have to look at the reconstructed growth factor. In each of the rows of Fig. 4.11, we show the constraints obtained for $D(z)$, $S_8(z)$ and $f\sigma_8(z)$ (in this order) as functions of redshift. In each panel, we compare the constraints obtained when using the ΛCDM model against the GP reconstruction. The left-column panels show analyses where only the 7×2-pt data was used while the right column shows the respective analyses when RSD data were included. Moreover, each panel has a subpanel showing the evolution of the 1σ confidence interval of each function over redshift.

Let us begin the discussion by focusing on the top panels of Fig. 4.11, which show the evolution of the growth factor. In these panels, we can see how introducing a GP to reconstruct the growth factor from the data increases the error bars in the predictions of $D(z)$ by one to two orders of magnitude when compared to the ΛCDM prediction based on the same data. Including RSD data significantly contributes to constraining $D(z)$ regardless of the model considered. In the case of ΛCDM analyses, we observe a 10% reduction in the standard deviation of $D(z)$ across redshift. The impact of RSD data is even bigger when we instead use a GP to reconstruct $D(z)$. In this case we observe a 20% to 30% percent improvement in the constraints depending on the redshift window. Nonetheless, we observe that in all four cases the obtained growth factor is compatible with the P18 prediction at all redshifts.

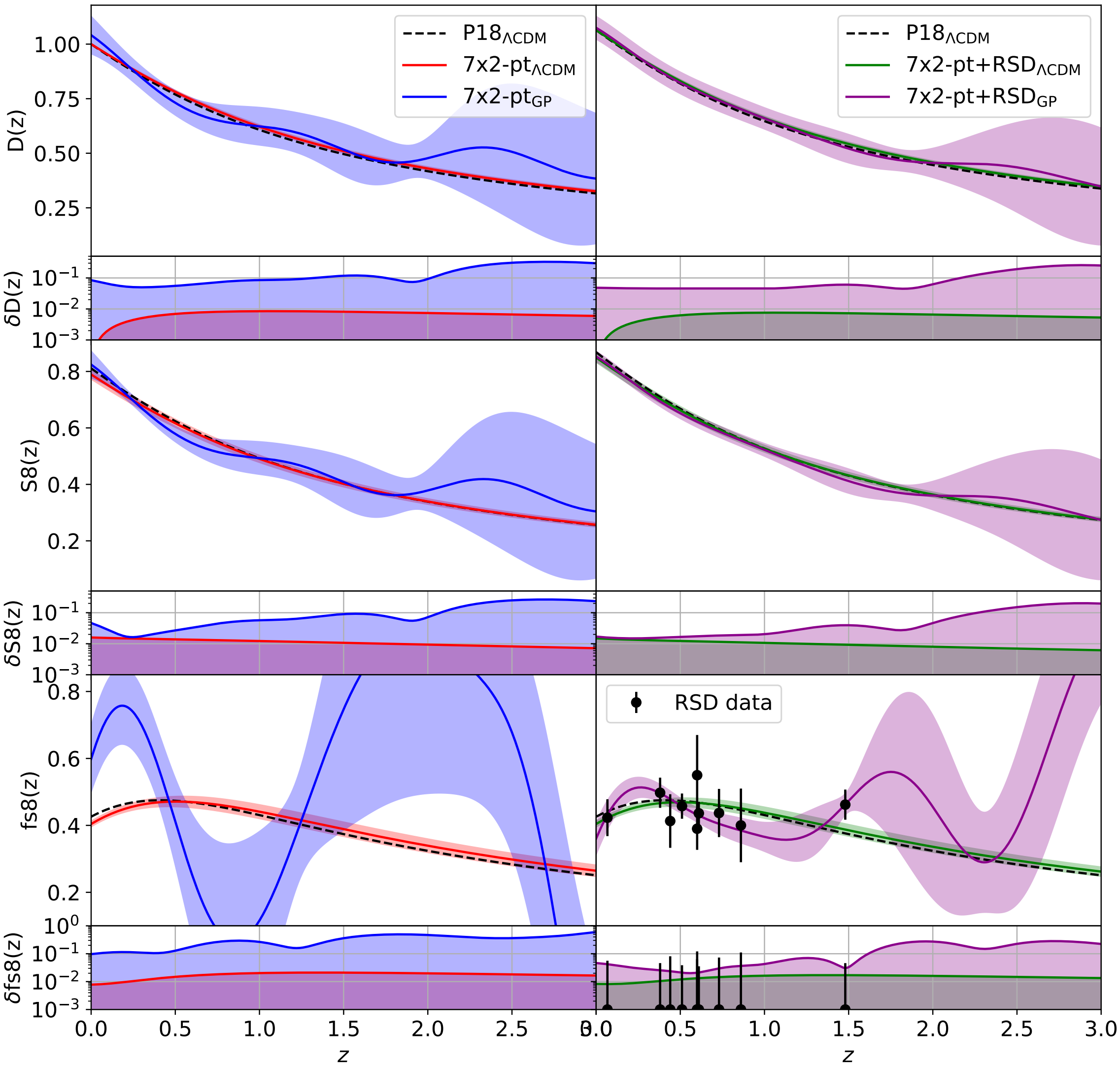

Figure 4.11: Evolution across redshift of the reconstructed $D(z)$ (first row), $S_8(z)$ (second row) and $f\sigma_8(z)$ (third row). Results based on 7×2-pt data are shown on the left column while results combining 7×2-pt plus RSD data are shown in the right column. In each panel we over plot the prediction of the ΛCDM model for the given data as well as the result of the GP reconstruction. Moreover, we also overplot the P18 ΛCDM prediction using a black dashed line for reference. Finally, each panel counts with a subpanel where the we show the evolution of the 1σ confidence intervals of the plotted functions across redshift.

It is important to note that the uncertainty of the $D(z)$ ΛCDM prediction actually falls to zero at $z = 0$ since in this model the growth factor is re-scaled to be precisely one at this redshift. Therefore, by looking at the ΛCDM prediction of $D(z)$ we are omitting a large contribution to the uncertainty of the growth factor in this model, its amplitude. Therefore, it is useful to consider quantities such as $S_8(z)$ that do incorporate the uncertainty in the amplitude of the growth factor in the ΛCDM model, encapsulated in the parameter σ_8. In the second row of panels of 4.11 we show the associated $S_8(z)$ constraints for the four scenarios considered. In these panels, we can see that the uncertainty in $S_8(z)$ of the GP reconstruction at low redshift is actually comparable to that of the ΛCDM model the uncertainty in the amplitude is taken into account. However, the uncertainty of the GP reconstruction quickly increases once the data becomes sparse as seen in the $D(z)$ panels.

It is also interesting to note that, as expected, fixing the correlation length of the GP induces oscillations in the reconstructed growth factor and $S_8(z)$. This is visible in the GP reconstruction of $D(z)$ based on 7×2-pt data. However, including RSD data nullifies these oscillations. In order to understand these behaviours we need to look at the third rows of panels of Fig. 4.11. In these panels we show the associated $f\sigma_8(z)$ predictions as a function of redshift. In the left panel we can see how the oscillations on the growth factor induced by our assumptions on the GP hyperparameters result in non-physical predictions for $f\sigma_8(z)$ despite the growth factor itself being perfectly compatible with the P18 ΛCDM prediction. Including RSD data solves this problem by directly constraining the possible values of $f\sigma_8(z)$ given the current data. These constraints on $f\sigma_8(z)$ translate into strict demands for the evolution of the gradient of the growth factor. Therefore, when RSD data are included, we can see how the oscillatory behaviour present in the reconstruction from only 7×2-pt data disappears. This means that when the data is constraining enough the GP reconstruction remains unbiased by the assumptions introduced by fixing the hyperparameters.

Finally it is relevant to compare the growth factor reconstruction performed in this work using GPs against the reconstruction performed by CGG21 from the same data using splines. In Fig. 4.12 we can observe that our GP reconstruction stands in good statistical agreement with the four constraints on $S_8(z)$ at $z = [0.24, 0.53, 0.83, 1.5]$ obtained by CGG21. This reinforces the notion that these different model-agnostic approaches are data-driven, and not biased by their own individual assumptions.

Table 4.6: Posterior distributions of the Growth Factor Reconstruction analyses. Error bars denote the 1σ confidence interval.

Growth Factor Reconstruction Posteriors				
Param.	7×2-pt $_{\Lambda\mathrm{CDM}}$	7×2-pt +RSD$_{\Lambda\mathrm{CDM}}$	7×2-pt $_{\mathrm{GP}}$	7×2-pt +RSD$_{\mathrm{GP}}$
χ^2	658	645	652	635
$\frac{\chi^2}{d-p}$	1.03	0.99	1.21	1.16
Ω_m	$0.287^{+0.026}_{-0.030}$	0.277 ± 0.021	$0.289^{+0.025}_{-0.031}$	$0.286^{+0.022}_{-0.026}$
Ω_b	$0.0450^{+0.0086}_{-0.012}$	$0.0433^{+0.0084}_{-0.012}$	$0.0448^{+0.0082}_{-0.012}$	$0.0444^{+0.0086}_{-0.012}$
h	$0.725^{+0.042}_{-0.050}$	0.733 ± 0.039	0.726 ± 0.040	0.731 ± 0.040
n_s	0.960 ± 0.056	0.965 ± 0.059	$0.961^{+0.058}_{-0.069}$	$0.948^{+0.054}_{-0.078}$
σ_8	$0.809^{+0.043}_{-0.051}$	$0.827^{+0.033}_{-0.037}$	$0.839^{+0.056}_{-0.079}$	0.815 ± 0.039
S_8	0.789 ± 0.016	$0.793^{+0.016}_{-0.014}$	$0.820^{+0.036}_{-0.052}$	0.793 ± 0.017
b^1_{gc}	1.46 ± 0.10	1.418 ± 0.082	1.46 ± 0.10	1.444 ± 0.090
b^2_{gc}	1.77 ± 0.11	1.729 ± 0.084	$1.84^{+0.12}_{-0.13}$	$1.771^{+0.094}_{-0.11}$
b^3_{gc}	1.75 ± 0.10	1.705 ± 0.078	1.87 ± 0.14	$1.752^{+0.092}_{-0.11}$
b^4_{gc}	2.13 ± 0.12	2.081 ± 0.095	$2.29^{+0.18}_{-0.26}$	$2.13^{+0.11}_{-0.14}$
b^5_{gc}	2.19 ± 0.13	2.14 ± 0.11	$2.31^{+0.22}_{-0.34}$	$2.18^{+0.13}_{-0.15}$
Δz^1_{gc}	0.0031 ± 0.0067	0.0040 ± 0.0066	0.0027 ± 0.0067	0.0039 ± 0.0068
Δz^2_{gc}	$0.0029^{+0.0076}_{-0.0066}$	0.0030 ± 0.0067	0.0016 ± 0.0067	0.0025 ± 0.0066
Δz^3_{gc}	0.0013 ± 0.0056	0.0006 ± 0.0060	0.0004 ± 0.0057	0.0007 ± 0.0056
Δz^4_{gc}	$0.0025^{+0.0085}_{-0.0097}$	0.0028 ± 0.0096	0.0019 ± 0.0095	0.0024 ± 0.0090
Δz^5_{gc}	-0.0013 ± 0.0095	-0.001 ± 0.010	-0.0011 ± 0.0099	-0.0010 ± 0.0099
m^1_{wl}	0.020 ± 0.023	0.022 ± 0.023	0.016 ± 0.022	0.022 ± 0.022
m^2_{wl}	0.014 ± 0.021	0.009 ± 0.022	0.012 ± 0.022	0.009 ± 0.022
m^3_{wl}	$0.011^{+0.022}_{-0.020}$	0.009 ± 0.021	0.012 ± 0.020	0.009 ± 0.020
m^4_{wl}	0.003 ± 0.022	0.006 ± 0.020	0.007 ± 0.021	0.006 ± 0.021
Δz^1_{wl}	0.006 ± 0.012	0.009 ± 0.013	0.003 ± 0.013	0.009 ± 0.013
Δz^2_{wl}	-0.030 ± 0.010	-0.035 ± 0.011	-0.031 ± 0.011	-0.035 ± 0.011
Δz^3_{wl}	0.0064 ± 0.0099	0.007 ± 0.010	0.0093 ± 0.0099	0.0070 ± 0.0096
Δz^4_{wl}	$-0.025^{+0.021}_{-0.018}$	-0.025 ± 0.019	-0.020 ± 0.020	-0.024 ± 0.018
b^1_{QSO}	$2.24^{+0.19}_{-0.22}$	2.18 ± 0.17	$2.18^{+0.29}_{-0.44}$	$2.26^{+0.23}_{-0.26}$
b^2_{QSO}	2.42 ± 0.22	2.36 ± 0.19	$2.28^{+0.32}_{-0.42}$	$2.48^{+0.28}_{-0.33}$
A_{IA}	$0.28^{+0.16}_{-0.19}$	0.31 ± 0.19	$0.28^{+0.17}_{-0.19}$	$0.31^{+0.17}_{-0.20}$
α_{IA}	0.2 ± 2.3	0.1 ± 2.5	$-0.2^{+2.5}_{-3.2}$	0.0 ± 2.5

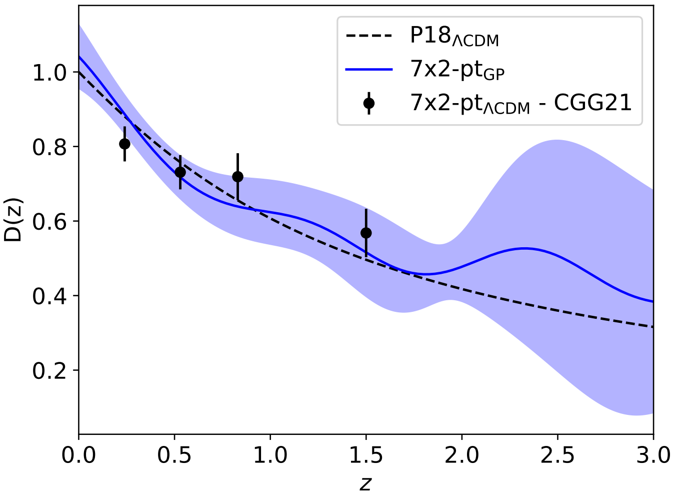

Figure 4.12: Comparison between the reconstructed growth factor in this work using a GP (blue) and the CGG21 reconstruction based on splines (black dots) of the same 7×2-pt data. We also overplot the P18 ΛCDM prediction for the growth factor (black dashed line) for reference.

4.4 Conclusions

In this work, we have presented `LimberJack.jl`, an auto-differentiable cosmological code to compute angular power spectra fully written in `Julia`. The goal of `LimberJack.jl` is to enable the use of gradient-based inference methods in cosmological analyses. `LimberJack.jl` core strength's are:

- **Auto-Differentiablility**: Every step in between the input cosmological parameters and the output theoretical prediction in `LimberJack.jl` is compatible with `Julia`'s auto-differentiation (AD) libraries `ForwardDiff.jl` and `ReverseDiff.jl`. These methods result in gradients up to an order of magnitude faster than when using finite differences. This is the key feature that makes the use of gradient-based inference methods possible with `LimberJack.jl`

- **Speed**: `LimberJack.jl` is equipped with a native implementation of the matter power spectrum emulator `EmuPk` [216] which makes it orders of magnitude faster than `CLASS` or `CAMB`. When the emulator is not used, `Julia`'s C-like speed makes the performance of `LimberJack.jl` comparable to those of other cosmological codes written in C or FORTRAN.

- **Accuracy**: The `LimberJack.jl` theoretical predictions have been thoroughly tested against those of the well-established cosmological code `CCL`, as well as the quality of the AD gradients against the more costly finite difference results.

- **Interoperability**: Thanks to its modular structure, `LimberJack.jl` can be easily interfaced with other `Julia` libraries to increase its capabilities. For example, `LimberJack.jl` can be interfaced with the `Bolt.jl` library [210] to gain access to the first auto-differentiable Boltzmann code.

Furthermore, we presented two examples of how `LimberJack.jl` can be employed to perform present and future cosmological analyses using the gradient-based inference algorithm NUTS [77], a self-tuning formulation of the Hamiltonian Monte Carlo algorithm. In the first example we reproduced two analyses of DES Y1 3x2-point data performed using the well established `Cobaya` library to ensure the reliability of `LimberJack.jl`. `LimberJack.jl`'s constraints were found virtually identical to those obtained with `Cobaya` regardless of whether the matter power spectrum was computed using the Eisenstein and Hu [25, 26] formula or `EmuPk` [216] to emulate `CLASS` [22]. Moreover, the NUTS sampler proved to be 1.5 times more efficient (measured as effective samples per likelihood call) than the MH sampler used by Cobaya. However, this improvement in efficiency proved to be not enough to compensate for the cost of computing the gradient of the likelihood despite using AD methods. Further work is necessary to determine the point at which gradient-based inference methods out-weight the cost of computing the likelihood gradient in angular power spectra analyses.

In the second example, we showcased the unique capabilities of `LimberJack.jl` by performing a Gaussian process (GP) reconstruction of the growth factor across redshift adding to a total of 128 parameters. In order to constrain this model we employed a combination of DES Y1 galaxy clustering and weak lensing data, eBOSS QSO's and CMB lensing (referred to as 7×2-pt in the text) as well as a collection of the latest RSD measurements of $f\sigma_8$. We started by considering a ΛCDM analysis of the aforementioned data to establish a baseline for the GP reconstruction. Our ΛCDM analysis of the 7×2-pt data found $S_8 = 0.789 \pm 0.016$. Adding the RSD data yielded $S_8 = 0.793 \pm 0.015$. Regardless of whether or not RSD data are included, our results are in 2σ disagreement with the Planck 2018 results which found $S_8 = 0.832 \pm 0.013$ [See Tab. 1 of 13]. Performing the GP reconstruction instead yielded $S_8 = 0.839 \pm 0.067$ when 7×2-pt alone is considered and $S_8 = 0.793 \pm 0.017$ when RSD data is included. Looking at the reconstructed growth factor we observed a reasonable agreement between the GP and the Planck 2018 ΛCDM prediction regardless of the data combination used. However, including RSD data significantly smoothed the reconstruction of the growth factor, disfavouring large oscillations. Moreover, it improved the constraints on the reconstructed growth factor by 20% on average across redshift. This stresses the importance of including RSD data in future cosmological analyses, specially given the up coming DESI survey [198]. In terms of sampling efficiency, our GP analysis of 7×2-pt +RSD data using the gradient-based NUTS sampler managed to achieve the same sampling efficiency as our reference ΛCDM analysis of 7×2-pt data. Moreover, `LimberJack.jl`'s AD methods reduced the cost of the likelihood's gradient by an order magnitude with respect finite differences. In combination, these two achievements made a previously unfeasible analysis computationally possible, taking $O(10^2)$ CPU hours to reach convergence.

Auto-differentiable and gradient-based inference methods will play a crucial role speeding up future cosmological analyses as well as enabling entirely new science. For instance, analyses of multiple auto- and cross-correlations between stage-IV surveys may contain up to a hundred free parameters. This will be the case even when performing traditional ΛCDM analyses with minimalistic modelling of systematics simply due to the large number of tomographic bins that will be involved. Future surveys will, however, provide unprecedented measurements of small scales. In order to fit these scales and further our understanding of non-linear cosmology, more complex modelling of baryonic effects will have to be included. Similarly, the constraining power provided by the new data will enable analyses with beyond ΛCDM physics as well as model-agnostic reconstructions such as the one presented in CGG21 and this work which will have a similar effect in the number of free parameters. In addition to this, gradient-based inference methods are already indispensable to undertake field level inference cosmology [225–227] and they will become more so in the future.

Finally, while `LimberJack.jl` is already a fully functional tool, there are several avenues for future improvement:

1. **Improved predictions**: the methods currently implemented in `LimberJack.jl` provide enough accuracy to analyse DES Y1 data. However, these methods will need to be improved in order to analyse DES Y3 data or future data sets such as LSST. Here we present a non-exhaustive list of possible extensions:
    - Non-linear corrections to the matter power spectrum beyond the `Halofit` formula.
    - Small scale baryonic effects on the galaxy and matter power spectra.
    - Angular power spectra beyond the Limber approximation [206].
    - Scale-dependent growth of structure.
    - Perturbatory expansion models for the matter-galaxy bias [228].

2. **Parallelisation**: currently the threading parallelisation of `LimberJack.jl` is suboptimal. This is because the default `Julia` threading parallelisation scheme does not handle the shared memory between the threads efficiently enough. At the moment computing resources are best spent running different instances of `LimberJack.jl` in parallel as opposed to parallelising one instance. Future works could study alternative parallelisation schemes or manually managing the memory between the threads to improve the multi-core performance of `LimberJack.jl`.

3. **GPUs**: `LimberJack.jl` currently cannot run on GPUs which are known to significantly speed-up cosmological inference. This is due to `LimberJack.jl` performing scalar indexing operations at several points of the computation of angular power spectra. Future work could study how to bypass these operations and to make `LimberJack.jl` compatible with `Julia`'s GPU libraries such as `CUDA.jl`.

4. **Backwards-AD**: currently `LimberJack.jl`'s preferred AD mode is forward-AD. However, statistical inference preferred AD mode is backwards-AD, specially as the number of parameters increases. Future works could look into making `LimberJack.jl` compatible with the latest `Julia` AD libraries such as `Zygote.jl` or `Enzyme.jl` to implement efficient backwards-AD.

Chapter 5

Analytical Marginalisation over Nuisance Parameters in Tomographic Analyses of Large Scale Structure

5.1 Introduction

As discussed in Sect. 1.1.5, unprecedentedly precise observations in Cosmology are driving equally as precise models for different systematic effects. These models for systematics are parametrised in terms of additional free parameters which threat to overwhelm our current inference methods. These parameters are often referred to as nuisance parameters. This is because we are normally not interested in learning their particular distribution, but rather in their impact of the cosmological parameters of interest.

This problem can be addressed through various approaches. In Chapter 4, we focused on gradient-based sampling algorithms, such as Hamiltonian Monte-Carlo [76, 77]. In this chapter, we will focus on the analytical marginalisation schemes, namely, the Laplace approximation described in Sect. 1.2.3.

The goal of this chapter is to exhaustively validate this approximate marginalisation scheme in the context of 3x2-pt and cosmic shear analyses. We will start by proving that we are able to replicate the constraints on cosmological parameters of DES-Y1 3×2-pt analysis [51] using the Laplace approximation. We will also show that the approximation holds for futurist data-sets that resemble what Stage-IV surveys will provide us with. Then, we will turn our attention to the marginalisation of photometric uncertainties in the radial distribution of galaxies, $p(z)$, in cosmic shear analyses. We consider both a simple parametrisations of these uncertainties, in terms of shifts to the mean of the distribution, as well as a completely general, non-parametric model. In the later model, we treat the amplitude of the $p(z)$ in narrowly-spaced intervals of z as free parameters, leading to a model with more than ~ 100 nuisance parameters. In order to numerically marginalise over such large parameter spaces, we use the recently developed `LimberJack.jl` (See Chapter 4) in combination with gradient-based inference algorithms (See Sect. 1.2.2). Similarly to the 3x2-pt case, we also consider a futuristic Stage-IV survey, where photometric redshift uncertainties will likely make up a large fraction of the total error budget.

It is worth noting that similar methods have been put forward in the past [e.g., 82], with a variety of applications in mind [229]. Moreover, analytical marginalisation schemes for photometric redshift uncertainties have already been proposed in the literature. In Stölzner *et al.* [230] an analytic marginalisation scheme for photometric redshift uncertainties was proposed based on Gaussian mixture models and applied to the analysis of KV450 data [231]. In Zhang *et al.* [232] a resampling approach to marginalise over these uncertainties was proposed and applied to the analysis HSC data. Similarly, Reischke [233] proposed an approach based on functional derivatives was proposed and applied to Euclid-like [201] and KV450 [231] data. Alternatively, Cordero *et al.* [234] proposed an approach using hyperrank, a method based on ranking discrete samples from the space of all possible redshift distributions of discrete realisations, and applied it to DESY3 data [235].

This chapter is structured as follows. In Sect. 5.2 we describe how the Laplace approximation is applied in practice to the different nuisance parameters of 3x2-pt analyses. Sect. 5.3 presents the Dark Energy Survey data used to produce realistic source redshift distributions and their associated uncertainties, as well as the models used to simulate future datasets. Finally in Sect. 5.4 presents our results, quantifying the performance of analytical marginalisation methods in the two scenarios. Finally, we present our conclusions in Sect. 5.5.

5.2 Methods

Nuisance parameters can be divided into two groups based on their prior distributions: calibrated and non-calibrated parameters. The non-calibrated parameters can only be constrained by the data and, as such, typically have largely non-constraining priors. On the other hand, we can place tighter priors on the calibrated parameters, either by accurately characterising the instrument measurements or by using independent external observations. It is important to stress that whether a parameter is calibrated or not is not inherent to the parameter but to whether a sufficiently tight prior on it can be placed (e.g. from external data). In the context of 3x2-pt analyses, mass-galaxy biases [48, 49] and the impact of galaxy intrinsic alignments [52] are often a non-calibrated systematic. On the calibrated side, the two best examples are multiplicative shape measurement systematics, and the uncertainties in the redshift distribution of the target source galaxies [221, 230–232, 236]. In this section, we will discuss how the Laplace approximation introduce in Sect. 5 can be applied in practice to these two types of parameters.

5.2.1 Calibratable systematics

In the presence of tight priors, which we will further assume to be Gaussian, the nuisance parameters may not stray far from their prior mean. In that case, we can Taylor-expand the theory prediction as in Eq. (1.137)

$$\mathbf{t}(\boldsymbol{\Omega},\boldsymbol{n}) = \mathbf{t}_0(\boldsymbol{\Omega}) + \mathsf{T}\,(\boldsymbol{n}-\boldsymbol{n}_p), \tag{5.1}$$

where $\boldsymbol{n}_p$ is the prior mean, and

$$\mathbf{t}_0 \equiv \mathbf{t}(\boldsymbol{\Omega},\boldsymbol{n}_p),\;\; \mathsf{T} \equiv \left.\frac{\partial\mathbf{t}}{\partial\boldsymbol{n}}\right|_{\boldsymbol{n}_p}. \tag{5.2}$$

Since now the theory is linear with respect to $\boldsymbol{n}-\boldsymbol{n}_p$, we can then follow the procedure in Section 1.2.3.3 to analytically marginalise over those parameters. As we discussed, we simply modify the covariance of the data vector (in this case, a collection of power spectra) as in Eq. (1.143), and then sample the resulting Gaussian likelihood evaluating $\mathbf{t}$ at the prior mean of $\boldsymbol{n}$. Two important things should be noted. First, in doing this, we have neglected the parameter dependence on $\boldsymbol{\Omega}$ of the Laplace term, given by

$$\log\left[\det\left(\mathsf{T}^T\mathsf{C}^{-1}\mathsf{T}+\mathsf{C}_n^{-1}\right)\right] \tag{5.3}$$

and therefore omit it from the calculation, as it only adds a multiplicative constant in this approximation. If the prior is sufficiently tight, this term is dominated by the constant C_n^{-1} contribution, so the approximation is reasonable. Second, since the modified covariance matrix (Eq. (1.143)) now involves a term of the form $\mathsf{T}\mathsf{C}_n\mathsf{T}^T$, in principle it depends on $\boldsymbol{\Omega}$ through T. Calculating the covariance at each point in the likelihood may be computationally costly, depending on the size of T. Instead, we will simply evaluate T at the best-fit value of $\boldsymbol{\Omega}$, and ignore all parameter dependence on $\boldsymbol{\Omega}$ of the covariance. It was shown in Kodwani *et al.* [237] that the parameter dependence of the covariance can generally be neglected and, furthermore, for a sufficiently tight prior, the $\mathsf{TC}_n\mathsf{T}^T$ contribution should be subdominant. Hadzhiyska *et al.* [238] also showed that the choice of fiducial $\mathbf{\Omega}$ does not affect the final results as long as they are close to the centre of the posterior distribution. Adopting these two approximations (in addition to the Taylor expansion of $\mathbf{t}$) is therefore well justified and, as we will show in Section 5.4.1.1, leads to accurate results.

We stress that we can apply the same procedure to any other parameter that appears to behave linearly in the theoretical model. This is the case for any other nuisance parameters with a sufficiently tight prior, and in fact this procedure is routinely used for multiplicative bias parameters in cosmic shear analyses [231]. We note that the general procedure of using the Laplace approximation is exact in the case of truly linear parameters, regardless of their priors. A good example of this is the amplitude of shot noise or stochastic contributions to galaxy clustering auto-correlations [54].

Of these calibrated systematics the dominant source of uncertainty in photometric surveys is the accuracy of redshift distributions, which are known to strongly affect the accuracy of cosmological constraints. The vital quantity to determine is the redshift distribution of each tomographic sample of galaxies. Photometric galaxy uncertainties, $p(z)$, can be calibrated through various methods, e.g.: weighted direct calibration with a sufficiently complete spectroscopic sample [37, 38], clustering redshifts [39–42], and shear ratios [43, 44, 239]. This typically leads to relatively tight priors on the $p(z)$, but the residual uncertainties in this prior must be propagated into the final parameter constraints.

To characterise these uncertainties, we will make use of two different methods, which encompass the range of model complexity we may reasonably expect from current and future data.

- **Method 1: z shifts.** Most cosmic shear analyses to date [53, 60, 231, 240, 241, among others] have summarised the uncertainty in the calibrated $p_\alpha(z)$ into a single parameter Δz^α that shifts the mean of the redshift distribution; i.e. let $\hat{p}_\alpha(z)$ be the best-guess redshift distribution. The true redshift distribution is then given by Eq. 1.94. A prior on Δz^α can be derived using the calibration methods listed above. We will refer to this method as *parametric*.

  This simple model turns out to be relatively well suited to describe the impact of $p(z)$ uncertainties in the case of cosmic shear data [242] even from stage-IV surveys [243]. Since weak lensing is a radially cumulative effect, the amplitude of the weak lensing kernel (Eq. 1.90) is mostly sensitive to the mean redshift of the sample, and thus much of the effect on cosmic shear observables is well described by this parameter [244].

Other modes of $p(z)$ uncertainty, such as the distribution width, may be more relevant for galaxy clustering observables, or for the intrinsic alignment contribution to cosmic shear. Near-future cosmic shear samples may indeed require a more sophisticated description of the $p(z)$ uncertainty, and thus we turn to a more general method.

- **Method 2: $p(z)$ bin heights.** Most $p(z)$ calibration methods (e.g. direct calibration or clustering redshifts) will produce a binned measurement of the $p(z)$ with deterministic redshift bin ranges, and uncertain bin heights. The most general method to propagate these uncertainties is therefore to treat each bin height $p_i \equiv p(z_i)$ as a free parameter in the model, with a prior given by the calibration uncertainties. The latter may be in the form of individual 1σ errors for each bin height, if the uncertainties are approximately uncorrelated, or a full covariance matrix covering all bin heights.

  The resulting parametrisation thus sidesteps any attempt at summarising the uncertainty into effective parameters, and thus we will refer to this method as *non-parametric*. The method therefore fully propagates all calibration uncertainties into the final constraints with minimal approximations.

The key practical difference between both methods, in the context of error propagation, is the additional complexity they incur. The parametric approach (Method 1) introduces one free parameter per redshift bin. For $O(5)$ bins, this is already enough to significantly impact the performance of standard MCMC algorithms. In turn, the non-parametric approach (Method 2) introduces tens or hundreds of parameters per redshift bin, and one must resort to advanced sampling methods in order to fully explore the resulting model without assumptions.

5.2.2 Non-calibratable systematics

Most sources of astrophysical uncertainty cannot be well-constrained from external data, and thus must be constrained at the same time as the cosmological parameters, and marginalised over. In this case, the linearisation described in the previous section is not appropriate, and we must resort to numerical methods in order to obtain $\boldsymbol{n}_*$ and the Laplace contribution in Eq. (1.128).

In the model introduced in Sect. 1.1.5, these astrophysical uncertainties are described by the bias and intrinsic alignment parameters. Comparing Eq. (1.71) and Eq. (1.100), we see that we can describe both tomographic galaxy clustering and cosmic shear as a projected tracer with radial kernels q^α with the generic form

$$q^\alpha = \epsilon_q \, q^\alpha_M + \sum_k b^q_{\alpha,k} \, q^\alpha_k \,, \tag{5.4}$$

where $b^u_{\alpha,k}$ are bias parameters (specifying the tracer type q), q^α_M and q^α_k are projected quantities that depend only on cosmological observables (matter overdensities, comoving distances etc.), and ϵ_q is a Boolean variable that is either 1 if the tracer contains an unbiased contribution (as is the case of cosmic shear), and 0 otherwise (as is the case of galaxy clustering). The index α in Eq. (5.4) runs over the redshift bins, which allows for the general case of having redshift-dependent bias functions. If this is not the case, $b^u_{\alpha,k} \equiv b^u_k$ can optionally be assumed to not vary across redshift bins.

Although the bias/IA description used here covers a wide range of state-of-the-art physical models used in current 3×2-pt analyses, it is mathematically exceptionally simple. From Eq. (5.4), we see that the cross-correlation between any two such tracers (q^α, g^β) is a simple quadratic function of the bias parameters:

$$\begin{aligned} C_\ell^{q^\alpha, g^\beta} &= \epsilon_q \epsilon_g C_\ell^{q^\alpha_M, g^\beta_M} + \sum_i b^q_{\alpha,i} \epsilon_g C_\ell^{q^\alpha_i, g^\beta_M} \\ &+ \sum_j \epsilon_q b^g_{\beta,j} C_\ell^{q^\alpha_M, g^\beta_j} + \sum_{i,j} b^q_{\alpha,i} b^g_{\beta,j} C_\ell^{q^\alpha_i, g^\beta_j} \,. \end{aligned} \tag{5.5}$$

Here, $C_\ell^{q^\alpha_{M/i}, g^\beta_{M/j}}$ are the power spectra between the cosmological projected fields $q^\alpha_{M/i}$ and $g^\beta_{M/j}$, defined in Eq. (5.4), and the sums run over the associated bias terms. Since these only involve radial projections of purely cosmological quantities, they can be treated as templates that only depend on the cosmological parameters. Note that, in principle, these templates also depend on the calibratable nuisance parameters described in the previous section (e.g. through the modification in the radial kernels due to $p(z)$ uncertainties). However, we assume that we have been able to marginalise over these analytically as we described above, and therefore they can be treated as fixed for all intents and purposes.

The first derivative of the power spectrum with respect to the bias parameters is thus a linear polynomial:

$$\begin{aligned} \frac{\partial C_\ell^{q^\alpha,g^\beta}}{\partial b^r_{\gamma,k}} &= \delta^K_{\alpha,\gamma}\delta^K_{q,r}\left[\epsilon_g C_\ell^{q^\alpha_k,g^\beta_M} + \sum_j b^w_{\beta,j} C_\ell^{q^\alpha_k,g^\beta_j}\right] \\ &+ \delta^K_{\beta,\gamma}\delta^K_{g,r}\left[\epsilon_q C_\ell^{q^\alpha_M,g^\beta_k} + \sum_i b^u_{\alpha,i} C_\ell^{q^\alpha_i,g^\beta_k}\right], \end{aligned} \tag{5.6}$$

where δ^K is the Kronecker delta, and $b^r_{\gamma,k}$ denotes the kth bias term of tracer type r (i.e., galaxy overdensity or shear) in redshift bin γ. Finally, the Hessian is constant with respect to the bias parameters

$$\begin{aligned} \frac{\partial^2 C_\ell^{u^\alpha,w^\beta}}{\partial b^r_{\gamma,k}\partial b^s_{\sigma,m}} &= \\ &\left[\delta^K_{\gamma,\alpha}\delta^K_{\sigma,\beta}\delta^K_{r,u}\delta^K_{s,w} + \delta^K_{\gamma,\beta}\delta^K_{\sigma,\alpha}\delta^K_{r,w}\delta^K_{s,u}\right] C_\ell^{r^\gamma_k,s^\sigma_m}. \end{aligned} \tag{5.7}$$

These expressions are remarkably simple and fast to evaluate, and thus computing the χ^2 and its derivatives (needed for minimisation, and to calculate the Laplace contribution) can be done extremely efficiently. As we will see, in practice we find that finding the minimum of the χ^2 takes $O(10-100)$ Gauss-Newton iterations, each of which is orders of magnitude faster than recomputing the power spectrum templates when changing cosmological parameters. Computing the Laplace approximation to the marginalised posterior at each sample of the cosmological parameters is therefore virtually equivalent to evaluating the joint posterior for new cosmological + nuisance parameters if using brute-force marginalisation.

5.3 Data

5.3.1 DES Y1

In order to evaluate the performance of the Laplace approximation, we make use of data from the first-year of the Dark Energy Survey [DES-Y1 220] described in Sect. 4.3.1. We do so to demonstrate that the method can be successfully implemented in real data, with real-life complications (e.g. noisy $p(z)$s, numerical covariances, astrophysical and observational systematics). As advertised in Sect. 5.1, we will reproduce the 3x2-pt analysis. Moreover, we will also study the validity of the approximation in the context of cosmic shear analyses. In order to do so, we employ the angular power spectra provided by Nicola *et al.* [245]. A full description of the methods used to estimate these power spectra, and their associated covariance matrix, from the

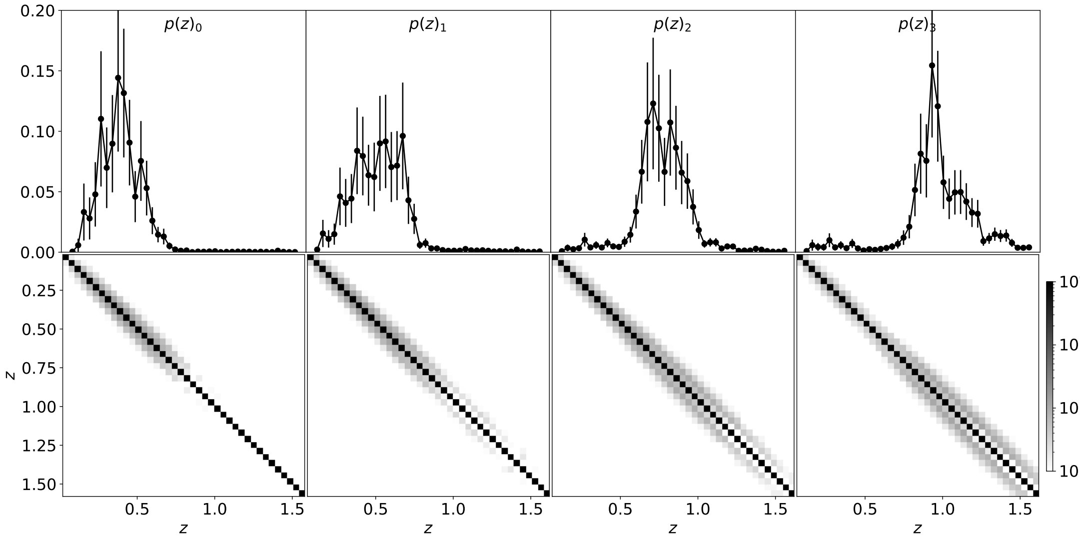

Figure 5.1: Top row: normalised galaxies' redshift distributions for each of the 4 redshift bins. Bottom row: correlation matrix obtained using the DIR algorithm for each of the 4 galaxies' redshift distributions. Note that for visualisation purposes we display the absolute values of the each correlation matrix in logarithmic scale. In this plot, we can see that the covariance matrices obtained through the DIR algorithm are mostly diagonal.

DES-Y1 data is provided by the authors. We use the calibrated redshift distributions of the `METACALIBRATION` sample provided by García-García *et al.* [246]. The $p(z)$'s were estimated via direct calibration [DIR 37], using the COSMOS 30-band catalogue [247] as a calibrating sample. The uncertainties of the measured redshift distributions were estimated analytically, as described in García-García *et al.* [246], accounting for both shot noise and sample variance, and represent a realistic level of $p(z)$ uncertainty achieved by current existing datasets. The redshift distributions were sampled on 40 bins of width $\delta z = 0.04$ covering the range $0 \leq z \leq 1.6$. Fig. 5.1 shows, in the first row, the redshift distributions of the four `METACALIBRATION` samples and their statistical uncertainties. Note that we estimated the full covariance matrix of the $p(z)$ bin heights. As can be seen in the bottom panels of Fig. 5.1, the covariance is dominated by the diagonal. Oddly, it is possible to observe a lack of nearest-neighbours correlations, specially at the high redshift tails of the two last tomographic bins. However, the calibration of these distributions is beyond the scope of this work. A full description of the methods used to estimate these power spectra, and their associated covariance matrix, from the DES-Y1 data is provided by the authors.

5.3.2 Synthetic Stage-IV data

We also consider a futuristic, idealised data set that resembles the characteristics of LSST. It is important to test our method in the low-noise regime, where the inferred posterior is even more sensitive to redshift distribution uncertainties or, in general, degeneracies between cosmological and nuisance parameters, and where the final error budget is more dominated by these effects.

To define the clustering and shear samples we follow the same procedure outlined in Nicola *et al.* [248]. The shear sample is defined following the LSST Science Requirements Document [243] (see Appendices D1 and D2). We divide this redshift distribution into 5 bins in photometric redshift space, each containing the same number of sources. We assume a Gaussian photometric redshift uncertainty with standard deviation $\sigma_z = 0.05(1+z)$, which thus defines the true-redshift tails of the distribution in each tomographic bin. The sample has an overall angular number density of 27 gals. arcmin^{-2}. For galaxy clustering, we define a sample extending out to $z \sim 1.5$ with a total density of $4\,\text{gals.}\,\text{arcmin}^{-2}$ (as would be expected of an LRG-like sample for LSST). This number density and the associated redshift distribution were estimated using measurements of the luminosity function for red galaxies as described in Alonso *et al.* [249]. The sample was divided into 6 redshift bins equi-spaced in photometric redshift space, and assuming a photometric redshift uncertainty of $\sigma_z = 0.02(1+z)$. To simplify the analysis, we assume a constant linear galaxy bias $b_1 = 1$, and set all higher-order bias coefficients to zero. The results obtained in the next section should be largely insensitive to this choice.

For simplicity, we use a Gaussian covariance to describe the uncertainties of the resulting data vector, calculated assuming a sky fraction $f_{\text{sky}} = 0.4$. The LSST data vector was generated assuming a true cosmology with parameters

$$(\Omega_m, \Omega_b, h, n_s, \sigma_8) = (0.3,\ 0.05,\ 0.7,\ 0.96,\ 0.8), \tag{5.8}$$

where Ω_m and Ω_b are the total matter and baryon fractions, h is the reduced Hubble parameter, n_s is the scalar spectral index, and σ_8 is the standard deviation of linear density perturbations smoothed on spheres of radius $8\,\text{Mpc}\,h^{-1}$ at redshift $z = 0$.

5.4 Results

5.4.1 DES Y1 3x2-pt

Let us start by validating the Laplace approximation. We will do so by reproducing the 3×2-pt DES Y1 analysis as well as a futuristic LSST-like 3-pt analysis based

Parameter priors			
Parameter	Prior	Parameter	Prior
Cosmology		**Redshift calibration**	
Ω_m	$U(0.07, 0.8)$	Δz_s^1	$\mathcal{N}(\Delta z^1_{s,*}, 0.016)$
Ω_b	$U(0.03, 0.07)$	Δz_s^2	$\mathcal{N}(\Delta z^2_{s,*}, 0.013)$
h	$U(0.55, 0.91)$	Δz_s^3	$\mathcal{N}(\Delta z^3_{s,*}, 0.011)$
n_s	$U(0.87, 1.07)$	$\Delta z_\mathrm{s}^{4,5}$	$\mathcal{N}(\Delta z^{4,5}_{s,*}, 0.022)$
σ_8	$U(0.5, 1.1)$	Δz_g^1	$\mathcal{N}(\Delta z^1_{g,*}, 0.007)$
Bias parameters		Δz_g^2	$\mathcal{N}(\Delta z^2_{g,*}, 0.007)$
b_1^i	$\mathcal{N}(1.5, 100)$	Δz_g^3	$\mathcal{N}(\Delta z^3_{g,*}, 0.006)$
$b^i_{2,s,\nabla}$	$\mathcal{N}(0, 100)$	$\Delta z_\mathrm{g}^{4,5,6}$	$\mathcal{N}(\Delta z^{4,5,6}_{g,*}, 0.01)$
$A_{\mathrm{IA},0}$	$\mathcal{N}(0, 100)$	$w^i_{z,\mathrm{g}}$	$\mathcal{N}(1.00, 0.08)$

Table 5.1: Prior distributions for the nuisance parameters entering our "3×2pt" analysis for each tracer. $U(a, b)$ and $\mathcal{N}(\mu, \sigma)$ describe a uniform distribution with boundaries (a, b) and a Gaussian distribution with mean μ and variance σ, respectively. The index i in b^i_g and m^i runs over the different redshift bins. Δz_* denotes the deviation from zero of the central/best-fit value of each redshift uncertainty parameter.

on synthetic data described in Sect. 5.3. We will start by marginalising over the calibratable nuisance parameters, namely the multiplicative bias and the photometric uncertainties. Then we will marginalise over the remaining non-calibratable nuisance parameters, i.e. mass-galaxy biases and intrinsic alignments.

To obtain constraints on cosmological parameters from both real and synthetic data, we will assume that the data vector (i.e. the clustering and shear power spectra), follows a Gaussian likelihood, with a parameter-independent covariance. The model will be described by 5 cosmological parameters, listed in Eq. (5.8), one or four linear parameters for each clustering redshift bins (for linear and PT bias respectively), one intrinsic alignment amplitude, one redshift shift parameter for each clustering and cosmic shear bin, and one redshift distribution width parameter for each clustering bin. The priors used for all parameters are provided in Table 5.1. The priors on cosmological parameters are roughly based on the choices made for the DES-Y1 analysis, except we sample over σ_8 instead of the scalar spectrum amplitude A_s. The priors on the redshift shift parameters are based on the calibration of the DES-Y1 data [221], and thus represent achievable calibration levels. In addition to the usual shift parameters, Δz, we also consider a series of width parameters, w_z, such as the final galaxy redshift distribution is given Eq. 1.94 . The priors on the redshift width parameters are commensurate with those used in the DES Year-3 analysis [61] (the DES-Y1 analysis did not introduce width parameters). Note that these width parameters are an extension of the *method 1* described in Sect. 5.2. Moreover, the no multiplicative biases in the shear data were considered. Finally, we place an uninformative Gaussian prior on all the bias parameters, centred at zero and with a standard deviation of 100. The choice of using a very broad Gaussian prior as opposed to simply a flat prior is intended to enforce a smooth distribution as a function of these parameters, and to potentially aid the Gauss-Newton iterator when minimising the χ^2.

We employ the `cobaya` MCMC sampler [250, 251] with a convergence condition that the Gelman-Rubin diagnostic, R, ought to satisfy $R - 1 < 0.01$. When using the Laplace approximation, we minimise the χ^2 over the nuisance parameters using a Gauss-Newton iterator, using the analytical derivatives with respect to bias parameters as described in Section 5.2.2, and modify the log-probability to be sampled by `cobaya` to be that of Eq. (1.128). We find that, in order to reduce the number of steps taken by the Gauss-Newton iterator, it is useful to determine a well educated global best-fit for the full parameter space before taking any samples, and to start the iterator from the corresponding best-fit value of the nuisance parameters.

Throughout, we made use of the fitting formula of Eisenstein and Hu [26] to calculate the linear matter power spectrum. We do this to speed up the calculations, and we have verified that the results obtained on the DES-Y1 data are insensitive to this choice compared to using a Boltzmann solver such as `CLASS` [252]. The non-linear matter power spectrum is then computed using `Halofit` [28].

5.4.1.1 Calibratable nuisance parameters

We begin by focusing on calibratable systematics, for which we will follow the procedure described in Section 1.2.3.3: we will linearise the dependence of the theoretical prediction with respect to these parameters, for which a relatively tight prior can be obtained, and simply modify the covariance matrix as in Eq. (1.143), with T evaluated at a fixed set of parameters (fixing all cosmological parameters to the best-fit values found by *Planck* [13] and all bias parameters to their best-fit values). At this stage, we will thus only consider the nuisance parameters describing the uncertainty in the redshift distributions of the tracers under study, described in the right column of Table 5.1. All other parameters (cosmological, bias, and intrinsic alignment parameters) will for now be marginalised "brute-force" (i.e. treating them as free parameters in the MCMC chains). For simplicity, we will consider only linear bias, using scales $k < 0.15\,\mathrm{Mpc}^{-1}$, as described in Section 1.2.3.3.

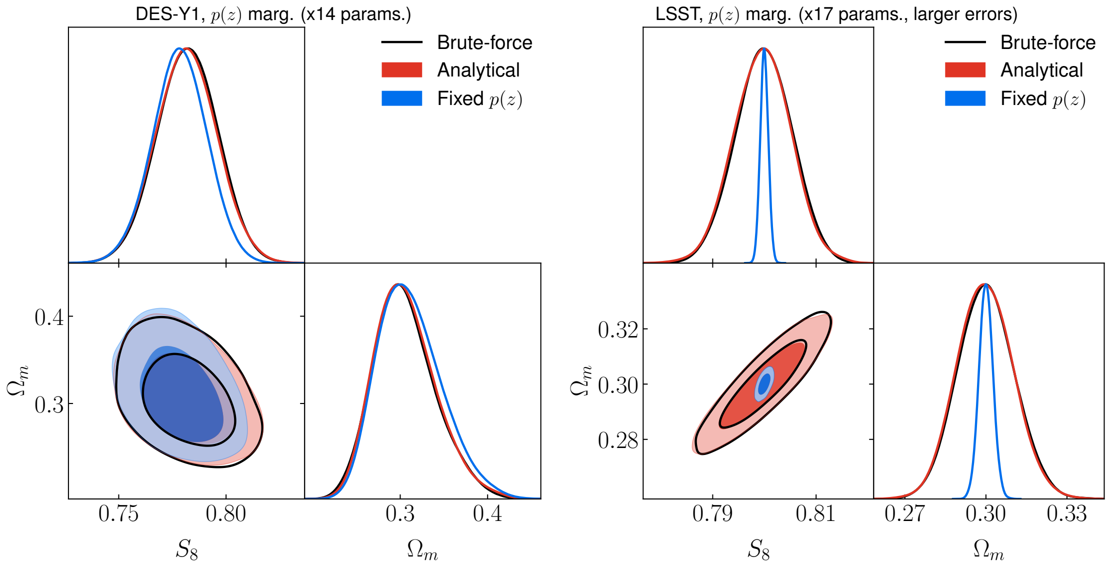

Figure 5.2: Contours comparing brute-force (black) with analytic marginalisation over photo-z uncertainties (red; see Sect. 1.2.3.3). Results are shown for the DES-Y1 data (left) and the LSST-like data (right). We find that the contours are virtually unchanged in both cases, demonstrating the benefit of using an efficient analytic marginalisation scheme. We also show for posterity the result of assuming negligible error on the photo-z distributions (blue). In the case of LSST-like data the errors on the nuisance parameters were quadrupled. Thus, we assert that the Laplace approximation holds even in the conservative scenario of large photometric uncertainties in futuristic data.

We first compare the performance of the method when applied to the 3×2pt analysis of DES-Y1 data (Section 5.3). The results are shown in the left panel of Fig. 5.2; with the exact results shown as black contour lines, and the results of the analytical marginalisation shown in red. We find that using the analytic marginalisation technique not only yields contours that are almost indistinguishable from those obtained by the traditional approach, but also does so significantly faster (by a factor ~ 10) with many fewer parameters. The blue contours in the figure show the constraints found by fixing the nuisance parameters to their prior means, instead of marginalising over them. In this case, we observe that the redshift distribution uncertainties cause only a mild broadening of the marginalised contours, which is not very challenging for the analytical approximate marginalisation to reproduce.

In order to explore a more challenging scenario, we now move on to the case of an LSST-like 3×2-pt dataset, as described in Sect. 5.3. Moreover, we will assume the same prior uncertainties are 4 times larger than the DES-Y1 ones. This will allow us to quantify the validity of the analytical marginalisation approach in a conservative scenario, in which, in spite of the much higher sensitivity of Stage-IV data, the precision with which we are able to calibrate redshift distributions has not improved with respect to the performance achieved with current data. The reasoning behind this test is two-fold: on one hand, our marginalisation method is an approximation that works in the regime where photometric uncertainties are linearisable and testing when that assumption breaks is essential; on the other hand, it is likely that, at the highest redshifts, and for the faintest samples, the $p(z)$ uncertainties will be somewhat larger for LSST than those of current surveys, especially at high redshifts. Hence, quadrupling the errors is an important worst-case scenario to consider.

It is important to note that the results in this section are not meant to be interpreted as forecasts on the constraining power of LSST on cosmological parameters, but only as a quantification of our ability to analytically marginalise over photometric uncertainties when inferring the underlying cosmology. A more thorough analysis would include a realistic treatment of the LSST true redshift distribution and noise, and a more careful treatment of galaxy bias and intrinsic alignments. As such, the results presented here give us a conservative estimate of the effect of analytic marginalisation on cosmology constraints.

We present the results of the LSST-like 3×2-pt analysis in the right panel of Fig. 5.2. The marginalised constraints on the cosmological parameters are virtually unchanged when we switch from the brute-force to the analytical marginalisation. This latter method is therefore successful at recovering accurate marginalised constraints on cosmological parameters. Also shown in blue are the constraints found assuming perfect knowledge of the redshift distributions (i.e. fixing all $p(z)$ parameters). In this case, the uncertainties on the redshift distributions have a much larger effects than for the DES-Y1 data, inflating the uncertainties on S_8 significantly. This is partially due to quadrupling the uncertainty in the redshift nuisance parameters. In spite of this, we find that the analytic marginalisation method not only yields virtually the same constraints on the cosmological parameters, but does so 3-10 times faster than the traditional approach. This implicitly validates the approximation that a first-order expansion of the theory data vector with respect to a change in redshift distribution is sufficient, even for prior uncertainties on $p(z)$ that are substantially worse than those achieved by current datasets.

5.4.1.2 Uncalibratable Nuisance Parameters

Let us now focus on the bias parameters. In this case, the absence of an informative prior prevents us from linearising the dependence of the theory on these parameters,

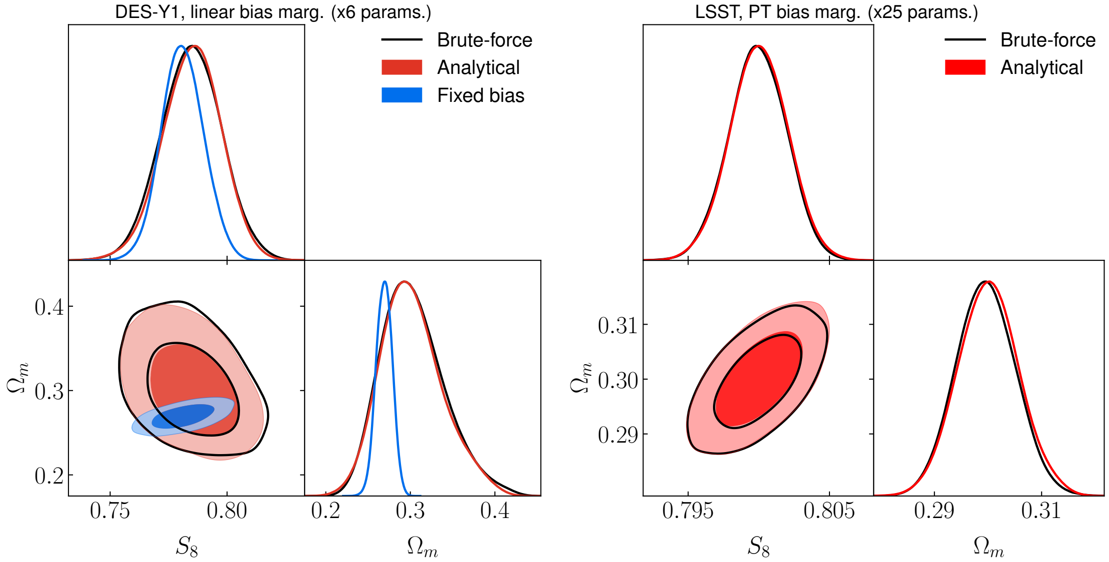

Figure 5.3: Same as Fig. 5.2, but we also marginalise over the bias parameters using the Laplace approximation. This reduces the parameter space to only the cosmological parameters without degrading the accuracy of the constraints. The agreement with the full numerical marginalisation (black) is almost perfect.

and we must therefore calculate the profile and Laplace terms numerically.

Similarly to Sect. 5.4.1.1, we make use of the DES-Y1 data, marginalising only over a single linear galaxy bias parameter per clustering bin, as well as an intrinsic alignment amplitude (i.e. a total of 6 nuisance bias parameters). The results are shown in the left panel of Fig. 5.3. The exact marginalised constrains (solid black contours) are accurately recovered by the Laplace approximation (red contours). While the former are obtained by running an MCMC with 11 free parameters (5 cosmological, 6 nuisance parameters), the latter involve only a 5-dimensional parameter space, which is therefore significantly simpler to explore. Concretely, the 5-parameter chain converged 3 times faster than the 11-parameter chain. The blue contours in the same figure shows the constraints obtained after fixing the bias parameters to the best-fit values found by DES [253]. Fixing the galaxy bias shifts the cosmological constraining power from the cosmic shear data to the higher signal-to-noise clustering data, thus significantly reducing the uncertainties. Note that, although the red contours show the result of the full Laplace approximation (i.e. profile + Laplace contributions), the Laplace contribution is negligible, and the profile term is enough to recover the marginalised posterior.

In order to explore the performance of the method with a significantly larger number of nuisance parameters, while still remaining in the regime where these parameters can be well constrained by the data, we now move to the LSST-like synthetic dataset, making use of the second-order perturbative expansion of Eq. (1.96) to describe galaxy bias. In this case, we include 4 free bias parameters in each of the 6 clustering redshift bins, adding up to a total of 25 nuisance parameters when combined with the intrinsic alignment amplitude. The results are shown in the right panel of Fig. 5.3, again as black lines for the brute-force marginalisation, i.e. considering all 30 free parameters, and as red contours for the 5-parameter Laplace approximation. As before, we find that the approximation is able to recover the marginalised constraints almost exactly.

Note that, in both of these cases, besides the bias parameters, we have also marginalised over a number of redshift distribution uncertainty parameters (14 for DES-Y1, 17 for LSST), thus reducing the model dimensionality from 47 and 28 for LSST and DES-Y1, respectively, to only 5 cosmological parameters. Obtaining converged MCMC chains for these 5 cosmological parameters takes approximately 2 hours for the LSST-like dataset, a factor ~ 6 faster than the chains with only the $p(z)$ parameters analytically marginalised over, and a factor ~ 30 times faster than the full brute-force chains. The magnitude of these speed gains, however impressive, must be taken with a pinch of salt. The performance of MCMC sampling may depend significantly upon the design of the likelihood code, and whether it allows the sampler to decompose the space between "fast" and "slow" parameters, over-sampling the former, and making use of "dragging" techniques [254, 255]. The fast-slow split allows one to effectively marginalise over the fast subspace, and becomes particularly powerful in the presence of a large number of nuisance parameters on which the likelihood has a computationally simple dependence (as is the case of the bias parameters in our model). To provide a fairer assessment of the computational gains obtained using the Laplace approximation, we wrote an optimised version of the 3×2-point likelihood that allows `cobaya` to exploit the fast nature of the bias parameters as efficiently as possible (assuming all $p(z)$ parameters are fixed). In this case, the speed-up factor for the cases explored here ranged between ~ 1.5 and ~ 11. Ultimately the performance difference depends on the design of the likelihood code, the complexity in the parameter dependence of the likelihood, and the efficiency of the χ^2 minimisation method used to calculate $\boldsymbol{n}_*$. On this latter point, we find that the Gauss-Newton method used here typically achieves convergence after only 10-20 iterations. Thus finding the best-fit bias parameters is significantly faster than calculating the cosmology-dependent power spectrum templates of Eq. (5.5).

Parameter priors			
Parameter	Prior	Parameter	Prior
Cosmology		**Redshift calibration**	
Ω_m	$U(0.1, 0.9)$	Δz_1	$\mathcal{N}0.0, 0.016)$
Ω_b	$U(0.03, 0.07)$	Δz_2	$\mathcal{N}(0.0, 0.017)$
h	$U(0.55, 0.91)$	Δz_3	$\mathcal{N}(0.0, 0.013)$
n_s	$U(0.87, 1.07)$	Δz_4	$\mathcal{N}(0.0, 0.015)$
σ_8	$U(0.6, 0.9)$	p_i	$\mathcal{N}(\bar{p}_i, \mathsf{C})$
Shear multiplicative bias			
m^i	0.012		

Table 5.2: Prior distributions for the parameters considered in this work. Note that the redshift calibration section contains the priors for both the Δz and $p_\alpha(z)$ models which are not sampled simultaneously.

5.4.2 DES-Y1 Shear Photometric Redshift Uncertainties

Let us now turn our attention to the study of photometric redshift uncertainties in cosmic shear analyses. As in the previous section, we will consider shear data from both the DES-Y1 survey and a LSST-like synthetic data set as described in Sect. 5.3. Moreover, we will study the two uncertainty models described in Sect. 5.2.1 Thus, we will either consider one redshift shift parameter Δz_α for redshift bin, when employing the parametric description of $p(z)$ uncertainties (Method 1), or a set of bin heights for each redshift bin determining $p_\alpha(z)$, when using the non-parametric approach (Method 2). The first case will introduce 4 new parameters to the model, while the latter will introduce $4 \times 40 = 160$ new amplitude parameters.

Table 5.2 shows the parameter priors used for this study. All cosmological parameters take uniform, largely uninformative priors. For simplicity, the multiplicative bias parameters were fixed at the centre of the Gaussian priors from the official analysis of DES-Y1 [51]. When using Method 1 to numerically marginalise over the $p(z)$ uncertainties, we used Gaussian priors on each of the shift parameters Δz_α, following those used by DES-Y1 [51]. When using Method 2 (marginalisation over $p(z)$ bin amplitudes) some considerations about the physical object being statistically modelled need to be taken into account. By virtue of describing a probability distribution, the samples of p_i have to add up to one. One way in which this constraint can be incorporated into the model is by imposing a Dirichlet prior on p_i which naturally imposes this requirement by sampling its values from a simplex[1][256]. However, applying

---

[1] A simplex is a series of numbers that add up to a constant

the Laplace approximation with a Dirichlet prior or auto-differentiating through it is not trivial. An alternative approach is to sample $p(z)$ from a Gaussian distribution and then impose the normalisation constraint a posteriori. This approach naively bypasses the above described problem. However, it is important to understand why. The distribution of galaxies per bin in inherently Poisson. For highly populated bins, the Poisson distribution can be approximated as Gaussian around a non-zero mean. However, this is not true in sparsely populated bins. In the later, the sampled number of galaxies for per bin cannot be negative. Moreover, galaxies removed from one bin need to be added to one of their neighbours. Thankfully, given the photometric calibration of DES Y1, these sparsely populated bin possess very narrow error bars such that they still disfavour values close to zero and we can ignore the above described problems. Thus we can assume a multi-variate Gaussian prior for $p(z)$ with a covariance matrix described in Sect. 5.3 and shown in Fig. 5.1. We then normalise samples a posterirori to enforce the fact that $p(z)$ describes a probability distribution.

For both $p(z)$ uncertainty models, when using analytical marginalisation, we use Eq. 1.144 and modify the covariance as in Eq. 1.143, with P given by the priors described above. When using numerical marginalisation, we simply explore the posterior distribution of the full model, including all the $p(z)$, p_i, parameters. In the case of Method 2, this involves sampling a distribution with 165 parameters, of which the bulk (160 parameters) describe the $p(z)$ uncertainty. This is not feasible for standard Metropolis-Hastings MCMC methods [70, 257] due to the curse of dimensionality. In order to overcome this issue, we write our pipeline using `LimberJack.jl` (See Sect. 4) which thanks to its auto-differentiable methods allows us to use the gradient-based sampler NUTS [NUTS 77] (See Sect. 1.2.2) to beat the curse of dimensionality. In particular, we use the NUTS implementation within the JULIA library for statistical inference `Turing.jl` [222].

5.4.2.1 Linearising Δz

Let us begin the discussion of our results by considering the simplest of the two models of the photometric uncertainties studied in this work, the Δz model (called Method 1 above). As discussed in the previous section, this model introduces 4 new shift parameters Δz (one per redshift bin) in addition to the 5 ΛCDM parameters. All other nuisance parameters are kept fixed. For the DES-Y1 and LSST-like datasets, we will compare the result of analytically marginalising over the Δz parameters against performing the full numerical marginalisation on the corresponding cosmological constraints. In order to quantify the contribution of redshift uncertainties to the total error budget, we will also present results for the case when the Δz parameters are fixed (i.e. assuming perfect knowledge of the redshift distributions).

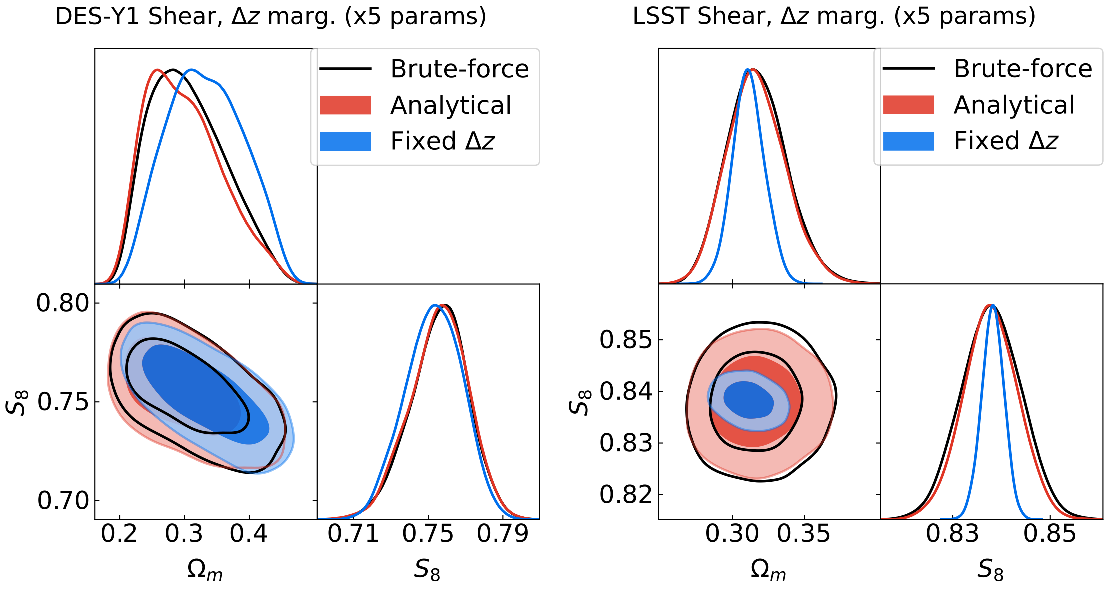

Figure 5.4: Marginalised posterior distributions for the combination of parameters $\Omega_{\rm m}$, σ_8 and S_8 obtained when considering the Δz model for photometric uncertainties for DES-Y1 data. The blue contours correspond to the case where the Δz parameter are fixed. The magenta contours are obtained when numerically marginalising over the Δz parameters. Finally, the black dashed contours are obtained when analytically marginalising over the Δz parameters. We can observe that the analytical and numerical marginalisation return nearly identical posteriors.

Our results for DES-Y1 data are shown in the left panel of Fig. 5.4, with the errors on all parameters listed in Table 5.3. On the one hand, we find that marginalising analytically or numerically over the Δz parameters leads to the same marginalised posterior for the cosmological parameters. On the other hand, fixing the Δz parameters returns a posterior distribution that is only mildly narrower than the marginal distribution. Thus, if we truly wish to study the effect of marginalising analytically as opposed to numerically over the Δz parameters we will have to consider futuristic LSST-like data, where the impact of these uncertainties will likely be higher.

We show results for futuristic LSST-like data on the right panel Fig. 5.4, with the parameter constraints listed in Table 5.3. First of all, in the case of LSST-like data we observe that not marginalising over the Δz parameters in the model results in significantly narrower posteriors, with the final uncertainties shrinking by a factor ~ 2. The impact of redshift distribution uncertainties in this case is thus much more relevant, and the accuracy of the analytical marginalisation scheme becomes

Δz model		Fixed	Numerical	Analytical
$\Omega_{\rm m}$	DES-Y1	0.333 ± 0.055	0.3 ± 0.056	0.306 ± 0.055
	LSST	0.311 ± 0.011	0.317 ± 0.02	0.317 ± 0.02
σ_8	DES-Y1	0.724 ± 0.072	0.765 ± 0.077	0.758 ± 0.076
	LSST	0.82 ± 0.015	0.821 ± 0.027	0.823 ± 0.027
S_8	DES-Y1	0.753 ± 0.015	0.756 ± 0.015	0.756 ± 0.015
	LSST	0.833 ± 0.002	0.833 ± 0.005	0.833 ± 0.006

Table 5.3: Numerical values for the mean and 1σ confidence intervals for the 1D marginalised posterior distributions of the cosmological parameters $\Omega_{\rm m}$, σ_8 and S_8 obtained when considering the first method (z shifts) to characterise the photometric redshift uncertainties. The first column shows the values obtained when the Δz parameters were kept fixed, the second column when they were marginalised numerically and the third column when they were marginalised analytically. In each row we display the constraints obtained when using DES-Y1 or LSST-like data to constrain the models.

paramount. However, comparing the contours obtained by numerical and analytical marginalisation, we observe that both methods return largely equivalent posterior distributions, with the final uncertainties changing by much less than 10%. This holds even in the case the Δz prior worsen by a factor 4 as seen in Figure 5.8, in Sect. 5.4.2.4. Therefore, linearising the likelihood around the Δz parameters will be a good enough approximation for LSST-data, at least for relatively simple parametrisations of the $p(z)$ uncertainty, which will allow us to reduce the dimensionality of the model and make parameter inference more efficient.

It is worth emphasising that the results in this section are not meant to be interpreted as forecasts on the constraining power of LSST on cosmological parameters, but only on our ability to analytically marginalise over photometric uncertainties in inferring the underlying cosmology. The recovered constraints depend strongly on assumptions such as the redshift calibration that LSST will be able to achieve for the different samples involved. As such, the results presented here are only a conservative estimate of the effect of analytic marginalisation on cosmological constraints.

5.4.2.2 Linearising $p_\alpha(z)$

In the previous section we have shown that, even for futuristic LSST-like data, it is possible to marginalise over redshift uncertainties analytically, assuming a relatively simple parametrisation of these uncertainties. We now turn to more complex models to characterise these uncertainties.

In order to do so, we consider the previously discussed $p_\alpha(z)$ model (called Method 2 above), which turns the height of each bin in the redshift distribution histograms into a free parameter. This results in 40 new free parameters per redshift bin with a total of 160 parameters for the data considered in this work.

We start by revisiting the DES-Y1 data analysis, presenting our results in the left panel of Fig. 5.5. As we observed in the previous section, we find that even for the far more general $p_\alpha(z)$ model there is no significant difference between numerically marginalising over the $p_\alpha(z)$, or doing so through our approximate analytical approach. Furthermore, as before, fixing the shape of the redshift distribution leads to only mildly tighter constraints. On the one hand, this means that the result found for the Δz model is not reliant on the simplicity of the model, but instead inherent to the sensitivity of DES-Y1 data. On the other hand, this also means that we must turn once again to futuristic LSST-like data to study the impact of a more general parametrisation of photometric uncertainties.

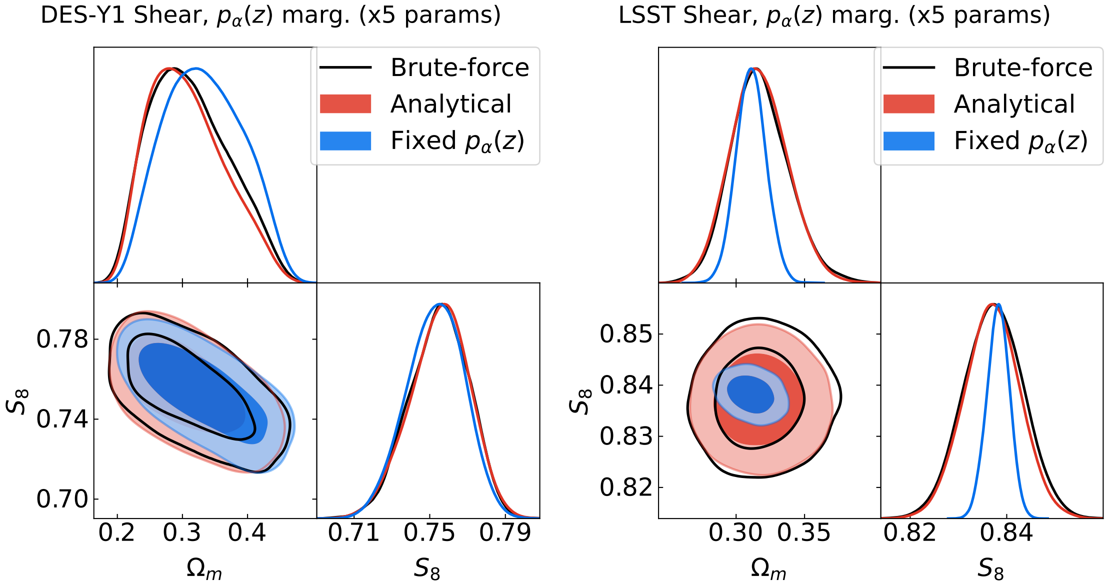

Figure 5.5: Marginalised posterior distributions for the combination of parameters Ω_m, σ_8 and S_8 obtained when considering the $p_\alpha(z)$ model for photometric uncertainties for DES-Y1 data. The blue contours correspond to the case where the $p_\alpha(z)$ parameter are fixed. The magenta contours are obtained when numerically marginalising over the $p_\alpha(z)$ parameters. Finally, the black dashed contours are obtained when analytically marginalising over the $p_\alpha(z)$ parameters. We can observe that the analytical and numerical marginalisation return nearly identical posteriors.

The results for futuristic LSST-like data are shown in the right panel of Fig. 5.5. As in the case of the Δz parametrisation, we find that, in the case of LSST-like data,

$p_\alpha(z)$ model		Fixed	Numerical	Analytical
Ω_m	DES-Y1	0.333 ± 0.056	0.308 ± 0.055	0.312 ± 0.057
	LSST	0.311 ± 0.011	0.317 ± 0.02	0.317 ± 0.021
σ_8	DES-Y1	0.723 ± 0.073	0.755 ± 0.075	0.75 ± 0.077
	LSST	0.824 ± 0.015	0.816 ± 0.026	0.815 ± 0.027
S_8	DES-Y1	0.753 ± 0.015	0.755 ± 0.015	0.755 ± 0.015
	LSST	0.838 ± 0.002	0.837 ± 0.006	0.837 ± 0.006

Table 5.4: Numerical values for the mean and 1σ confidence intervals for the 1D marginalised posterior distributions of the cosmological parameters Ω_m, σ_8 and S_8 obtained when considering the second method ($p(z)$ bin heights) to characterise the photometric redshift uncertainties. The first column shows the values obtained when the $p_\alpha(z)$ parameters are kept fixed, the second column when they are marginalised numerically, and the third column when they are marginalised analytically. In each row we display the constraints obtained when using DES-Y1 or LSST-like data to constrain the models.

not including the $p_\alpha(z)$ parameters in the model results in significantly narrower posteriors. By looking at the corresponding numerical values in Tab. 5.4, we see that the S_8 constraints become twice as tight when the $p_\alpha(z)$ parameters are fixed. Most importantly, we find that marginalising over the $p_\alpha(z)$ parameters analytically or numerically yields almost indistinguishable posteriors. Thus, the results found in Sect. 5.4.2.1 for the simple Δz parametrisation, in fact hold for significantly more general models of the uncertainty in the galaxy redshift distributions.

Finally, in Fig. 5.6 we present the constraints obtained for the 160 $p_\alpha(z)$ parameters for both the DES-Y1 (top panel) and LSST-like data (bottom panel) in colour bands. We observe that the posterior distributions are largely dominated by the prior (shown in dashed black line with error bars) and, thus, the redshift distribution is not significantly self-calibrated by the data in either case.

Before moving to the next section, it is worth stressing that constraining such a large parameter space has only been possible thanks to the auto-differentiable nature of the code used to obtain theoretical predictions, allowing us to use gradient-based samplers, much more efficient that standard samplers. The development of such auto-differentiable codes will therefore become imperative in the near future given the increasing complexity of models used in cosmological analyses.

5.4.2.3 Δz vs $p_\alpha(z)$

In the previous sections, we have focused in the impact of how we marginalise over the different parametrisations of photometric redshift uncertainties. In this section

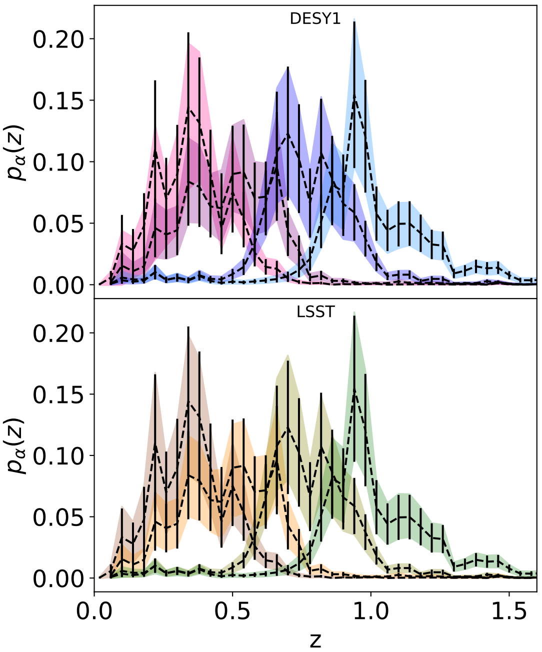

Figure 5.6: Posterior distributions for the $p_\alpha(z)$ parameters when considering DES-Y1 data (top row) and futuristic LSST-like data (bottom row). The black dashed line shows the mean of the Gaussian prior of the $p_\alpha(z)$ parameters. The error bars show their corresponding error.

we will focus instead on what we marginalise over, i.e. the impact of the choice of parametrisation. The question is then: Can a one-parameter-per-bin model (Δz model) capture all the meaningful modifications to photometric redshift distributions?

In order to answer this question, we constrain the cosmological parameters for the Δz and $p_\alpha(z)$ models in the case with futuristic LSST-like data. In both cases, we marginalise numerically over their respective nuisance parameters. As shown in Fig. 5.7 and Tables 5.3 and 5.4, both methods recover the same posterior distributions with small differences. Moreover, we observed no biases with respect to the fiducial cosmology regardless of the parametrisation of the photometric redshift uncertainty or the method used to marginalise over it.

Thus, it is in principle possible that even Stage-IV surveys will be able to use relatively simple models to describe the redshift distribution of cosmic shear samples[2]. Note however that this result is subject to the modelling of the rest of the systematic effects in the survey. More complex modelling of IA might require of better control on the $p(z)$ systematics. In such case, the $\Delta_{\rm z}$ parametrisation might become insufficient

---

[2] Note, however, this is likely not the case for photometric galaxy clustering studies where other properties of the redshift distribution (e.g. its width) have a stronger impact on the theoretical prediction [258].

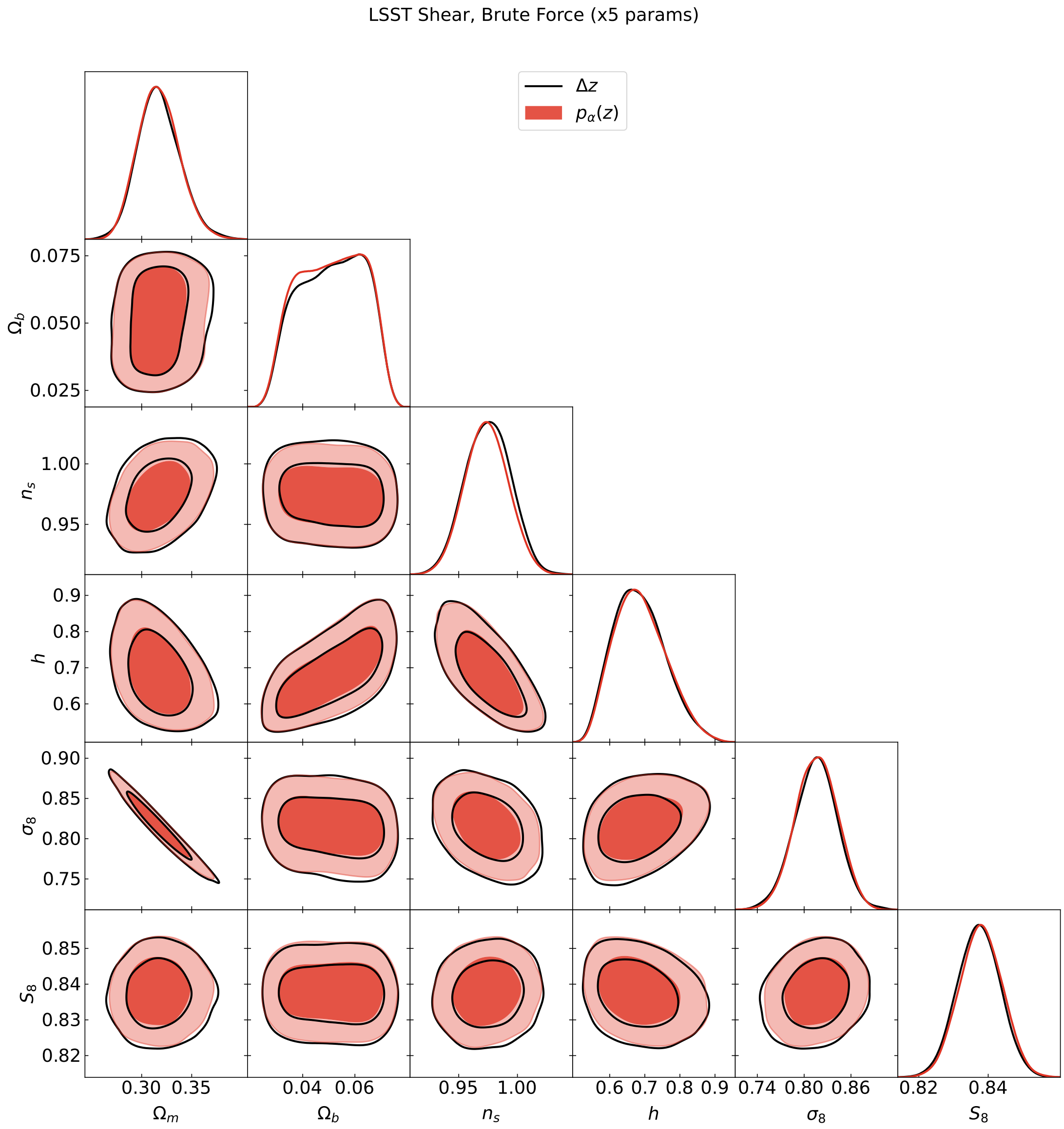

Figure 5.7: Comparison between the obtained marginalised posterior distributions of the cosmological parameters when numerically marginalising over the Δz (black dash-dotted) and $p_\alpha(z)$ (orange) photometric uncertainties models when applied to LSST-like futuristic data. Dotted lines mark the values of the fiducial cosmology used to generate the data. We can observe that both parametrisations of the photometric redshift uncertainties return identical posteriors for the cosmological parameters.

and one would expect a difference between said parametrisation and sampling the full $p_\alpha(z)$.

5.4.2.4 Stress-testing the approximation

As described in Sect. 5.2, the approximation used here to analytically marginalise over the redshift calibration parameters assumes a sufficiently tight prior on these parameters, such that the dependence of the theory prediction on them can be linearised. Testing whether this assumption might break in a realistic scenario, is therefore essential. This is important in the context of Stage-IV since, even though it is expected that spectroscopic samples and the associated calibration techniques will improve over time, the increase in depth that LSST-like surveys will represent may make the calibration of the faintest samples in the survey particularly challenging.

To further stress-test our approximate method, we repeat our analysis of the LSST-like futuristic data using the Δz model for redshift uncertainties with priors 4 times larger than used in our fiducial analysis (which themselves were based on existing calibration samples). The result of this test is shown in Fig. 5.8. Reassuringly, the results show that, despite quadrupling the uncertainty in the redshift nuisance parameters, the analytic marginalisation method yields virtually the same constraints on the cosmological parameters as the brute-force marginalisation, in spite of the significantly broader posterior contours. This implicit validates the approximation that a first-order expansion of the theory data vector with respect to a change in redshift distribution is sufficient over a conservative range of calibration priors.

5.5 Conclusions

Forecasts of the next decade in cosmology predict that meaningful constraints on fundamental unknowns such as the mass of neutrinos and the nature of dark matter and dark energy will come from multi-scale, multi-tracer efforts, encompassing a wide range of probes and redshifts. It is for this reason that the efficient analysis of joint data sets combining low- and high-redshift probes in an accurate manner is of great importance to the field of large-scale structure analysis. However, this comes at the cost of adding a colossal number of nuisance parameters characterising the observational and theoretical systematic uncertainties of all probes involved, which can noticeably slow down the sampling of the parameter space. In galaxy clustering and weak lensing joint studies, the most significant obstacles to overcome are the

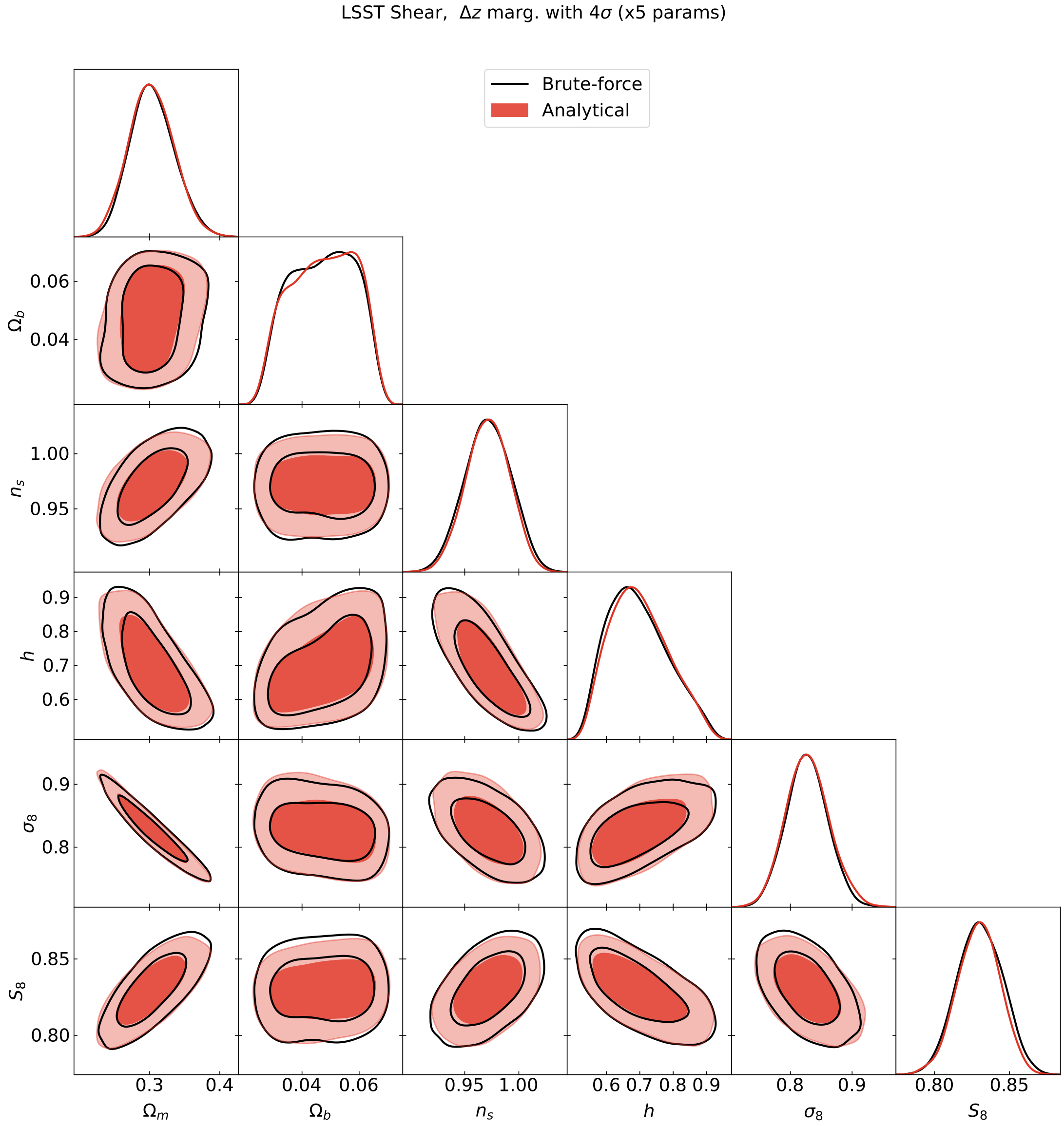

Figure 5.8: Comparison between the obtained marginalised posterior distributions of the cosmological parameters when analytically marginalising over the Δz (black dashed) and when performing the full numerical marginalisation (orange) when analysing LSST-like data. In both cases the Δz prior distributions where made 4 times wider. Dotted lines mark the values of the fiducial cosmology used to generate the data. We can observe that despite significantly broadening the prior distributions the analytical marginalisation returns virtually identical posteriors for the cosmological parameters.

accurate modelling of the redshift distribution, the galaxy bias relation, and intrinsic alignments.

In this chapter, we have introduced the Laplace approximation, an analytical marginalisation scheme for speeding up the sampling process in the presence of a large number of nuisance parameters, and investigated its accuracy when applied to photometric survey data. In particular, we studied the current DES-Y1 data set as well as a synthetic data vector from an LSST-like survey to validate whether the Laplace approximation is capable of reproducing the posterior contours and constraints one arrives at when adopting the traditional method of diligently varying tens of nuisance parameters. We focused in two particular cases of study. First, we aimed at reproducing current 3x2-pt analyses, marginalising over all sorts of nuisance parameters. Second, we focused on the impact of redshift photometric uncertainties in shear analyses and studied the equivalence between different parametrisations of said uncertainties.

Let us start by summarising our study of 3x2-pt analyses. Our results showed that the Laplace approximation produces indistinguishable posteriors for the cosmological parameters from those obtained performing full numerical marginalisation. Most importantly, the analytical marginalisation proved to be several fold faster than its numerical counterpart. This success was achieved by dividing the marginalisation process into two steps. First, we marginalised over the photometric redshift uncertainties by linearising their contribution to the likelihood function. In Sect. 1.2.3.3 it was shown that in such scenario, marginalising over these parameters amounts to precomputing a modification to the data covariance matrix. Having marginalised over the linearisable parameters, we turned to the bias an intrinsic alignments nuisance parameters. In order to marginalise these parameters, we exploited the fact that they contribute to the theory prediction (not the likelihood function) linearly. This allowed us to derive analytical expressions for the gradient of the likelihood function with respect to said parameters, significantly speeding up the optimisation process described in Sect. 1.2.3.1.

Concerning our study of photometric redshift uncertainties, our results show that, for present cosmic shear surveys, marginalising over these uncertainties has only a mild impact on the constraints on cosmological parameters. Nonetheless, our analytical approximation is able to capture their contribution accurately. This is true for the two parametrisations of the photometric uncertainties considered in this work, in terms of mean redshift shifts or redshift distribution histogram heights. However, the impact of redshift distribution uncertainties changes dramatically for future LSST-like surveys. In this case, redshift uncertainties commensurate with current calibration samples lead to a degradation in the final constraints on cosmological parameters up to a factor ~ 2. Capturing this effect for an arbitrarily complex parametrisation of the redshift distribution uncertainties is an a priori difficult task without resorting to a full exploration of the parameter space. Nevertheless, we find that the Laplace approximation is still able to recover the marginalised constraints on cosmological parameters to high fidelity, even after marginalising over more than 100 nuisance parameters. This means that, while future surveys will certainly have to account for these uncertainties, they will be able to do so using fast marginalisation methods without increasing the dimensionality of their astrophysical and cosmological models. Moreover we showed that simple parametrisations of the redshift distribution for cosmic shear samples, in terms of shifts in the mean redshift, are, surprisingly, able to reproduce the impact of the full uncertainty on $p(z)$ on the final constraints to high precision. Although this result will likely not hold for other probes (e.g. tomographic galaxy clustering), it should certainly simplify the analysis of future cosmic shear data.

Chapter 6

Conclusions

In this thesis, I have made a case for the need to develop auto-differentiable analysis pipelines for future cosmological surveys. I argue that the number of free parameters in future (and to some extent present) cosmological analyses will render traditional inference algorithms, such as Metropolis Hastings, computationally unfeasible. This will be due to two phenomena:

- **Systematic effects**: as the quality of data improves so will our treatment of its systematics. More complex models for these effects will bring along larger number of free parameters to capture the richer phenomenology (See Sect. 1.1.5).
- **Model-agnostic analyses**: in the face of the current inconsistencies in our theory of Cosmology, performing analysis that bypass these assumptions will become pressing. Removing these assumptions from our analyses will directly lead to a larger number of parameters to constrain (See Sect. 1.4).

In order to overcome this challenge, I have proposed and applied two possible solutions, one analytical and one numerical. Starting with the latter, gradient-based inference algorithms can characterise posterior distributions with thousands if not millions of dimensions (See Sect. 1.2.2). In particular, through this thesis I have made extensive use of the No-U-Turns sampler, a self-tuning version of traditional Hamiltonian Monte Carlo. However, they need to be provided with the gradient of the Bayesian likelihood function.

In Chapter 2, I developed an auto-differentiable framework to fit data from the expansion history, cosmological distances and the growth factor. This allowed me to develop a model-independent methodology to constrain Ω_{m} and $H(z)$. I obtained $\Omega_{\mathrm{m}} = 0.224 \pm 0.066$, a lower but statistically compatible value with the Planck 2018 constraints. Similarly, the Gaussian Process reconstruction of the expansion history proved to be compatible with the expectation from Planck despite fitting SH0ES supernova data. Instead, the tension manifested in the supernova absolute magnitude which was treated as a free parameter. I found $M = -19.43 \pm 0.026$ which is in 5σ tension with the SH0ES value.

In Chapter 3, I applied the same methodology to constrain possible deviations of General Relativity in the form of modifications to the Poisson equation. These can effectively be understood as a time-varying Newton's constant given by $\tilde{G} = G\mu(z)$ which enhances or depresses the growth of structure. I also studied the degeneracy of this type of modifications with modifications to the background Cosmology. Namely, we consider three possible background models. First, a traditional ΛCDM model. Second, wCDM parametrised in terms of the w_0 and w_a parameters. Third, a second Gaussian Process for $H(z)$, similarly to Sect. 2. My study found that $\mu(z)$ was largely independent on the modelling of the background expansion as opposed to what was formerly believed. This is because current measurements can effectively constrain the expansion history when ΛCDM or wCDM are considered. The story changes when a second Gaussian Process is used to model $H(z)$ since background data can no longer constrain Ω_{m}. In this scenario, Ω_{m} and $\mu(z)$ become completely degenerate. Thus we group them into the new variable $\tilde{\mu}(z) \equiv \Omega_{\mathrm{m}}\mu(z)$. Regardless of the background model considered, I found the constraints on $\mu(z)$ and $\tilde{\mu}(z)$ to be statistically compatible the General Relativity value with error bars ranging from 10% to 20%, depending on the specific model.

In Chapter 4, I developed and showcased `LimberJack.jl`, an open source library written `Julia` which extends the previously described auto-differentiable methods to angular power spectra analyses. Armed with this new tool, I performed a Gaussian Process reconstruction of the growth factor, $D(z)$, based on a combination of clustering, weak lensing and CMB lensing power spectra cross-correlations and the direct growth rate measurements considered in chapters 2 and 3. The results show that, despite being less than 2% of the data, growth rate measurements constitute 20% of the constraining power on the growth factor in the analyses. Moreover, the inclusion of growth data significantly smooths the reconstructed $D(z)$, disfavouring strong oscillations. Nonetheless, my results also tell a cautionary tale about deploying gradient-based inference algorithms at low dimensions. Thus, I found that, in these scenarios, the efficiency boost resulting from using the gradient of the likelihood can be outweighed by the cost of computing it, even when using auto-differentiable methods.

Finally, in Chapter 5, I showed how analytical marginalisation schemes can vastly reduce the number of free parameters in the analysis by getting rid of nuisance parameters. In particular, I studied the Laplace approximation which involves Taylor expanding the log-likelihood function around the best-fit value of the nuisance parameters. This expansion Gaussianises the posterior distribution along these parameters which can then be integrated analytically. Since the Laplace approximation entails optimising for the best-fit value of the nuisance parameters for a given cosmology, having access to the gradient of the likelihood function is also extremely beneficial. This reinforces the notion that auto-differentiable pipelines are an unavoidable step in order to deal with models with large numbers of parameters either numerically or analytically.

I first showed the reliability of the Laplace approximation by reproducing the DES-Y1 3x2-pt analysis as well as a futuristic 3x2-pt analysis. This proved that the Laplace approximation can be used to marginalise over all sorts of nuisances parameters when analysing both present and future data. Most importantly, using the Laplace approximation resulted in up to 15 times faster analyses. I then turned to explore different parametrisations of the photometric redshift uncertainties in cosmic shear analysis. I proposed a new model of this systematic effect in which each individual histogram bin in each of the tomographic bins of the galaxy distribution is treated as a free parameter with a Gaussian prior given by the covariance of the photometric calibration. I tested this model and compared it against the usual shift parameter model for the weak lensing data of the DES-Y1 survey. In this particular analysis the 4 tomographic bins of the weak lensing data contribute a total of 160 nuisance parameters. I used `LimberJack.jl` to numerically constrain the model and then compared the results with the analytical marginalisation performed using the Laplace approximation. My results showed that both the analytical and numerical approaches yielded indistinguishable posteriors for the cosmological parameters. Moreover, I showed that the current parametrisation of photometric uncertainty with one shift parameter per bin results in the same posteriors for the cosmological parameters as my more general model. Finally, the previous results were also found to hold when a futuristic cosmic shear analysis was considered instead of DES-Y1 data.

While I believe that the case made above is already compelling, there's nonetheless a plethora of avenues for improvement and continuation:

- **Precision measurements of the growth rate**: current measurements of the growth rate by SDSS collaboration [73, 145] and others are not precise enough to untangle the effects of varying Ω_{m} and σ_8. This will drastically change with the arrival of DESI data [198]. The forecasts I presented in Sect. 2 and Sect. 3 showed that DESI data will improve model-agnostic constraints of Ω_m and $\mu(z)$ several fold. Moreover, in Sect. 4, we showed that already current growth rate measurements are extremely powerful when included in angular power spectra analyses. Future DESI measurements will further improve this picture, making these RSD measurements part of the standard 3x2-pt cosmological analysis.

- **Better Auto-differentiable tools**: while `LimberJack.jl` is already a tool ready to perform current cosmological analyses, it can be improved in many ways. First, the implementation of auto-differentiation could be more efficient by making `LimberJack.jl` compatible with backwards auto-differentiation algorithms. Second, `LimberJack.jl` currently only runs on CPU. A significant speed boost would be obtained by making `LimberJack.jl` run on GPU's. Third, LB's current parallelisation routines have an underwhelming performance. On top of this, future cosmological analyses will require additional features such as better modelling of the non-linear matter power spectrum or beyond-Limber integration which currently not present in `LimberJack.jl` or similar libraries such as JAX-COSMO. This is due to most of these libraries being developed by independent teams with limited resources. In the future we expect large collaborations to start backing up these technologies leading to libraries of auto-differentiable methods akin to LSST's `CCL` among others.

- **Generic implementation of the Laplace approximation**: the implementations of the Laplace approximation shown in Sect. 5 were tailored to the particular analyses being undertaken. This is a major obstacle to a wider adoption of the method. Once again, this is mostly due to lack of resources to properly develop a generic, public implementation. In the future, we expect larger collaborations to adopt these methods, popularising them to a broader audience.

References

[1] J. Ruiz-Zapatero, C. García-García, D. Alonso, P. G. Ferreira, and R. D. P. Grumitt, mnras **512**, 1967 (2022), arXiv:2201.07025 [astro-ph.CO] .

[2] J. Ruiz-Zapatero, D. Alonso, P. G. Ferreira, and C. Garcia-Garcia, Phys. Rev. D **106**, 083523 (2022), arXiv:2207.09896 [astro-ph.CO] .

[3] J. Ruiz-Zapatero *et al.*, In prep, (2023).

[4] B. Hadzhiyska, K. Wolz, S. Azzoni, D. Alonso, C. García-García, J. Ruiz-Zapatero, and A. Slosar, The Open Journal of Astrophysics **6**, 23 (2023), arXiv:2301.11895 [astro-ph.CO] .

[5] J. Ruiz-Zapatero, B. Hadzhiyska, D. Alonso, P. G. Ferreira, C. García-García, and A. Mootoovaloo, MNRAS **522**, 5037 (2023), arXiv:2301.11978 [astro-ph.CO] .

[6] C. G. Böhmer, *Introduction to General Relativity and Cosmology.* (World Scientific, 2016).

[7] J. A. Peacock, *Cosmological Physics* (Cambridge University Press, 1998).

[8] S. Dodelson, *Modern Cosmology* (Academic Press, Elsevier Science, 2003).

[9] A. G. Lemaître, Monthly Notices of the Royal Astronomical Society **91**, 483 (1931), https://academic.oup.com/mnras/article-pdf/91/5/483/3079971/mnras91-0483.pdf .

[10] H. P. Robertson, ApJ **82**, 284 (1935).

[11] D. Scott, arXiv e-prints , arXiv:1804.01318 (2018), arXiv:1804.01318 [astro-ph.CO] .

[12] S. Aiola, E. Calabrese, L. Maurin, S. Naess, B. L. Schmitt, M. H. Abitbol, G. E. Addison, P. A. R. Ade, D. Alonso, M. Amiri, S. Amodeo, E. Angile, J. E. Austermann, T. Baildon, N. Battaglia, J. A. Beall, R. Bean, D. T. Becker, J. R. Bond, S. M. Bruno, V. Calafut, L. E. Campusano, F. Carrero, G. E. Chesmore, H.-m. Cho, S. K. Choi, S. E. Clark, N. F. Cothard, D. Crichton, K. T. Crowley, O. Darwish, R. Datta, E. V. Denison, M. J. Devlin, C. J. Duell, S. M. Duff, A. J. Duivenvoorden, J. Dunkley, R. Dünner, T. Essinger-Hileman,

M. Fankhanel, S. Ferraro, A. E. Fox, B. Fuzia, P. A. Gallardo, V. Gluscevic, J. E. Golec, E. Grace, M. Gralla, Y. Guan, K. Hall, M. Halpern, D. Han, P. Hargrave, M. Hasselfield, J. M. Helton, S. Henderson, B. Hensley, J. C. Hill, G. C. Hilton, M. Hilton, A. D. Hincks, R. Hložek, S.-P. P. Ho, J. Hubmayr, K. M. Huffenberger, J. P. Hughes, L. Infante, K. Irwin, R. Jackson, J. Klein, K. Knowles, B. Koopman, A. Kosowsky, V. Lakey, D. Li, Y. Li, Z. Li, M. Lokken, T. Louis, M. Lungu, A. MacInnis, M. Madhavacheril, F. Maldonado, M. Mallaby-Kay, D. Marsden, J. McMahon, F. Menanteau, K. Moodley, T. Morton, T. Namikawa, F. Nati, L. Newburgh, J. P. Nibarger, A. Nicola, M. D. Niemack, M. R. Nolta, J. Orlowski-Sherer, L. A. Page, C. G. Pappas, B. Partridge, P. Phakathi, G. Pisano, H. Prince, R. Puddu, F. J. Qu, J. Rivera, N. Robertson, F. Rojas, M. Salatino, E. Schaan, A. Schillaci, N. Sehgal, B. D. Sherwin, C. Sierra, J. Sievers, C. Sifon, P. Sikhosana, S. Simon, D. N. Spergel, S. T. Staggs, J. Stevens, E. Storer, D. D. Sunder, E. R. Switzer, B. Thorne, R. Thornton, H. Trac, J. Treu, C. Tucker, L. R. Vale, A. Van Engelen, J. Van Lanen, E. M. Vavagiakis, K. Wagoner, Y. Wang, J. T. Ward, E. J. Wollack, Z. Xu, F. Zago, and N. Zhu, JCAP **2020**, 047 (2020), arXiv:2007.07288 [astro-ph.CO] .

[13] Planck Collaboration, N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. J. Banday, R. B. Barreiro, N. Bartolo, S. Basak, R. Battye, K. Benabed, J. P. Bernard, M. Bersanelli, P. Bielewicz, J. J. Bock, J. R. Bond, J. Borrill, F. R. Bouchet, F. Boulanger, M. Bucher, C. Burigana, R. C. Butler, E. Calabrese, J. F. Cardoso, J. Carron, A. Challinor, H. C. Chiang, J. Chluba, L. P. L. Colombo, C. Combet, D. Contreras, B. P. Crill, F. Cuttaia, P. de Bernardis, G. de Zotti, J. Delabrouille, J. M. Delouis, E. Di Valentino, J. M. Diego, O. Doré, M. Douspis, A. Ducout, X. Dupac, S. Dusini, G. Efstathiou, F. Elsner, T. A. Enßlin, H. K. Eriksen, Y. Fantaye, M. Farhang, J. Fergusson, R. Fernandez-Cobos, F. Finelli, F. Forastieri, M. Frailis, A. A. Fraisse, E. Franceschi, A. Frolov, S. Galeotta, S. Galli, K. Ganga, R. T. Génova-Santos, M. Gerbino, T. Ghosh, J. González-Nuevo, K. M. Górski, S. Gratton, A. Gruppuso, J. E. Gudmundsson, J. Hamann, W. Handley, F. K. Hansen, D. Herranz, S. R. Hildebrandt, E. Hivon, Z. Huang, A. H. Jaffe, W. C. Jones, A. Karakci, E. Keihänen, R. Keskitalo, K. Kiiveri, J. Kim, T. S. Kisner, L. Knox, N. Krachmalnicoff, M. Kunz, H. Kurki-Suonio, G. Lagache, J. M. Lamarre, A. Lasenby, M. Lattanzi, C. R. Lawrence, M. Le Jeune, P. Lemos,

J. Lesgourgues, F. Levrier, A. Lewis, M. Liguori, P. B. Lilje, M. Lilley, V. Lindholm, M. López-Caniego, P. M. Lubin, Y. Z. Ma, J. F. Macías-Pérez, G. Maggio, D. Maino, N. Mandolesi, A. Mangilli, A. Marcos-Caballero, M. Maris, P. G. Martin, M. Martinelli, E. Martínez-González, S. Matarrese, N. Mauri, J. D. McEwen, P. R. Meinhold, A. Melchiorri, A. Mennella, M. Migliaccio, M. Millea, S. Mitra, M. A. Miville-Deschênes, D. Molinari, L. Montier, G. Morgante, A. Moss, P. Natoli, H. U. Nørgaard-Nielsen, L. Pagano, D. Paoletti, B. Partridge, G. Patanchon, H. V. Peiris, F. Perrotta, V. Pettorino, F. Piacentini, L. Polastri, G. Polenta, J. L. Puget, J. P. Rachen, M. Reinecke, M. Remazeilles, A. Renzi, G. Rocha, C. Rosset, G. Roudier, J. A. Rubiño-Martín, B. Ruiz-Granados, L. Salvati, M. Sandri, M. Savelainen, D. Scott, E. P. S. Shellard, C. Sirignano, G. Sirri, L. D. Spencer, R. Sunyaev, A. S. Suur-Uski, J. A. Tauber, D. Tavagnacco, M. Tenti, L. Toffolatti, M. Tomasi, T. Trombetti, L. Valenziano, J. Valiviita, B. Van Tent, L. Vibert, P. Vielva, F. Villa, N. Vittorio, B. D. Wandelt, I. K. Wehus, M. White, S. D. M. White, A. Zacchei, and A. Zonca, A&A **641**, A6 (2020), arXiv:1807.06209 [astro-ph.CO] .

[14] M. Einasto, L. J. Liivamägi, E. Tempel, E. Saar, E. Tago, P. Einasto, I. Enkvist, J. Einasto, V. J. Martínez, P. Heinämäki, and P. Nurmi, ApJ **736**, 51 (2011), arXiv:1105.1632 [astro-ph.CO] .

[15] Planck Collaboration, P. A. R. Ade, N. Aghanim, M. Arnaud, M. Ashdown, F. Atrio-Barandela, J. Aumont, C. Baccigalupi, A. Balbi, A. J. Banday, R. B. Barreiro, J. G. Bartlett, E. Battaner, K. Benabed, A. Benoît, J. P. Bernard, M. Bersanelli, I. Bikmaev, H. Böhringer, A. Bonaldi, J. R. Bond, J. Borrill, F. R. Bouchet, H. Bourdin, M. L. Brown, S. D. Brown, R. Burenin, C. Burigana, P. Cabella, J. F. Cardoso, P. Carvalho, A. Catalano, L. Cayón, L. Y. Chiang, G. Chon, P. R. Christensen, E. Churazov, D. L. Clements, S. Colafrancesco, L. P. L. Colombo, A. Coulais, B. P. Crill, F. Cuttaia, A. Da Silva, H. Dahle, L. Danese, R. J. Davis, P. de Bernardis, G. de Gasperis, A. de Rosa, G. de Zotti, J. Delabrouille, J. Démoclès, F. X. Désert, C. Dickinson, J. M. Diego, K. Dolag, H. Dole, S. Donzelli, O. Doré, U. Dörl, M. Douspis, X. Dupac, T. A. Enßlin, H. K. Eriksen, F. Finelli, I. Flores-Cacho, O. Forni, M. Frailis, E. Franceschi, M. Frommert, S. Galeotta, K. Ganga, R. T. Génova-Santos, M. Giard, M. Gilfanov, J. González-Nuevo, K. M. Górski, A. Gregorio, A. Gruppuso, F. K. Hansen, D. Harrison, S. Henrot-Versillé, C. Hernández-Monteagudo, S. R. Hildebrandt, E. Hivon, M. Hobson, W. A. Holmes, A. Hornstrup, W. Hovest, K. M. Huffenberger, G. Hurier, T. R. Jaffe, T. Jagemann,

W. C. Jones, M. Juvela, E. Keihänen, I. Khamitov, R. Kneissl, J. Knoche, L. Knox, M. Kunz, H. Kurki-Suonio, G. Lagache, A. Lähteenmäki, J. M. Lamarre, A. Lasenby, C. R. Lawrence, M. Le Jeune, R. Leonardi, P. B. Lilje, M. Linden-Vørnle, M. López-Caniego, P. M. Lubin, J. F. Macías-Pérez, B. Maffei, D. Maino, N. Mandolesi, M. Maris, F. Marleau, E. Martínez-González, S. Masi, M. Massardi, S. Matarrese, F. Matthai, P. Mazzotta, S. Mei, A. Melchiorri, J. B. Melin, L. Mendes, A. Mennella, S. Mitra, M. A. Miville-Deschênes, A. Moneti, L. Montier, G. Morgante, D. Munshi, J. A. Murphy, P. Naselsky, P. Natoli, H. U. Nørgaard-Nielsen, F. Noviello, D. Novikov, I. Novikov, S. Osborne, F. Pajot, D. Paoletti, O. Perdereau, F. Perrotta, F. Piacentini, M. Piat, E. Pierpaoli, R. Piffaretti, S. Plaszczynski, E. Pointecouteau, G. Polenta, N. Ponthieu, L. Popa, T. Poutanen, G. W. Pratt, S. Prunet, J. L. Puget, J. P. Rachen, R. Rebolo, M. Reinecke, M. Remazeilles, C. Renault, S. Ricciardi, T. Riller, I. Ristorcelli, G. Rocha, M. Roman, C. Rosset, M. Rossetti, J. A. Rubiño-Martín, L. Rudnick, B. Rusholme, M. Sandri, G. Savini, B. M. Schaefer, D. Scott, G. F. Smoot, F. Stivoli, R. Sudiwala, R. Sunyaev, D. Sutton, A. S. Suur-Uski, J. F. Sygnet, J. A. Tauber, L. Terenzi, L. Toffolatti, M. Tomasi, M. Tristram, J. Tuovinen, M. Türler, G. Umana, L. Valenziano, B. Van Tent, J. Varis, P. Vielva, F. Villa, N. Vittorio, L. A. Wade, B. D. Wandelt, N. Welikala, S. D. M. White, D. Yvon, A. Zacchei, S. Zaroubi, and A. Zonca, A&A **554**, A140 (2013), arXiv:1208.3611 [astro-ph.CO] .

[16] J. R. Bond, L. Kofman, and D. Pogosyan, Nature Astrophysics **380**, 603 (1996), arXiv:astro-ph/9512141 [astro-ph] .

[17] Y. B. Zel'dovich, Astronomy and Astrophysics **500**, 13 (1970).

[18] Planck Collaboration, P. A. R. Ade, N. Aghanim, M. Arnaud, F. Arroja, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. J. Banday, R. B. Barreiro, N. Bartolo, E. Battaner, K. Benabed, A. Benoît, A. Benoit-Lévy, J. P. Bernard, M. Bersanelli, P. Bielewicz, J. J. Bock, A. Bonaldi, L. Bonavera, J. R. Bond, J. Borrill, F. R. Bouchet, F. Boulanger, M. Bucher, C. Burigana, R. C. Butler, E. Calabrese, J. F. Cardoso, A. Catalano, A. Challinor, A. Chamballu, R. R. Chary, H. C. Chiang, P. R. Christensen, S. Church, D. L. Clements, S. Colombi, L. P. L. Colombo, C. Combet, D. Contreras, F. Couchot, A. Coulais, B. P. Crill, A. Curto, F. Cuttaia, L. Danese, R. D. Davies, R. J. Davis, P. de Bernardis, A. de Rosa, G. de Zotti, J. Delabrouille, F. X. Désert, J. M. Diego, H. Dole, S. Donzelli, O. Doré, M. Douspis, A. Ducout, X. Dupac, G. Efstathiou, F. Elsner, T. A. Enßlin, H. K. Eriksen, J. Fergusson,

F. Finelli, O. Forni, M. Frailis, A. A. Fraisse, E. Franceschi, A. Frejsel, A. Frolov, S. Galeotta, S. Galli, K. Ganga, C. Gauthier, M. Giard, Y. Giraud-Héraud, E. Gjerløw, J. González-Nuevo, K. M. Górski, S. Gratton, A. Gregorio, A. Gruppuso, J. E. Gudmundsson, J. Hamann, W. Handley, F. K. Hansen, D. Hanson, D. L. Harrison, S. Henrot-Versillé, C. Hernández-Monteagudo, D. Herranz, S. R. Hildebrandt, E. Hivon, M. Hobson, W. A. Holmes, A. Hornstrup, W. Hovest, Z. Huang, K. M. Huffenberger, G. Hurier, A. H. Jaffe, T. R. Jaffe, W. C. Jones, M. Juvela, E. Keihänen, R. Keskitalo, J. Kim, T. S. Kisner, R. Kneissl, J. Knoche, M. Kunz, H. Kurki-Suonio, G. Lagache, A. Lähteenmäki, J. M. Lamarre, A. Lasenby, M. Lattanzi, C. R. Lawrence, R. Leonardi, J. Lesgourgues, F. Levrier, A. Lewis, M. Liguori, P. B. Lilje, M. Linden-Vørnle, M. López-Caniego, P. M. Lubin, Y. Z. Ma, J. F. Macías-Pérez, G. Maggio, D. Maino, N. Mandolesi, A. Mangilli, M. Maris, P. G. Martin, E. Martínez-González, S. Masi, S. Matarrese, P. McGehee, P. R. Meinhold, A. Melchiorri, L. Mendes, A. Mennella, M. Migliaccio, S. Mitra, M. A. Miville-Deschênes, D. Molinari, A. Moneti, L. Montier, G. Morgante, D. Mortlock, A. Moss, M. Münchmeyer, D. Munshi, J. A. Murphy, P. Naselsky, F. Nati, P. Natoli, C. B. Netterfield, H. U. Nørgaard-Nielsen, F. Noviello, D. Novikov, I. Novikov, C. A. Oxborrow, F. Paci, L. Pagano, F. Pajot, R. Paladini, S. Pandolfi, D. Paoletti, F. Pasian, G. Patanchon, T. J. Pearson, H. V. Peiris, O. Perdereau, L. Perotto, F. Perrotta, V. Pettorino, F. Piacentini, M. Piat, E. Pierpaoli, D. Pietrobon, S. Plaszczynski, E. Pointecouteau, G. Polenta, L. Popa, G. W. Pratt, G. Prézeau, S. Prunet, J. L. Puget, J. P. Rachen, W. T. Reach, R. Rebolo, M. Reinecke, M. Remazeilles, C. Renault, A. Renzi, I. Ristorcelli, G. Rocha, C. Rosset, M. Rossetti, G. Roudier, M. Rowan-Robinson, J. A. Rubiño-Martín, B. Rusholme, M. Sandri, D. Santos, M. Savelainen, G. Savini, D. Scott, M. D. Seiffert, E. P. S. Shellard, M. Shiraishi, L. D. Spencer, V. Stolyarov, R. Stompor, R. Sudiwala, R. Sunyaev, D. Sutton, A. S. Suur-Uski, J. F. Sygnet, J. A. Tauber, L. Terenzi, L. Toffolatti, M. Tomasi, M. Tristram, T. Trombetti, M. Tucci, J. Tuovinen, L. Valenziano, J. Valiviita, B. Van Tent, P. Vielva, F. Villa, L. A. Wade, B. D. Wandelt, I. K. Wehus, M. White, D. Yvon, A. Zacchei, J. P. Zibin, and A. Zonca, A&A **594**, A20 (2016), arXiv:1502.02114 [astro-ph.CO] .

[19] P. Campeti and E. Komatsu, ApJ **941**, 110 (2022), arXiv:2205.05617 [astro-ph.CO] .

[20] D. Alonso, J. Sanchez, A. Slosar, and LSST Dark Energy Science Collaboration, MNRAS **484**, 4127 (2019), arXiv:1809.09603 [astro-ph.CO] .

[21] U. Seljak and M. Zaldarriaga, ApJ **469**, 437 (1996), arXiv:astro-ph/9603033 [astro-ph] .

[22] J. Lesgourgues, arXiv e-prints , arXiv:1104.2932 (2011), arXiv:1104.2932 [astro-ph.IM] .

[23] A. Lewis, A. Challinor, and A. Lasenby, ApJ **538**, 473 (2000), arXiv:astro-ph/9911177 [astro-ph] .

[24] J. M. Bardeen, J. R. Bond, N. Kaiser, and A. S. Szalay, ApJ **304**, 15 (1986).

[25] D. J. Eisenstein and W. Hu, ApJ **511**, 5 (1999), arXiv:astro-ph/9710252 [astro-ph] .

[26] D. J. Eisenstein and W. Hu, ApJ **496**, 605 (1998), arXiv:astro-ph/9709112 [astro-ph] .

[27] R. E. Smith, J. A. Peacock, A. Jenkins, S. D. M. White, C. S. Frenk, F. R. Pearce, P. A. Thomas, G. Efstathiou, and H. M. P. Couchman, MNRAS **341**, 1311 (2003), arXiv:astro-ph/0207664 [astro-ph] .

[28] R. Takahashi, M. Sato, T. Nishimichi, A. Taruya, and M. Oguri, ApJ **761**, 152 (2012), arXiv:1208.2701 [astro-ph.CO] .

[29] A. Heavens, MNRAS **343**, 1327 (2003), arXiv:astro-ph/0304151 [astro-ph] .

[30] M. Bartelmann and P. Schneider, Phys. Rep. **340**, 291 (2001), arXiv:astro-ph/9912508 [astro-ph] .

[31] D. N. Limber, ApJ **117**, 134 (1953).

[32] M. LoVerde and N. Afshordi, Phys. Rev. D **78**, 123506 (2008), arXiv:0809.5112 [astro-ph] .

[33] H. J. Mo and S. D. M. White, MNRAS **282**, 347 (1996), arXiv:astro-ph/9512127 [astro-ph] .

[34] M. Kilbinger, Reports on Progress in Physics **78**, 086901 (2015), arXiv:1411.0115 [astro-ph.CO] .

[35] A. Fernández-Soto, K. M. Lanzetta, H.-W. Chen, S. M. Pascarelle, and N. Yahata, ApJS **135**, 41 (2001), arXiv:astro-ph/0007447 [astro-ph] .

[36] J. Zuntz, F. Lanusse, A. I. Malz, A. H. Wright, A. Slosar, B. Abolfathi, D. Alonso, A. Bault, C. R. Bom, M. Brescia, A. Broussard, J.-E. Campagne, S. Cavuoti, E. S. Cypriano, B. M. O. Fraga, E. Gawiser, E. J. Gonzalez, D. Green, P. Hatfield, K. Iyer, D. Kirkby, A. Nicola, E. Nourbakhsh, A. Park, G. Teixeira, K. Heitmann, E. Kovacs, Y.-Y. Mao, and LSST Dark Energy Science Collaboration, The Open Journal of Astrophysics **4**, 13 (2021), arXiv:2108.13418 [astro-ph.IM] .

[37] M. Lima, C. E. Cunha, H. Oyaizu, J. Frieman, H. Lin, and E. S. Sheldon, MNRAS **390**, 118 (2008), arXiv:0801.3822 [astro-ph] .

[38] A. H. Wright, H. Hildebrandt, J. L. van den Busch, and C. Heymans, A&A **637**, A100 (2020), arXiv:1909.09632 [astro-ph.CO] .

[39] M. Schneider, L. Knox, H. Zhan, and A. Connolly, ApJ **651**, 14 (2006), arXiv:astro-ph/0606098 [astro-ph] .

[40] J. A. Newman, ApJ **684**, 88 (2008), arXiv:0805.1409 [astro-ph] .

[41] D. J. Matthews and J. A. Newman, ApJ **721**, 456 (2010), arXiv:1003.0687 [astro-ph.CO] .

[42] S. J. Schmidt, B. Ménard, R. Scranton, C. Morrison, and C. K. McBride, MNRAS **431**, 3307 (2013), arXiv:1303.0292 [astro-ph.CO] .

[43] J. Prat, C. Sánchez, Y. Fang, D. Gruen, J. Elvin-Poole, N. Kokron, L. F. Secco, B. Jain, R. Miquel, N. MacCrann, M. A. Troxel, A. Alarcon, D. Bacon, G. M. Bernstein, J. Blazek, R. Cawthon, C. Chang, M. Crocce, C. Davis, J. De Vicente, J. P. Dietrich, A. Drlica-Wagner, O. Friedrich, M. Gatti, W. G. Hartley, B. Hoyle, E. M. Huff, M. Jarvis, M. M. Rau, R. P. Rollins, A. J. Ross, E. Rozo, E. S. Rykoff, S. Samuroff, E. Sheldon, T. N. Varga, P. Vielzeuf, J. Zuntz, T. M. C. Abbott, F. B. Abdalla, S. Allam, J. Annis, K. Bechtol, A. Benoit-Lévy, E. Bertin, D. Brooks, E. Buckley-Geer, D. L. Burke, A. Carnero Rosell, M. Carrasco Kind, J. Carretero, F. J. Castander, C. E. Cunha, C. B. D'Andrea, L. N. da Costa, S. Desai, H. T. Diehl, S. Dodelson, T. F. Eifler, E. Fernandez, B. Flaugher, P. Fosalba, J. Frieman, J. García-Bellido, E. Gaztanaga, D. W. Gerdes, T. Giannantonio, D. A. Goldstein, R. A. Gruendl, J. Gschwend, G. Gutierrez, K. Honscheid, D. J. James, T. Jeltema, M. W. G. Johnson, M. D. Johnson, D. Kirk, E. Krause, K. Kuehn, S. Kuhlmann, O. Lahav, T. S. Li, M. Lima, M. A. G. Maia, M. March, J. L. Marshall, P. Martini, P. Melchior, F. Menanteau, J. J. Mohr, R. C. Nichol, B. Nord, A. A. Plazas, A. K. Romer, A. Roodman, M. Sako, E. Sanchez, V. Scarpine, R. Schindler, M. Schubnell, I. Sevilla-Noarbe, M. Smith, R. C. Smith, M. Soares-Santos, F. Sobreira, E. Suchyta, M. E. C. Swanson, G. Tarle, D. Thomas, D. L. Tucker, V. Vikram, A. R. Walker, R. H. Wechsler, B. Yanny, Y. Zhang, and DES Collaboration, Phys. Rev. D **98**, 042005 (2018), arXiv:1708.01537 [astro-ph.CO] .

[44] C. Sánchez, J. Prat, G. Zacharegkas, S. Pandey, E. Baxter, G. M. Bernstein, J. Blazek, R. Cawthon, C. Chang, E. Krause, P. Lemos, Y. Park, M. Raveri, J. Sanchez, M. A. Troxel, A. Amon, X. Fang, O. Friedrich, D. Gruen, A. Porredon, L. F. Secco, S. Samuroff, A. Alarcon, O. Alves, F. Andrade-Oliveira,

K. Bechtol, M. R. Becker, H. Camacho, A. Campos, A. Carnero Rosell, M. Carrasco Kind, R. Chen, A. Choi, M. Crocce, C. Davis, J. De Vicente, J. DeRose, E. Di Valentino, H. T. Diehl, S. Dodelson, C. Doux, A. Drlica-Wagner, K. Eckert, T. F. Eifler, F. Elsner, J. Elvin-Poole, S. Everett, A. Ferté, P. Fosalba, M. Gatti, G. Giannini, R. A. Gruendl, I. Harrison, W. G. Hartley, K. Herner, E. M. Huff, D. Huterer, M. Jarvis, B. Jain, N. Kuropatkin, P. F. Leget, N. MacCrann, J. McCullough, J. Muir, J. Myles, A. Navarro-Alsina, R. P. Rollins, A. Roodman, R. Rosenfeld, E. S. Rykoff, I. Sevilla-Noarbe, E. Sheldon, T. Shin, A. Troja, I. Tutusaus, T. N. Varga, R. H. Wechsler, B. Yanny, B. Yin, Y. Zhang, J. Zuntz, T. M. C. Abbott, M. Aguena, S. Allam, D. Bacon, E. Bertin, S. Bhargava, D. Brooks, E. Buckley-Geer, D. L. Burke, J. Carretero, M. Costanzi, L. N. da Costa, M. E. S. Pereira, S. Desai, J. P. Dietrich, P. Doel, A. E. Evrard, I. Ferrero, B. Flaugher, J. Frieman, J. García-Bellido, E. Gaztanaga, D. W. Gerdes, T. Giannantonio, J. Gschwend, G. Gutierrez, S. R. Hinton, D. L. Hollowood, K. Honscheid, B. Hoyle, D. J. James, K. Kuehn, O. Lahav, M. Lima, H. Lin, M. A. G. Maia, J. L. Marshall, P. Martini, P. Melchior, F. Menanteau, R. Miquel, J. J. Mohr, R. Morgan, A. Palmese, F. Paz-Chinchón, D. Petravick, A. Pieres, A. A. Plazas Malagón, M. Rodriguez-Monroy, E. Sanchez, V. Scarpine, M. Schubnell, S. Serrano, M. Smith, M. Soares-Santos, E. Suchyta, M. E. C. Swanson, G. Tarle, D. Thomas, C. To, and DES Collaboration, Phys. Rev. D **105**, 083529 (2022), arXiv:2105.13542 [astro-ph.CO] .

[45] Z. J. Zhang, C. Chang, P. Larsen, L. F. Secco, J. Zuntz, and LSST Dark Energy Science Collaboration, MNRAS **514**, 2181 (2022), arXiv:2111.04917 [astro-ph.CO] .

[46] J. A. Newman and D. Gruen, ARA&A **60**, 363 (2022), arXiv:2206.13633 [astro-ph.CO] .

[47] G. Cabass, M. M. Ivanov, M. Lewandowski, M. Mirbabayi, and M. Simonović, Science Direct **40**, 101193 (2023), arXiv:2203.08232 [astro-ph.CO] .

[48] P. McDonald and A. Roy, JCAP **2009**, 020 (2009), arXiv:0902.0991 [astro-ph.CO] .

[49] M. M. Abidi and T. Baldauf, JCAP **2018**, 029 (2018), arXiv:1802.07622 [astro-ph.CO] .

[50] M. Zennaro, R. E. Angulo, M. Pellejero-Ibáñez, J. Stücker, S. Contreras, and G. Aricò, arXiv e-prints , arXiv:2101.12187 (2021), arXiv:2101.12187 [astro-ph.CO] .

[51] T. M. C. Abbott, F. B. Abdalla, A. Alarcon, J. Aleksić, S. Allam, S. Allen, A. Amara, J. Annis, J. Asorey, S. Avila, D. Bacon, E. Balbinot, M. Banerji, N. Banik, W. Barkhouse, M. Baumer, E. Baxter, K. Bechtol, M. R. Becker, A. Benoit-Lévy, B. A. Benson, G. M. Bernstein, E. Bertin, J. Blazek, S. L. Bridle, D. Brooks, D. Brout, E. Buckley-Geer, D. L. Burke, M. T. Busha, A. Campos, D. Capozzi, A. Carnero Rosell, M. Carrasco Kind, J. Carretero, F. J. Castander, R. Cawthon, C. Chang, N. Chen, M. Childress, A. Choi, C. Conselice, R. Crittenden, M. Crocce, C. E. Cunha, C. B. D'Andrea, L. N. da Costa, R. Das, T. M. Davis, C. Davis, J. De Vicente, D. L. DePoy, J. DeRose, S. Desai, H. T. Diehl, J. P. Dietrich, S. Dodelson, P. Doel, A. Drlica-Wagner, T. F. Eifler, A. E. Elliott, F. Elsner, J. Elvin-Poole, J. Estrada, A. E. Evrard, Y. Fang, E. Fernandez, A. Ferté, D. A. Finley, B. Flaugher, P. Fosalba, O. Friedrich, J. Frieman, J. García-Bellido, M. Garcia-Fernandez, M. Gatti, E. Gaztanaga, D. W. Gerdes, T. Giannantonio, M. S. S. Gill, K. Glazebrook, D. A. Goldstein, D. Gruen, R. A. Gruendl, J. Gschwend, G. Gutierrez, S. Hamilton, W. G. Hartley, S. R. Hinton, K. Honscheid, B. Hoyle, D. Huterer, B. Jain, D. J. James, M. Jarvis, T. Jeltema, M. D. Johnson, M. W. G. Johnson, T. Kacprzak, S. Kent, A. G. Kim, A. King, D. Kirk, N. Kokron, A. Kovacs, E. Krause, C. Krawiec, A. Kremin, K. Kuehn, S. Kuhlmann, N. Kuropatkin, F. Lacasa, O. Lahav, T. S. Li, A. R. Liddle, C. Lidman, M. Lima, H. Lin, N. MacCrann, M. A. G. Maia, M. Makler, M. Manera, M. March, J. L. Marshall, P. Martini, R. G. McMahon, P. Melchior, F. Menanteau, R. Miquel, V. Miranda, D. Mudd, J. Muir, A. Möller, E. Neilsen, R. C. Nichol, B. Nord, P. Nugent, R. L. C. Ogando, A. Palmese, J. Peacock, H. V. Peiris, J. Peoples, W. J. Percival, D. Petravick, A. A. Plazas, A. Porredon, J. Prat, A. Pujol, M. M. Rau, A. Refregier, P. M. Ricker, N. Roe, R. P. Rollins, A. K. Romer, A. Roodman, R. Rosenfeld, A. J. Ross, E. Rozo, E. S. Rykoff, M. Sako, A. I. Salvador, S. Samuroff, C. Sánchez, E. Sanchez, B. Santiago, V. Scarpine, R. Schindler, D. Scolnic, L. F. Secco, S. Serrano, I. Sevilla-Noarbe, E. Sheldon, R. C. Smith, M. Smith, J. Smith, M. Soares-Santos, F. Sobreira, E. Suchyta, G. Tarle, D. Thomas, M. A. Troxel, D. L. Tucker, B. E. Tucker, S. A. Uddin, T. N. Varga, P. Vielzeuf, V. Vikram, A. K. Vivas, A. R. Walker, M. Wang, R. H. Wechsler, J. Weller, W. Wester, R. C. Wolf, B. Yanny, F. Yuan, A. Zenteno, B. Zhang, Y. Zhang, J. Zuntz, and Dark Energy Survey Collaboration, Phys. Rev. D **98**, 043526 (2018), arXiv:1708.01530 [astro-ph.CO] .

[52] C. M. Hirata and U. Seljak, Phys. Rev. D **70**, 063526 (2004), arXiv:astro-ph/0406275 [astro-ph] .

[53] M. A. Troxel, N. MacCrann, J. Zuntz, T. F. Eifler, E. Krause, S. Dodelson, D. Gruen, J. Blazek, O. Friedrich, S. Samuroff, J. Prat, L. F. Secco, C. Davis, A. Ferté, J. DeRose, A. Alarcon, A. Amara, E. Baxter, M. R. Becker, G. M. Bernstein, S. L. Bridle, R. Cawthon, C. Chang, A. Choi, J. De Vicente, A. Drlica-Wagner, J. Elvin-Poole, J. Frieman, M. Gatti, W. G. Hartley, K. Honscheid, B. Hoyle, E. M. Huff, D. Huterer, B. Jain, M. Jarvis, T. Kacprzak, D. Kirk, N. Kokron, C. Krawiec, O. Lahav, A. R. Liddle, J. Peacock, M. M. Rau, A. Refregier, R. P. Rollins, E. Rozo, E. S. Rykoff, C. Sánchez, I. Sevilla-Noarbe, E. Sheldon, A. Stebbins, T. N. Varga, P. Vielzeuf, M. Wang, R. H. Wechsler, B. Yanny, T. M. C. Abbott, F. B. Abdalla, S. Allam, J. Annis, K. Bechtol, A. Benoit-Lévy, E. Bertin, D. Brooks, E. Buckley-Geer, D. L. Burke, A. Carnero Rosell, M. Carrasco Kind, J. Carretero, F. J. Castander, M. Crocce, C. E. Cunha, C. B. D'Andrea, L. N. da Costa, D. L. DePoy, S. Desai, H. T. Diehl, J. P. Dietrich, P. Doel, E. Fernandez, B. Flaugher, P. Fosalba, J. García-Bellido, E. Gaztanaga, D. W. Gerdes, T. Giannantonio, D. A. Goldstein, R. A. Gruendl, J. Gschwend, G. Gutierrez, D. J. James, T. Jeltema, M. W. G. Johnson, M. D. Johnson, S. Kent, K. Kuehn, S. Kuhlmann, N. Kuropatkin, T. S. Li, M. Lima, H. Lin, M. A. G. Maia, M. March, J. L. Marshall, P. Martini, P. Melchior, F. Menanteau, R. Miquel, J. J. Mohr, E. Neilsen, R. C. Nichol, B. Nord, D. Petravick, A. A. Plazas, A. K. Romer, A. Roodman, M. Sako, E. Sanchez, V. Scarpine, R. Schindler, M. Schubnell, M. Smith, R. C. Smith, M. Soares-Santos, F. Sobreira, E. Suchyta, M. E. C. Swanson, G. Tarle, D. Thomas, D. L. Tucker, V. Vikram, A. R. Walker, J. Weller, Y. Zhang, and DES Collaboration, Phys. Rev. D **98**, 043528 (2018), arXiv:1708.01538 [astro-ph.CO] .

[54] C. García-García, J. Ruiz Zapatero, D. Alonso, E. Bellini, P. G. Ferreira, E.-M. Mueller, A. Nicola, and P. Ruiz-Lapuente, arXiv e-prints , arXiv:2105.12108 (2021), arXiv:2105.12108 [astro-ph.CO] .

[55] B. Jego, J. Ruiz-Zapatero, C. García-García, N. Koukoufilippas, and D. Alonso, arXiv e-prints , arXiv:2206.15394 (2022), arXiv:2206.15394 [astro-ph.GA] .

[56] B. Jego, D. Alonso, C. García-García, and J. Ruiz-Zapatero, MNRAS **520**, 583 (2023), arXiv:2209.05472 [astro-ph.CO] .

[57] T. Ferreira, D. Alonso, C. Garcia-Garcia, and N. E. Chisari, arXiv e-prints , arXiv:2309.11129 (2023), arXiv:2309.11129 [astro-ph.CO] .

[58] C. García-García, D. Alonso, and E. Bellini, JCAP **2019**, 043 (2019), arXiv:1906.11765 [astro-ph.CO] .

[59] C. Garcia-Garcia, (In Prep.).

[60] C. Heymans, T. Tröster, M. Asgari, C. Blake, H. Hildebrandt, B. Joachimi, K. Kuijken, C.-A. Lin, A. G. Sánchez, J. L. van den Busch, A. H. Wright, A. Amon, M. Bilicki, J. de Jong, M. Crocce, A. Dvornik, T. Erben, M. C. Fortuna, F. Getman, B. Giblin, K. Glazebrook, H. Hoekstra, S. Joudaki, A. Kannawadi, F. Köhlinger, C. Lidman, L. Miller, N. R. Napolitano, D. Parkinson, P. Schneider, H. Shan, E. A. Valentijn, G. Verdoes Kleijn, and C. Wolf, A&A **646**, A140 (2021), arXiv:2007.15632 [astro-ph.CO] .

[61] T. M. C. Abbott, M. Aguena, A. Alarcon, S. Allam, O. Alves, A. Amon, F. Andrade-Oliveira, J. Annis, S. Avila, D. Bacon, E. Baxter, K. Bechtol, M. R. Becker, G. M. Bernstein, S. Bhargava, S. Birrer, J. Blazek, A. Brandao-Souza, S. L. Bridle, D. Brooks, E. Buckley-Geer, D. L. Burke, H. Camacho, A. Campos, A. Carnero Rosell, M. Carrasco Kind, J. Carretero, F. J. Castander, R. Cawthon, C. Chang, A. Chen, R. Chen, A. Choi, C. Conselice, J. Cordero, M. Costanzi, M. Crocce, L. N. da Costa, M. E. da Silva Pereira, C. Davis, T. M. Davis, J. De Vicente, J. DeRose, S. Desai, E. Di Valentino, H. T. Diehl, J. P. Dietrich, S. Dodelson, P. Doel, C. Doux, A. Drlica-Wagner, K. Eckert, T. F. Eifler, F. Elsner, J. Elvin-Poole, S. Everett, A. E. Evrard, X. Fang, A. Farahi, E. Fernandez, I. Ferrero, A. Ferté, P. Fosalba, O. Friedrich, J. Frieman, J. García-Bellido, M. Gatti, E. Gaztanaga, D. W. Gerdes, T. Giannantonio, G. Giannini, D. Gruen, R. A. Gruendl, J. Gschwend, G. Gutierrez, I. Harrison, W. G. Hartley, K. Herner, S. R. Hinton, D. L. Hollowood, K. Honscheid, B. Hoyle, E. M. Huff, D. Huterer, B. Jain, D. J. James, M. Jarvis, N. Jeffrey, T. Jeltema, A. Kovacs, E. Krause, R. Kron, K. Kuehn, N. Kuropatkin, O. Lahav, P. F. Leget, P. Lemos, A. R. Liddle, C. Lidman, M. Lima, H. Lin, N. MacCrann, M. A. G. Maia, J. L. Marshall, P. Martini, J. McCullough, P. Melchior, J. Mena-Fernández, F. Menanteau, R. Miquel, J. J. Mohr, R. Morgan, J. Muir, J. Myles, S. Nadathur, A. Navarro-Alsina, R. C. Nichol, R. L. C. Ogando, Y. Omori, A. Palmese, S. Pandey, Y. Park, F. Paz-Chinchón, D. Petravick, A. Pieres, A. A. Plazas Malagón, A. Porredon, J. Prat, M. Raveri, M. Rodriguez-Monroy, R. P. Rollins, A. K. Romer, A. Roodman, R. Rosenfeld, A. J. Ross, E. S. Rykoff, S. Samuroff, C. Sánchez, E. Sanchez, J. Sanchez, D. Sanchez Cid, V. Scarpine, M. Schubnell, D. Scolnic, L. F. Secco, S. Serrano, I. Sevilla-Noarbe, E. Sheldon, T. Shin, M. Smith, M. Soares-Santos, E. Suchyta,

M. E. C. Swanson, M. Tabbutt, G. Tarle, D. Thomas, C. To, A. Troja, M. A. Troxel, D. L. Tucker, I. Tutusaus, T. N. Varga, A. R. Walker, N. Weaverdyck, R. Wechsler, J. Weller, B. Yanny, B. Yin, Y. Zhang, J. Zuntz, and DES Collaboration, Phys. Rev. D **105**, 023520 (2022), arXiv:2105.13549 [astro-ph.CO] .

[62] M. White, R. Zhou, J. DeRose, S. Ferraro, S.-F. Chen, N. Kokron, S. Bailey, D. Brooks, J. García-Bellido, J. Guy, K. Honscheid, R. Kehoe, A. Kremin, M. Levi, N. Palanque-Delabrouille, C. Poppett, D. Schlegel, and G. Tarle, JCAP **2022**, 007 (2022), arXiv:2111.09898 [astro-ph.CO] .

[63] G. Cowan, *Statistical data analysis* (Oxford University Press, USA, 1998).

[64] C. P. Robert, J.-M. Marin, and J. Rousseau, arXiv e-prints , arXiv:1002.2080 (2010), arXiv:1002.2080 [stat.ME] .

[65] A. Gelman, Bayesian Analysis **3**, 445 (2008).

[66] M. Betancourt, arXiv e-prints , arXiv:1701.02434 (2017), arXiv:1701.02434 [stat.ME] .

[67] R. E. Caflisch, Acta Numerica **7**, 1–49 (1998).

[68] G. Grimmett and D. Stirzaker, *Probability and random processes*, Vol. 80 (Oxford university press, 2001).

[69] A. Gelman and D. B. Rubin, Statistical Science **7**, 457 (1992).

[70] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).

[71] C. Geyer, "Introduction to markov chain monte carlo," in *Handbook of Markov Chain Monte Carlo* (CRC Press, 2011) pp. 3–48.

[72] C. K. Carter and R. Kohn, Biometrika **81**, 541 (1994).

[73] S. Alam, M. Ata, S. Bailey, F. Beutler, D. Bizyaev, J. A. Blazek, A. S. Bolton, J. R. Brownstein, A. Burden, C.-H. Chuang, J. Comparat, A. J. Cuesta, K. S. Dawson, D. J. Eisenstein, S. Escoffier, H. Gil-Marín, J. N. Grieb, N. Hand, S. Ho, K. Kinemuchi, D. Kirkby, F. Kitaura, E. Malanushenko, V. Malanushenko, C. Maraston, C. K. McBride, R. C. Nichol, M. D. Olmstead, D. Oravetz, N. Padmanabhan, N. Palanque-Delabrouille, K. Pan, M. Pellejero-Ibanez, W. J. Percival, P. Petitjean, F. Prada, A. M. Price-Whelan, B. A. Reid, S. A. Rodríguez-Torres, N. A. Roe, A. J. Ross, N. P. Ross, G. Rossi, J. A. Rubiño-Martín, S. Saito, S. Salazar-Albornoz, L. Samushia, A. G. Sánchez, S. Satpathy, D. J. Schlegel, D. P. Schneider, C. G. Scóccola, H.-J. Seo, E. S. Sheldon, A. Simmons, A. Slosar, M. A. Strauss, M. E. C. Swanson, D. Thomas, J. L. Tinker, R. Tojeiro, M. V. Magaña, J. A. Vazquez, L. Verde, D. A. Wake,

Y. Wang, D. H. Weinberg, M. White, W. M. Wood-Vasey, C. Yèche, I. Zehavi, Z. Zhai, and G.-B. Zhao, MNRAS **470**, 2617 (2017), arXiv:1607.03155 [astro-ph.CO] .

[74] J. Goodman and J. Weare, Communications in Applied Mathematics and Computational Science **5**, 65 (2010).

[75] D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman, PASP **125**, 306 (2013), arXiv:1202.3665 [astro-ph.IM] .

[76] D. J. C. MacKay, *Information Theory, Inference & Learning Algorithms* (Cambridge University Press, Cambridge University Press, Shaftesbury Road Cambridge, CB2 8BS, United Kingdom, 2002).

[77] M. D. Hoffman and A. Gelman, arXiv e-prints , arXiv:1111.4246 (2011), arXiv:1111.4246 [stat.CO] .

[78] L. Zhang, B. Carpenter, A. Gelman, and A. Vehtari, arXiv e-prints , arXiv:2108.03782 (2021), arXiv:2108.03782 [stat.ML] .

[79] J. Robnik, G. B. De Luca, E. Silverstein, and U. Seljak, arXiv e-prints , arXiv:2212.08549 (2022), arXiv:2212.08549 [stat.CO] .

[80] J. Robnik and U. Seljak, arXiv e-prints , arXiv:2303.18221 (2023), arXiv:2303.18221 [hep-lat] .

[81] R. E. Kass, L. Tierney, and J. J. B. Kadane (1990).

[82] A. N. Taylor and T. D. Kitching, MNRAS **408**, 865 (2010), arXiv:1003.1136 [astro-ph.CO] .

[83] S. R. Cole, H. Chu, and S. Greenland, American Journal of Epidemiology **179**, 252 (2013), https://academic.oup.com/aje/article-pdf/179/2/252/287443/kwt245.pdf .

[84] R. D. Cousins, Am. J. Phys. **63**, 398 (1995).

[85] J. Hamann, S. Hannestad, G. G. Raffelt, and Y. Y. Y. Wong, JCAP **2007**, 021 (2007), arXiv:0705.0440 [astro-ph] .

[86] L. Herold, E. G. M. Ferreira, and E. Komatsu, ApJ **929**, L16 (2022), arXiv:2112.12140 [astro-ph.CO] .

[87] O. Barndorff-Nielsen, Biometrika **70**, 343 (1983).

[88] J. R. Bond, A. H. Jaffe, and L. Knox, Phys. Rev. D **57**, 2117 (1998), arXiv:astro-ph/9708203 [astro-ph] .

[89] D. Jimenez Rezende and S. Mohamed, arXiv e-prints , arXiv:1505.05770 (2015), arXiv:1505.05770 [stat.ML] .

[90] G. B. Rybicki and W. H. Press, ApJ **398**, 169 (1992).

[91] M. Woodbury, in *Memorandum Rept. 42, Statistical Research Group* (Princeton Univ., 1950) p. 4.

[92] G. W. Leibniz, "Monadology," (1714), 1.

[93] M. Spivak, *Calculus On Manifolds: A Modern Approach To Classical Theorems Of Advanced Calculus*, Mathematics monograph series (Avalon Publishing, 1971).

[94] A. Griewank and A. Walther, *Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation*, 2nd ed. (Society for Industrial and Applied Mathematics, USA, 2008).

[95] N. Porqueres, A. Heavens, D. Mortlock, and G. Lavaux, MNRAS **509**, 3194 (2022), arXiv:2108.04825 [astro-ph.CO] .

[96] R. E. Wengert, Communications of the ACM **7**, 463 (1964).

[97] M. Bartholomew-Biggs, S. Brown, B. Christianson, and L. Dixon, Journal of Computational and Applied Mathematics **124**, 171 (2000), numerical Analysis 2000. Vol. IV: Optimization and Nonlinear Equations.

[98] S. Linnainmaa, BIT **16**, 146–160 (1976).

[99] C. Bischof and H. Bucker, NIC Series **3** (2000).

[100] B. van Merriënboer, O. Breuleux, A. Bergeron, and P. Lamblin, arXiv e-prints , arXiv:1810.11530 (2018), arXiv:1810.11530 [cs.LG] .

[101] W. S. Moses, S. H. K. Narayanan, L. Paehler, V. Churavy, M. Schanen, J. Hückelheim, J. Doerfert, and P. Hovland, in *Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis*, SC '22 (IEEE Press, 2022).

[102] P. Shah, P. Lemos, and O. Lahav, A&A Rev. **29**, 9 (2021), arXiv:2109.01161 [astro-ph.CO] .

[103] G. Efstathiou, MNRAS **505**, 3866 (2021), arXiv:2103.08723 [astro-ph.CO] .

[104] A. G. Riess, S. Casertano, W. Yuan, J. B. Bowers, L. Macri, J. C. Zinn, and D. Scolnic, ApJ **908**, L6 (2021), arXiv:2012.08534 [astro-ph.CO] .

[105] D. Brout, D. Scolnic, B. Popovic, A. G. Riess, A. Carr, J. Zuntz, R. Kessler, T. M. Davis, S. Hinton, D. Jones, W. D. Kenworthy, E. R. Peterson, K. Said, G. Taylor, N. Ali, P. Armstrong, P. Charvu, A. Dwomoh, C. Meldorf, A. Palmese, H. Qu, B. M. Rose, B. Sanchez, C. W. Stubbs, M. Vincenzi, C. M. Wood, P. J. Brown, R. Chen, K. Chambers, D. A. Coulter, M. Dai, G. Dimitriadis, A. V. Filippenko, R. J. Foley, S. W. Jha, L. Kelsey, R. P. Kirshner, A. Möller, J. Muir, S. Nadathur, Y.-C. Pan, A. Rest, C. Rojas-Bravo, M. Sako,

M. R. Siebert, M. Smith, B. E. Stahl, and P. Wiseman, ApJ **938**, 110 (2022), arXiv:2202.04077 [astro-ph.CO] .

[106] N. MacCrann, J. Zuntz, S. Bridle, B. Jain, and M. R. Becker, MNRAS **451**, 2877 (2015), arXiv:1408.4742 [astro-ph.CO] .

[107] S. Joudaki *et al.*, Mon. Not. Roy. Astron. Soc. **474**, 4894 (2018), arXiv:1707.06627 [astro-ph.CO] .

[108] E. Mörtsell and S. Dhawan, 2018 IOP Publishing Ltd and Sissa Medialab Journal of Cosmology and Astroparticle Physics **2018** (2018), 10.1088/1475-7516/2018/09/025.

[109] E. V. Linder, Phys. Rev. Lett. **90**, 091301 (2003), arXiv:astro-ph/0208512 [astro-ph] .

[110] E. Di Valentino, O. Mena, S. Pan, L. Visinelli, W. Yang, A. Melchiorri, D. F. Mota, A. G. Riess, and J. Silk, Classical and Quantum Gravity **38**, 153001 (2021), arXiv:2103.01183 [astro-ph.CO] .

[111] M. Cranmer, arXiv e-prints , arXiv:2305.01582 (2023), arXiv:2305.01582 [astro-ph.IM] .

[112] D. J. Bartlett, H. Desmond, and P. G. Ferreira, arXiv e-prints , arXiv:2211.11461 (2022), arXiv:2211.11461 [astro-ph.CO] .

[113] C. E. Rasmussen and C. K. I. Williams, *Gaussian processes for machine learning.*, Adaptive computation and machine learning (MIT Press, USA, Cambridge, MA 02142, 2006) pp. I–XVIII, 1–248.

[114] M. Seikel, C. Clarkson, and M. Smith, jcap **2012**, 036 (2012), arXiv:1204.2832 [astro-ph.CO] .

[115] A. Gómez-Valent and L. Amendola, JCAP **2018**, 051 (2018), arXiv:1802.01505 [astro-ph.CO] .

[116] Y.-F. Cai, M. Khurshudyan, and E. N. Saridakis, ApJ **888**, 62 (2020), arXiv:1907.10813 [astro-ph.CO] .

[117] K. Liao, A. Shafieloo, R. E. Keeley, and E. V. Linder, ApJ **895**, L29 (2020), arXiv:2002.10605 [astro-ph.CO] .

[118] A. Bonilla, S. Kumar, and R. C. Nunes, European Physical Journal C **81**, 127 (2021), arXiv:2011.07140 [astro-ph.CO] .

[119] D. Benisty, Physics of the Dark Universe **31**, 100766 (2021), arXiv:2005.03751 [astro-ph.CO] .

[120] B. L'Huillier, A. Shafieloo, and H. Kim, mnras **476**, 3263 (2018), arXiv:1712.04865 [astro-ph.CO] .

[121] S. Vagnozzi, A. Loeb, and M. Moresco, ApJ **908**, 84 (2021), arXiv:2011.11645 [astro-ph.CO] .

[122] Y. Yang and Y. Gong, MNRAS **504**, 3092 (2021), arXiv:2007.05714 [astro-ph.CO] .

[123] A. Shafieloo, B. L'Huillier, and A. A. Starobinsky, Phys. Rev. D **98**, 083526 (2018), arXiv:1804.04320 [astro-ph.CO] .

[124] F. Gerardi, M. Martinelli, and A. Silvestri, JCAP **2019**, 042 (2019), arXiv:1902.09423 [astro-ph.CO] .

[125] M.-J. Zhang and H. Li, European Physical Journal C **78**, 460 (2018), arXiv:1806.02981 [astro-ph.CO] .

[126] L. Perenon, M. Martinelli, S. Ilić, R. Maartens, M. Lochner, and C. Clarkson, arXiv e-prints , arXiv:2105.01613 (2021), arXiv:2105.01613 [astro-ph.CO] .

[127] J. R. Bond, A. H. Jaffe, and L. Knox, ApJ **533**, 19 (2000), arXiv:astro-ph/9808264 [astro-ph] .

[128] P. J. E. Peebles, *The large-scale structure of the universe* (Princeton Press, 1980).

[129] W. L. Freedman, B. F. Madore, T. Hoyt, I. S. Jang, R. Beaton, M. G. Lee, A. Monson, J. Neeley, and J. Rich, ApJ **891**, 57 (2020), arXiv:2002.01550 [astro-ph.GA] .

[130] D. W. Pesce, J. A. Braatz, M. J. Reid, A. G. Riess, D. Scolnic, J. J. Condon, F. Gao, C. Henkel, C. M. V. Impellizzeri, C. Y. Kuo, and K. Y. Lo, ApJ **891**, L1 (2020), arXiv:2001.09213 [astro-ph.CO] .

[131] M. Asgari, C.-A. Lin, B. Joachimi, B. Giblin, C. Heymans, H. Hildebrandt, A. Kannawadi, B. Stölzner, T. Tröster, J. L. van den Busch, A. H. Wright, M. Bilicki, C. Blake, J. de Jong, A. Dvornik, T. Erben, F. Getman, H. Hoekstra, F. Köhlinger, K. Kuijken, L. Miller, M. Radovich, P. Schneider, H. Shan, and E. Valentijn, A&A **645**, A104 (2021), arXiv:2007.15633 [astro-ph.CO] .

[132] J. Alsing, A. Heavens, and A. H. Jaffe, MNRAS **466**, 3272 (2017), arXiv:1607.00008 [astro-ph.CO] .

[133] R. C. Nunes and S. Vagnozzi, MNRAS **505**, 5427 (2021), arXiv:2106.01208 [astro-ph.CO] .

[134] J. Ruiz-Zapatero, B. Stölzner, B. Joachimi, M. Asgari, M. Bilicki, A. Dvornik, B. Giblin, C. Heymans, H. Hildebrandt, A. Kannawadi, K. Kuijken, TilmanTröster, J. L. van den Busch, and A. H. Wright, arXiv e-prints , arXiv:2105.09545 (2021), arXiv:2105.09545 [astro-ph.CO] .

[135] M. White, R. Zhou, J. DeRose, S. Ferraro, S.-F. Chen, N. Kokron, S. Bailey, D. Brooks, J. Garcia-Bellido, J. Guy, K. Honscheid, R. Kehoe, A. Kremin, M. Levi, N. Palanque-Delabrouille, C. Poppett, D. Schlegel, and G. Tarle, arXiv e-prints , arXiv:2111.09898 (2021), arXiv:2111.09898 [astro-ph.CO] .

[136] E.-K. Li, M. Du, Z.-H. Zhou, H. Zhang, and L. Xu, MNRAS **501**, 4452 (2021), arXiv:1911.12076 [astro-ph.CO] .

[137] A. Shafieloo, A. G. Kim, and E. V. Linder, Phys. Rev. D **85**, 123530 (2012), arXiv:1204.2272 [astro-ph.CO] .

[138] K. Dialektopoulos, J. Levi Said, J. Mifsud, J. Sultana, and K. Zarb Adami, arXiv e-prints , arXiv:2111.11462 (2021), arXiv:2111.11462 [astro-ph.CO] .

[139] J. Salvatier, T. Wiecki, and C. Fonnesbeck, arXiv e-prints , arXiv:1507.08050 (2015), arXiv:1507.08050 [stat.CO] .

[140] The Theano Development Team, R. Al-Rfou, G. Alain, A. Almahairi, C. Angermueller, D. Bahdanau, N. Ballas, F. Bastien, J. Bayer, A. Belikov, A. Belopolsky, Y. Bengio, A. Bergeron, J. Bergstra, V. Bisson, J. Bleecher Snyder, N. Bouchard, N. Boulanger-Lewandowski, X. Bouthillier, A. de Brébisson, O. Breuleux, P.-L. Carrier, K. Cho, J. Chorowski, P. Christiano, T. Cooijmans, M.-A. Côté, M. Côté, A. Courville, Y. N. Dauphin, O. Delalleau, J. Demouth, G. Desjardins, S. Dieleman, L. Dinh, M. Ducoffe, V. Dumoulin, S. Ebrahimi Kahou, D. Erhan, Z. Fan, O. Firat, M. Germain, X. Glorot, I. Goodfellow, M. Graham, C. Gulcehre, P. Hamel, I. Harlouchet, J.-P. Heng, B. Hidasi, S. Honari, A. Jain, S. Jean, K. Jia, M. Korobov, V. Kulkarni, A. Lamb, P. Lamblin, E. Larsen, C. Laurent, S. Lee, S. Lefrancois, S. Lemieux, N. Léonard, Z. Lin, J. A. Livezey, C. Lorenz, J. Lowin, Q. Ma, P.-A. Manzagol, O. Mastropietro, R. T. McGibbon, R. Memisevic, B. van Merriënboer, V. Michalski, M. Mirza, A. Orlandi, C. Pal, R. Pascanu, M. Pezeshki, C. Raffel, D. Renshaw, M. Rocklin, A. Romero, M. Roth, P. Sadowski, J. Salvatier, F. Savard, J. Schlüter, J. Schulman, G. Schwartz, I. Vlad Serban, D. Serdyuk, S. Shabanian, É. Simon, S. Spieckermann, S. Ramana Subramanyam, J. Sygnowski, J. Tanguay, G. van Tulder, J. Turian, S. Urban, P. Vincent, F. Visin, H. de Vries, D. Warde-Farley, D. J. Webb, M. Willson, K. Xu, L. Xue, L. Yao, S. Zhang, and Y. Zhang, arXiv e-prints , arXiv:1605.02688 (2016), arXiv:1605.02688 [cs.SC] .

[141] C. C. Margossian, arXiv e-prints , arXiv:1811.05031 (2018), arXiv:1811.05031 [cs.MS] .

[142] S. Joudaki, M. Kaplinghat, R. Keeley, and D. Kirkby, Phys. Rev. D **97**, 123501 (2018), arXiv:1710.04236 [astro-ph.CO] .

[143] A. Gelman, D. Simpson, and M. Betancourt, Entropy **19**, 555 (2017), arXiv:1708.07487 [stat.ME] .

[144] D. M. Scolnic, D. O. Jones, A. Rest, Y. C. Pan, R. Chornock, R. J. Foley, M. E. Huber, R. Kessler, G. Narayan, A. G. Riess, S. Rodney, E. Berger, D. J. Brout, P. J. Challis, M. Drout, D. Finkbeiner, R. Lunnan, R. P. Kirshner, N. E. Sanders, E. Schlafly, S. Smartt, C. W. Stubbs, J. Tonry, W. M. Wood-Vasey, M. Foley, J. Hand, E. Johnson, W. S. Burgett, K. C. Chambers, P. W. Draper, K. W. Hodapp, N. Kaiser, R. P. Kudritzki, E. A. Magnier, N. Metcalfe, F. Bresolin, E. Gall, R. Kotak, M. McCrum, and K. W. Smith, ApJ **859**, 101 (2018), arXiv:1710.00845 [astro-ph.CO] .

[145] S. Alam, M. Aubert, S. Avila, C. Balland, J. E. Bautista, M. A. Bershady, D. Bizyaev, M. R. Blanton, A. S. Bolton, J. Bovy, J. Brinkmann, J. R. Brownstein, E. Burtin, S. Chabanier, M. J. Chapman, P. D. Choi, C.-H. Chuang, J. Comparat, M.-C. Cousinou, A. Cuceu, K. S. Dawson, S. de la Torre, A. de Mattia, V. d. S. Agathe, H. d. M. des Bourboux, S. Escoffier, T. Etourneau, J. Farr, A. Font-Ribera, P. M. Frinchaboy, S. Fromenteau, H. Gil-Marín, J.-M. Le Goff, A. X. Gonzalez-Morales, V. Gonzalez-Perez, K. Grabowski, J. Guy, A. J. Hawken, J. Hou, H. Kong, J. Parker, M. Klaene, J.-P. Kneib, S. Lin, D. Long, B. W. Lyke, A. de la Macorra, P. Martini, K. Masters, F. G. Mohammad, J. Moon, E.-M. Mueller, A. Muñoz-Gutiérrez, A. D. Myers, S. Nadathur, R. Neveux, J. A. Newman, P. Noterdaeme, A. Oravetz, D. Oravetz, N. Palanque-Delabrouille, K. Pan, R. Paviot, W. J. Percival, I. Pérez-Ràfols, P. Petitjean, M. M. Pieri, A. Prakash, A. Raichoor, C. Ravoux, M. Rezaie, J. Rich, A. J. Ross, G. Rossi, R. Ruggeri, V. Ruhlmann-Kleider, A. G. Sánchez, F. J. Sánchez, J. R. Sánchez-Gallego, C. Sayres, D. P. Schneider, H.-J. Seo, A. Shafieloo, A. Slosar, A. Smith, J. Stermer, A. Tamone, J. L. Tinker, R. Tojeiro, M. Vargas-Magaña, A. Variu, Y. Wang, B. A. Weaver, A.-M. Weijmans, C. Yèche, P. Zarrouk, C. Zhao, G.-B. Zhao, and Z. Zheng, Phys. Rev. D **103**, 083533 (2021), arXiv:2007.08991 [astro-ph.CO] .

[146] C. Blake, S. Brough, M. Colless, C. Contreras, W. Couch, S. Croom, D. Croton, T. M. Davis, M. J. Drinkwater, K. Forster, D. Gilbank, M. Gladders, K. Glazebrook, B. Jelliffe, R. J. Jurek, I. h. Li, B. Madore, D. C. Martin, K. Pimbblet, G. B. Poole, M. Pracy, R. Sharp, E. Wisnioski, D. Woods, T. K. Wyder, and H. K. C. Yee, MNRAS **425**, 405 (2012), arXiv:1204.3674 [astro-ph.CO] .

[147] B. E. Stahl, T. de Jaeger, S. S. Boruah, W. Zheng, A. V. Filippenko, and M. J. Hudson, mnras **505**, 2349 (2021), arXiv:2105.05185 [astro-ph.CO] .

[148] A. Font-Ribera, P. McDonald, N. Mostek, B. A. Reid, H.-J. Seo, and A. Slosar, JCAP **2014**, 023 (2014), arXiv:1308.4164 [astro-ph.CO] .

[149] R. Jimenez and A. Loeb, ApJ **573**, 37 (2002), arXiv:astro-ph/0106145 [astro-ph] .

[150] A. Cimatti, E. Daddi, and A. Renzini, aap **453**, L29 (2006), arXiv:astro-ph/0605353 [astro-ph] .

[151] J. Thomas, R. P. Saglia, R. Bender, D. Thomas, K. Gebhardt, J. Magorrian, E. M. Corsini, G. Wegner, and S. Seitz, MNRAS **415**, 545 (2011), arXiv:1103.3414 [astro-ph.CO] .

[152] M. Moresco, MNRAS **450**, L16 (2015), arXiv:1503.01116 [astro-ph.CO] .

[153] M. Moresco, R. Jimenez, L. Verde, L. Pozzetti, A. Cimatti, and A. Citro, ApJ **868**, 84 (2018), arXiv:1804.05864 [astro-ph.CO] .

[154] M. Moresco, R. Jimenez, L. Verde, A. Cimatti, and L. Pozzetti, ApJ **898**, 82 (2020), arXiv:2003.07362 [astro-ph.GA] .

[155] F. Hoyle and W. A. Fowler, ApJ **132**, 565 (1960).

[156] S. A. Colgate and C. McKee, ApJ **157**, 623 (1969).

[157] M. M. Phillips, P. Lira, N. B. Suntzeff, R. A. Schommer, M. Hamuy, and J. Maza, AJ **118**, 1766 (1999), arXiv:astro-ph/9907052 [astro-ph] .

[158] W. L. Freedman, B. F. Madore, B. K. Gibson, L. Ferrarese, D. D. Kelson, S. Sakai, J. R. Mould, J. Kennicutt, Robert C., H. C. Ford, J. A. Graham, J. P. Huchra, S. M. G. Hughes, G. D. Illingworth, L. M. Macri, and P. B. Stetson, ApJ **553**, 47 (2001), arXiv:astro-ph/0012376 [astro-ph] .

[159] P. Shah, P. Lemos, and O. Lahav, arXiv e-prints , arXiv:2109.01161 (2021), arXiv:2109.01161 [astro-ph.CO] .

[160] P. J. E. Peebles and J. T. Yu, ApJ **162**, 815 (1970).

[161] W. Hu and S. Dodelson, ARA&A **40**, 171 (2002), arXiv:astro-ph/0110414 [astro-ph] .

[162] E. Komatsu, J. Dunkley, M. R. Nolta, C. L. Bennett, B. Gold, G. Hinshaw, N. Jarosik, D. Larson, M. Limon, L. Page, D. N. Spergel, M. Halpern, R. S. Hill, A. Kogut, S. S. Meyer, G. S. Tucker, J. L. Weiland, E. Wollack, and E. L. Wright, ApJS **180**, 330 (2009), arXiv:0803.0547 [astro-ph] .

[163] M. R. Blanton, M. A. Bershady, B. Abolfathi, F. D. Albareti, C. Allende Prieto, A. Almeida, J. Alonso-García, F. Anders, S. F. Anderson, B. Andrews, E. Aquino-Ortíz, A. Aragón-Salamanca, M. Argudo-Fernández, E. Armengaud, E. Aubourg, V. Avila-Reese, C. Badenes, S. Bailey, K. A. Barger, J. Barrera-Ballesteros, C. Bartosz, D. Bates, F. Baumgarten, J. Bautista, R. Beaton,

T. C. Beers, F. Belfiore, C. F. Bender, A. A. Berlind, M. Bernardi, F. Beutler, J. C. Bird, D. Bizyaev, G. A. Blanc, M. Blomqvist, A. S. Bolton, M. Boquien, J. Borissova, R. van den Bosch, J. Bovy, W. N. Brandt, J. Brinkmann, J. R. Brownstein, K. Bundy, A. J. Burgasser, E. Burtin, N. G. Busca, M. Cappellari, M. L. Delgado Carigi, J. K. Carlberg, A. Carnero Rosell, R. Carrera, N. J. Chanover, B. Cherinka, E. Cheung, Y. Gómez Maqueo Chew, C. Chiappini, P. D. Choi, D. Chojnowski, C.-H. Chuang, H. Chung, R. F. Cirolini, N. Clerc, R. E. Cohen, J. Comparat, L. da Costa, M.-C. Cousinou, K. Covey, J. D. Crane, R. A. C. Croft, I. Cruz-Gonzalez, D. Garrido Cuadra, K. Cunha, G. J. Damke, J. Darling, R. Davies, K. Dawson, A. de la Macorra, F. Dell'Agli, N. De Lee, T. Delubac, F. Di Mille, A. Diamond-Stanic, M. Cano-Díaz, J. Donor, J. J. Downes, N. Drory, H. du Mas des Bourboux, C. J. Duckworth, T. Dwelly, J. Dyer, G. Ebelke, A. D. Eigenbrot, D. J. Eisenstein, E. Emsellem, M. Eracleous, S. Escoffier, M. L. Evans, X. Fan, E. Fernández-Alvar, J. G. Fernandez-Trincado, D. K. Feuillet, A. Finoguenov, S. W. Fleming, A. Font-Ribera, A. Fredrickson, G. Freischlad, P. M. Frinchaboy, C. E. Fuentes, L. Galbany, R. Garcia-Dias, D. A. García-Hernández, P. Gaulme, D. Geisler, J. D. Gelfand, H. Gil-Marín, B. A. Gillespie, D. Goddard, V. Gonzalez-Perez, K. Grabowski, P. J. Green, C. J. Grier, J. E. Gunn, H. Guo, J. Guy, A. Hagen, C. Hahn, M. Hall, P. Harding, S. Hasselquist, S. L. Hawley, F. Hearty, J. I. Gonzalez Hernández, S. Ho, D. W. Hogg, K. Holley-Bockelmann, J. A. Holtzman, P. H. Holzer, J. Huehnerhoff, T. A. Hutchinson, H. S. Hwang, H. J. Ibarra-Medel, G. da Silva Ilha, I. I. Ivans, K. Ivory, K. Jackson, T. W. Jensen, J. A. Johnson, A. Jones, H. Jönsson, E. Jullo, V. Kamble, K. Kinemuchi, D. Kirkby, F.-S. Kitaura, M. Klaene, G. R. Knapp, J.-P. Kneib, J. A. Kollmeier, I. Lacerna, R. R. Lane, D. Lang, D. R. Law, D. Lazarz, Y. Lee, J.-M. Le Goff, F.-H. Liang, C. Li, H. Li, J. Lian, M. Lima, L. Lin, Y.-T. Lin, S. Bertran de Lis, C. Liu, M. A. C. de Icaza Lizaola, D. Long, S. Lucatello, B. Lundgren, N. K. MacDonald, A. Deconto Machado, C. L. MacLeod, S. Mahadevan, M. A. Geimba Maia, R. Maiolino, S. R. Majewski, E. Malanushenko, V. Malanushenko, A. Manchado, S. Mao, C. Maraston, R. Marques-Chaves, T. Masseron, K. L. Masters, C. K. McBride, R. M. McDermid, B. McGrath, I. D. McGreer, N. Medina Peña, M. Melendez, A. Merloni, M. R. Merrifield, S. Meszaros, A. Meza, I. Minchev, D. Minniti, T. Miyaji, S. More, J. Mulchaey, F. Müller-Sánchez, D. Muna, R. R. Munoz, A. D. Myers, P. Nair, K. Nandra, J. Correa do Nascimento, A. Negrete, M. Ness, J. A. Newman, R. C. Nichol, D. L. Nidever, C. Nitschelm,

P. Ntelis, J. E. O'Connell, R. J. Oelkers, A. Oravetz, D. Oravetz, Z. Pace, N. Padilla, N. Palanque-Delabrouille, P. Alonso Palicio, K. Pan, J. K. Parejko, T. Parikh, I. Pâris, C. Park, A. Y. Patten, S. Peirani, M. Pellejero-Ibanez, S. Penny, W. J. Percival, I. Perez-Fournon, P. Petitjean, M. M. Pieri, M. Pinsonneault, A. Pisani, R. Poleski, F. Prada, A. Prakash, A. B. d. A. Queiroz, M. J. Raddick, A. Raichoor, S. Barboza Rembold, H. Richstein, R. A. Riffel, R. Riffel, H.-W. Rix, A. C. Robin, C. M. Rockosi, S. Rodríguez-Torres, A. Roman-Lopes, C. Román-Zúñiga, M. Rosado, A. J. Ross, G. Rossi, J. Ruan, R. Ruggeri, E. S. Rykoff, S. Salazar-Albornoz, M. Salvato, A. G. Sánchez, D. S. Aguado, J. R. Sánchez-Gallego, F. A. Santana, B. X. Santiago, C. Sayres, R. P. Schiavon, J. da Silva Schimoia, E. F. Schlafly, D. J. Schlegel, D. P. Schneider, M. Schultheis, W. J. Schuster, A. Schwope, H.-J. Seo, Z. Shao, S. Shen, M. Shetrone, M. Shull, J. D. Simon, D. Skinner, M. F. Skrutskie, A. Slosar, V. V. Smith, J. S. Sobeck, F. Sobreira, G. Somers, D. Souto, D. V. Stark, K. Stassun, F. Stauffer, M. Steinmetz, T. Storchi-Bergmann, A. Streblyanska, G. S. Stringfellow, G. Suárez, J. Sun, N. Suzuki, L. Szigeti, M. Taghizadeh-Popp, B. Tang, C. Tao, J. Tayar, M. Tembe, J. Teske, A. R. Thakar, D. Thomas, B. A. Thompson, J. L. Tinker, P. Tissera, R. Tojeiro, H. Hernandez Toledo, S. de la Torre, C. Tremonti, N. W. Troup, O. Valenzuela, I. Martinez Valpuesta, J. Vargas-González, M. Vargas-Magaña, J. A. Vazquez, S. Villanova, M. Vivek, N. Vogt, D. Wake, R. Walterbos, Y. Wang, B. A. Weaver, A.-M. Weijmans, D. H. Weinberg, K. B. Westfall, D. G. Whelan, V. Wild, J. Wilson, W. M. Wood-Vasey, D. Wylezalek, T. Xiao, R. Yan, M. Yang, J. E. Ybarra, C. Yèche, N. Zakamska, O. Zamora, P. Zarrouk, G. Zasowski, K. Zhang, G.-B. Zhao, Z. Zheng, Z. Zheng, X. Zhou, Z.-M. Zhou, G. B. Zhu, M. Zoccali, and H. Zou, AJ **154**, 28 (2017), arXiv:1703.00052 [astro-ph.GA] .

[164] K. S. Dawson, J.-P. Kneib, W. J. Percival, S. Alam, F. D. Albareti, S. F. Anderson, E. Armengaud, É. Aubourg, S. Bailey, J. E. Bautista, A. A. Berlind, M. A. Bershady, F. Beutler, D. Bizyaev, M. R. Blanton, M. Blomqvist, A. S. Bolton, J. Bovy, W. N. Brandt, J. Brinkmann, J. R. Brownstein, E. Burtin, N. G. Busca, Z. Cai, C.-H. Chuang, N. Clerc, J. Comparat, F. Cope, R. A. C. Croft, I. Cruz-Gonzalez, L. N. da Costa, M.-C. Cousinou, J. Darling, A. de la Macorra, S. de la Torre, T. Delubac, H. du Mas des Bourboux, T. Dwelly, A. Ealet, D. J. Eisenstein, M. Eracleous, S. Escoffier, X. Fan, A. Finoguenov, A. Font-Ribera, P. Frinchaboy, P. Gaulme, A. Georgakakis, P. Green, H. Guo, J. Guy, S. Ho, D. Holder, J. Huehnerhoff, T. Hutchinson, Y. Jing, E. Jullo,

V. Kamble, K. Kinemuchi, D. Kirkby, F.-S. Kitaura, M. A. Klaene, R. R. Laher, D. Lang, P. Laurent, J.-M. Le Goff, C. Li, Y. Liang, M. Lima, Q. Lin, W. Lin, Y.-T. Lin, D. C. Long, B. Lundgren, N. MacDonald, M. A. Geimba Maia, E. Malanushenko, V. Malanushenko, V. Mariappan, C. K. McBride, I. D. McGreer, B. Ménard, A. Merloni, A. Meza, A. D. Montero-Dorta, D. Muna, A. D. Myers, K. Nandra, T. Naugle, J. A. Newman, P. Noterdaeme, P. Nugent, R. Ogando, M. D. Olmstead, A. Oravetz, D. J. Oravetz, N. Padmanabhan, N. Palanque-Delabrouille, K. Pan, J. K. Parejko, I. Pâris, J. A. Peacock, P. Petitjean, M. M. Pieri, A. Pisani, F. Prada, A. Prakash, A. Raichoor, B. Reid, J. Rich, J. Ridl, S. Rodriguez-Torres, A. Carnero Rosell, A. J. Ross, G. Rossi, J. Ruan, M. Salvato, C. Sayres, D. P. Schneider, D. J. Schlegel, U. Seljak, H.-J. Seo, B. Sesar, S. Shandera, Y. Shu, A. Slosar, F. Sobreira, A. Streblyanska, N. Suzuki, D. Taylor, C. Tao, J. L. Tinker, R. Tojeiro, M. Vargas-Magaña, Y. Wang, B. A. Weaver, D. H. Weinberg, M. White, W. M. Wood-Vasey, C. Yeche, Z. Zhai, C. Zhao, G.-b. Zhao, Z. Zheng, G. Ben Zhu, and H. Zou, AJ **151**, 44 (2016), arXiv:1508.04473 [astro-ph.CO] .

[165] J. Hou, A. G. Sánchez, A. J. Ross, A. Smith, R. Neveux, J. Bautista, E. Burtin, C. Zhao, R. Scoccimarro, K. S. Dawson, A. de Mattia, A. de la Macorra, H. du Mas des Bourboux, D. J. Eisenstein, H. Gil-Marín, B. W. Lyke, F. G. Mohammad, E.-M. Mueller, W. J. Percival, G. Rossi, M. Vargas Magaña, P. Zarrouk, G.-B. Zhao, J. Brinkmann, J. R. Brownstein, C.-H. Chuang, A. D. Myers, J. A. Newman, D. P. Schneider, and M. Vivek, MNRAS **500**, 1201 (2021), arXiv:2007.08998 [astro-ph.CO] .

[166] N. Kaiser, MNRAS **227**, 1 (1987).

[167] A. J. S. Hamilton, "Linear Redshift Distortions: a Review," in *The Evolving Universe*, Vol. 231 (in "The Evolving Universe" ed. D. Hamilton, Kluwer Academic, p. 185-275 (1998), 1998) p. 185.

[168] F. Beutler, H.-J. Seo, S. Saito, C.-H. Chuang, A. J. Cuesta, D. J. Eisenstein, H. Gil-Marín, J. N. Grieb, N. Hand, F.-S. Kitaura, C. Modi, R. C. Nichol, M. D. Olmstead, W. J. Percival, F. Prada, A. G. Sánchez, S. Rodriguez-Torres, A. J. Ross, N. P. Ross, D. P. Schneider, J. Tinker, R. Tojeiro, and M. Vargas-Magaña, MNRAS **466**, 2242 (2017), arXiv:1607.03150 [astro-ph.CO] .

[169] A. G. Riess, in *Cosmic Flows Workshop*, Astronomical Society of the Pacific Conference Series, Vol. 201, edited by S. Courteau and J. Willick (2000) p. 80, arXiv:astro-ph/9908237 [astro-ph] .

[170] R. B. Tully, H. M. Courtois, and J. G. Sorce, AJ **152**, 50 (2016), arXiv:1605.01765 [astro-ph.CO] .

[171] M. J. Drinkwater, R. J. Jurek, C. Blake, D. Woods, K. A. Pimbblet, K. Glazebrook, R. Sharp, M. B. Pracy, S. Brough, M. Colless, W. J. Couch, S. M. Croom, T. M. Davis, D. Forbes, K. Forster, D. G. Gilbank, M. Gladders, B. Jelliffe, N. Jones, I. H. Li, B. Madore, D. C. Martin, G. B. Poole, T. Small, E. Wisnioski, T. Wyder, and H. K. C. Yee, MNRAS **401**, 1429 (2010), arXiv:0911.4246 [astro-ph.CO] .

[172] A. Krolewski, S. Ferraro, and M. White, arXiv e-prints , arXiv:2105.03421 (2021), arXiv:2105.03421 [astro-ph.CO] .

[173] Q. Hang, S. Alam, J. A. Peacock, and Y.-C. Cai, MNRAS **501**, 1481 (2021), arXiv:2010.00466 [astro-ph.CO] .

[174] M. Raveri, P. Bull, A. Silvestri, and L. Pogosian, Phys. Rev. D **96**, 083509 (2017), arXiv:1703.05297 [astro-ph.CO] .

[175] J. Espejo, S. Peirone, M. Raveri, K. Koyama, L. Pogosian, and A. Silvestri, Phys. Rev. D **99**, 023512 (2019), arXiv:1809.01121 [astro-ph.CO] .

[176] M. Park, M. Raveri, and B. Jain, Phys. Rev. D **103**, 103530 (2021), arXiv:2101.04666 [astro-ph.CO] .

[177] M. Raveri, L. Pogosian, K. Koyama, M. Martinelli, A. Silvestri, G.-B. Zhao, J. Li, S. Peirone, and A. Zucca, arXiv e-prints , arXiv:2107.12990 (2021), arXiv:2107.12990 [astro-ph.CO] .

[178] L. Pogosian, M. Raveri, K. Koyama, M. Martinelli, A. Silvestri, and G.-B. Zhao, arXiv e-prints , arXiv:2107.12992 (2021), arXiv:2107.12992 [astro-ph.CO] .

[179] T. Baker, P. Ferreira, and C. Skordis, prd **89**, 024026 (2014), arXiv:1310.1086 [astro-ph.CO] .

[180] V. Rajpaul, arXiv e-prints , arXiv:1202.1643 (2012), arXiv:1202.1643 [astro-ph.IM] .

[181] S. Nesseris, in *Journal of Physics Conference Series*, Journal of Physics Conference Series, Vol. 283 (2011) p. 012025, arXiv:1011.1859 [astro-ph.CO] .

[182] I. Gómez-Vargas, J. A. Vázquez, R. Medel Esquivel, and R. García-Salcedo, arXiv e-prints , arXiv:2104.00595 (2021), arXiv:2104.00595 [astro-ph.CO] .

[183] P. G. Ferreira and C. Skordis, Phys. Rev. D **81**, 104020 (2010), arXiv:1003.4231 [astro-ph.CO] .

[184] P. G. Ferreira, Ann. Rev. Astron. Astrophys. **57**, 335 (2019), arXiv:1902.10503 [astro-ph.CO] .

[185] E. G. Adelberger, B. R. Heckel, and A. E. Nelson, Ann. Rev. Nucl. Part. Sci. **53**, 77 (2003), arXiv:hep-ph/0307284 .

[186] T. Clifton, P. G. Ferreira, A. Padilla, and C. Skordis, Phys. Rept. **513**, 1 (2012), arXiv:1106.2476 [astro-ph.CO] .

[187] A. Joyce, B. Jain, J. Khoury, and M. Trodden, Phys. Rept. **568**, 1 (2015), arXiv:1407.0059 [astro-ph.CO] .

[188] F. Simpson and J. A. Peacock, Phys. Rev. D **81**, 043512 (2010), arXiv:0910.3834 [astro-ph.CO] .

[189] E.-M. Mueller, W. Percival, E. Linder, S. Alam, G.-B. Zhao, A. G. Sánchez, F. Beutler, and J. Brinkmann, Mon. Not. Roy. Astron. Soc. **475**, 2122 (2018), arXiv:1612.00812 [astro-ph.CO] .

[190] L. Pèrenon, F. Piazza, C. Marinoni, and L. Hui, JCAP **2015**, 029 (2015), arXiv:1506.03047 [astro-ph.CO] .

[191] D. Traykova, E. Bellini, P. G. Ferreira, C. García-García, J. Noller, and M. Zumalacárregui, Phys. Rev. D **104**, 083502 (2021), arXiv:2103.11195 [astro-ph.CO] .

[192] D. Harrison, arXiv e-prints , arXiv:2110.06209 (2021), arXiv:2110.06209 [cs.LG] .

[193] A. Aizpuru, R. Arjona, and S. Nesseris, Phys. Rev. D **104**, 043521 (2021), arXiv:2106.00428 [astro-ph.CO] .

[194] A. Pezzotta, S. de la Torre, J. Bel, B. R. Granett, L. Guzzo, J. A. Peacock, B. Garilli, M. Scodeggio, M. Bolzonella, U. Abbas, C. Adami, D. Bottini, A. Cappi, O. Cucciati, I. Davidzon, P. Franzetti, A. Fritz, A. Iovino, J. Krywult, V. Le Brun, O. Le Fèvre, D. Maccagni, K. Małek, F. Marulli, M. Polletta, A. Pollo, L. A. M. Tasca, R. Tojeiro, D. Vergani, A. Zanichelli, S. Arnouts, E. Branchini, J. Coupon, G. De Lucia, J. Koda, O. Ilbert, F. Mohammad, T. Moutard, and L. Moscardini, aap **604**, A33 (2017), arXiv:1612.05645 [astro-ph.CO] .

[195] F. Beutler, C. Blake, M. Colless, D. H. Jones, L. Staveley-Smith, G. B. Poole, L. Campbell, Q. Parker, W. Saunders, and F. Watson, mnras **423**, 3430 (2012), arXiv:1204.4725 [astro-ph.CO] .

[196] T. Okumura, C. Hikage, T. Totani, M. Tonegawa, H. Okada, K. Glazebrook, C. Blake, P. G. Ferreira, S. More, A. Taruya, S. Tsujikawa, M. Akiyama, G. Dalton, T. Goto, T. Ishikawa, F. Iwamuro, T. Matsubara, T. Nishimichi, K. Ohta, I. Shimizu, R. Takahashi, N. Takato, N. Tamura, K. Yabe, and N. Yoshida, PASJ **68**, 38 (2016), arXiv:1511.08083 [astro-ph.CO] .

[197] S.-g. Hwang, B. L'Huillier, R. E. Keeley, M. J. Jee, and A. Shafieloo, arXiv e-prints , arXiv:2206.15081 (2022), arXiv:2206.15081 [astro-ph.CO] .

[198] DESI Collaboration, A. Aghamousa, J. Aguilar, S. Ahlen, S. Alam, L. E. Allen, C. Allende Prieto, J. Annis, S. Bailey, C. Balland, O. Ballester, C. Baltay, L. Beaufore, C. Bebek, T. C. Beers, E. F. Bell, J. L. Bernal, R. Besuner, F. Beutler, C. Blake, H. Bleuler, M. Blomqvist, R. Blum, A. S. Bolton, C. Briceno, D. Brooks, J. R. Brownstein, E. Buckley-Geer, A. Burden, E. Burtin, N. G. Busca, R. N. Cahn, Y.-C. Cai, L. Cardiel-Sas, R. G. Carlberg, P.-H. Carton, R. Casas, F. J. Castander, J. L. Cervantes-Cota, T. M. Claybaugh, M. Close, C. T. Coker, S. Cole, J. Comparat, A. P. Cooper, M. C. Cousinou, M. Crocce, J.-G. Cuby, D. P. Cunningham, T. M. Davis, K. S. Dawson, A. de la Macorra, J. De Vicente, T. Delubac, M. Derwent, A. Dey, G. Dhungana, Z. Ding, P. Doel, Y. T. Duan, A. Ealet, J. Edelstein, S. Eftekharzadeh, D. J. Eisenstein, A. Elliott, S. Escoffier, M. Evatt, P. Fagrelius, X. Fan, K. Fanning, A. Farahi, J. Farihi, G. Favole, Y. Feng, E. Fernandez, J. R. Findlay, D. P. Finkbeiner, M. J. Fitzpatrick, B. Flaugher, S. Flender, A. Font-Ribera, J. E. Forero-Romero, P. Fosalba, C. S. Frenk, M. Fumagalli, B. T. Gaensicke, G. Gallo, J. Garcia-Bellido, E. Gaztanaga, N. Pietro Gentile Fusillo, T. Gerard, I. Gershkovich, T. Giannantonio, D. Gillet, G. Gonzalez-de-Rivera, V. Gonzalez-Perez, S. Gott, O. Graur, G. Gutierrez, J. Guy, S. Habib, H. Heetderks, I. Heetderks, K. Heitmann, W. A. Hellwing, D. A. Herrera, S. Ho, S. Holland, K. Honscheid, E. Huff, T. A. Hutchinson, D. Huterer, H. S. Hwang, J. M. Illa Laguna, Y. Ishikawa, D. Jacobs, N. Jeffrey, P. Jelinsky, E. Jennings, L. Jiang, J. Jimenez, J. Johnson, R. Joyce, E. Jullo, S. Juneau, S. Kama, A. Karcher, S. Karkar, R. Kehoe, N. Kennamer, S. Kent, M. Kilbinger, A. G. Kim, D. Kirkby, T. Kisner, E. Kitanidis, J.-P. Kneib, S. Koposov, E. Kovacs, K. Koyama, A. Kremin, R. Kron, L. Kronig, A. Kueter-Young, C. G. Lacey, R. Lafever, O. Lahav, A. Lambert, M. Lampton, M. Landriau, D. Lang, T. R. Lauer, J.-M. Le Goff, L. Le Guillou, A. Le Van Suu, J. H. Lee, S.-J. Lee, D. Leitner, M. Lesser, M. E. Levi, B. L'Huillier, B. Li, M. Liang, H. Lin, E. Linder, S. R. Loebman, Z. Lukić, J. Ma, N. MacCrann, C. Magneville, L. Makarem, M. Manera, C. J. Manser, R. Marshall, P. Martini, R. Massey, T. Matheson, J. McCauley, P. McDonald, I. D. McGreer, A. Meisner, N. Metcalfe, T. N. Miller, R. Miquel, J. Moustakas, A. Myers, M. Naik, J. A. Newman, R. C. Nichol, A. Nicola, L. Nicolati da Costa, J. Nie, G. Niz, P. Norberg, B. Nord, D. Norman, P. Nugent, T. O'Brien, M. Oh,

K. A. G. Olsen, C. Padilla, H. Padmanabhan, N. Padmanabhan, N. Palanque-Delabrouille, A. Palmese, D. Pappalardo, I. Pâris, C. Park, A. Patej, J. A. Peacock, H. V. Peiris, X. Peng, W. J. Percival, S. Perruchot, M. M. Pieri, R. Pogge, J. E. Pollack, C. Poppett, F. Prada, A. Prakash, R. G. Probst, D. Rabinowitz, A. Raichoor, C. H. Ree, A. Refregier, X. Regal, B. Reid, K. Reil, M. Rezaie, C. M. Rockosi, N. Roe, S. Ronayette, A. Roodman, A. J. Ross, N. P. Ross, G. Rossi, E. Rozo, V. Ruhlmann-Kleider, E. S. Rykoff, C. Sabiu, L. Samushia, E. Sanchez, J. Sanchez, D. J. Schlegel, M. Schneider, M. Schubnell, A. Secroun, U. Seljak, H.-J. Seo, S. Serrano, A. Shafieloo, H. Shan, R. Sharples, M. J. Sholl, W. V. Shourt, J. H. Silber, D. R. Silva, M. M. Sirk, A. Slosar, A. Smith, G. F. Smoot, D. Som, Y.-S. Song, D. Sprayberry, R. Staten, A. Stefanik, G. Tarle, S. Sien Tie, J. L. Tinker, R. Tojeiro, F. Valdes, O. Valenzuela, M. Valluri, M. Vargas-Magana, L. Verde, A. R. Walker, J. Wang, Y. Wang, B. A. Weaver, C. Weaverdyck, R. H. Wechsler, D. H. Weinberg, M. White, Q. Yang, C. Yeche, T. Zhang, G.-B. Zhao, Y. Zheng, X. Zhou, Z. Zhou, Y. Zhu, H. Zou, and Y. Zu, arXiv e-prints , arXiv:1611.00036 (2016), arXiv:1611.00036 [astro-ph.IM] .

[199] Ž. Ivezić, S. M. Kahn, J. A. Tyson, B. Abel, E. Acosta, R. Allsman, D. Alonso, Y. AlSayyad, S. F. Anderson, J. Andrew, J. R. P. Angel, G. Z. Angeli, R. Ansari, P. Antilogus, C. Araujo, R. Armstrong, K. T. Arndt, P. Astier, É. Aubourg, N. Auza, T. S. Axelrod, D. J. Bard, J. D. Barr, A. Barrau, J. G. Bartlett, A. E. Bauer, B. J. Bauman, S. Baumont, E. Bechtol, K. Bechtol, A. C. Becker, J. Becla, C. Beldica, S. Bellavia, F. B. Bianco, R. Biswas, G. Blanc, J. Blazek, R. D. Blandford, J. S. Bloom, J. Bogart, T. W. Bond, M. T. Booth, A. W. Borgland, K. Borne, J. F. Bosch, D. Boutigny, C. A. Brackett, A. Bradshaw, W. N. Brandt, M. E. Brown, J. S. Bullock, P. Burchat, D. L. Burke, G. Cagnoli, D. Calabrese, S. Callahan, A. L. Callen, J. L. Carlin, E. L. Carlson, S. Chandrasekharan, G. Charles-Emerson, S. Chesley, E. C. Cheu, H.-F. Chiang, J. Chiang, C. Chirino, D. Chow, D. R. Ciardi, C. F. Claver, J. Cohen-Tanugi, J. J. Cockrum, R. Coles, A. J. Connolly, K. H. Cook, A. Cooray, K. R. Covey, C. Cribbs, W. Cui, R. Cutri, P. N. Daly, S. F. Daniel, F. Daruich, G. Daubard, G. Daues, W. Dawson, F. Delgado, A. Dellapenna, R. de Peyster, M. de Val-Borro, S. W. Digel, P. Doherty, R. Dubois, G. P. Dubois-Felsmann, J. Durech, F. Economou, T. Eifler, M. Eracleous, B. L. Emmons, A. Fausti Neto, H. Ferguson, E. Figueroa, M. Fisher-Levine, W. Focke, M. D. Foss, J. Frank, M. D. Freemon, E. Gangler, E. Gawiser, J. C. Geary, P. Gee, M. Geha, C. J. B. Gessner, R. R. Gibson, D. K. Gilmore, T. Glanzman, W. Glick, T. Goldina,

D. A. Goldstein, I. Goodenow, M. L. Graham, W. J. Gressler, P. Gris, L. P. Guy, A. Guyonnet, G. Haller, R. Harris, P. A. Hascall, J. Haupt, F. Hernandez, S. Herrmann, E. Hileman, J. Hoblitt, J. A. Hodgson, C. Hogan, J. D. Howard, D. Huang, M. E. Huffer, P. Ingraham, W. R. Innes, S. H. Jacoby, B. Jain, F. Jammes, M. J. Jee, T. Jenness, G. Jernigan, D. Jevremović, K. Johns, A. S. Johnson, M. W. G. Johnson, R. L. Jones, C. Juramy-Gilles, M. Jurić, J. S. Kalirai, N. J. Kallivayalil, B. Kalmbach, J. P. Kantor, P. Karst, M. M. Kasliwal, H. Kelly, R. Kessler, V. Kinnison, D. Kirkby, L. Knox, I. V. Kotov, V. L. Krabbendam, K. S. Krughoff, P. Kubánek, J. Kuczewski, S. Kulkarni, J. Ku, N. R. Kurita, C. S. Lage, R. Lambert, T. Lange, J. B. Langton, L. Le Guillou, D. Levine, M. Liang, K.-T. Lim, C. J. Lintott, K. E. Long, M. Lopez, P. J. Lotz, R. H. Lupton, N. B. Lust, L. A. MacArthur, A. Mahabal, R. Mandelbaum, T. W. Markiewicz, D. S. Marsh, P. J. Marshall, S. Marshall, M. May, R. McKercher, M. McQueen, J. Meyers, M. Migliore, M. Miller, D. J. Mills, C. Miraval, J. Moeyens, F. E. Moolekamp, D. G. Monet, M. Moniez, S. Monkewitz, C. Montgomery, C. B. Morrison, F. Mueller, G. P. Muller, F. Muñoz Arancibia, D. R. Neill, S. P. Newbry, J.-Y. Nief, A. Nomerotski, M. Nordby, P. O'Connor, J. Oliver, S. S. Olivier, K. Olsen, W. O'Mullane, S. Ortiz, S. Osier, R. E. Owen, R. Pain, P. E. Palecek, J. K. Parejko, J. B. Parsons, N. M. Pease, J. M. Peterson, J. R. Peterson, D. L. Petravick, M. E. Libby Petrick, C. E. Petry, F. Pierfederici, S. Pietrowicz, R. Pike, P. A. Pinto, R. Plante, S. Plate, J. P. Plutchak, P. A. Price, M. Prouza, V. Radeka, J. Rajagopal, A. P. Rasmussen, N. Regnault, K. A. Reil, D. J. Reiss, M. A. Reuter, S. T. Ridgway, V. J. Riot, S. Ritz, S. Robinson, W. Roby, A. Roodman, W. Rosing, C. Roucelle, M. R. Rumore, S. Russo, A. Saha, B. Sassolas, T. L. Schalk, P. Schellart, R. H. Schindler, S. Schmidt, D. P. Schneider, M. D. Schneider, W. Schoening, G. Schumacher, M. E. Schwamb, J. Sebag, B. Selvy, G. H. Sembroski, L. G. Seppala, A. Serio, E. Serrano, R. A. Shaw, I. Shipsey, J. Sick, N. Silvestri, C. T. Slater, J. A. Smith, R. C. Smith, S. Sobhani, C. Soldahl, L. Storrie-Lombardi, E. Stover, M. A. Strauss, R. A. Street, C. W. Stubbs, I. S. Sullivan, D. Sweeney, J. D. Swinbank, A. Szalay, P. Takacs, S. A. Tether, J. J. Thaler, J. G. Thayer, S. Thomas, A. J. Thornton, V. Thukral, J. Tice, D. E. Trilling, M. Turri, R. Van Berg, D. Vanden Berk, K. Vetter, F. Virieux, T. Vucina, W. Wahl, L. Walkowicz, B. Walsh, C. W. Walter, D. L. Wang, S.-Y. Wang, M. Warner, O. Wiecha, B. Willman, S. E. Winters, D. Wittman, S. C. Wolff,

W. M. Wood-Vasey, X. Wu, B. Xin, P. Yoachim, and H. Zhan, ApJ **873**, 111 (2019), arXiv:0805.2366 [astro-ph] .

[200] R. Laureijs, J. Amiaux, S. Arduini, J. L. Augueres, J. Brinchmann, R. Cole, M. Cropper, C. Dabin, L. Duvet, A. Ealet, B. Garilli, P. Gondoin, L. Guzzo, J. Hoar, H. Hoekstra, R. Holmes, T. Kitching, T. Maciaszek, Y. Mellier, F. Pasian, W. Percival, J. Rhodes, G. Saavedra Criado, M. Sauvage, R. Scaramella, L. Valenziano, S. Warren, R. Bender, F. Castander, A. Cimatti, O. Le Fèvre, H. Kurki-Suonio, M. Levi, P. Lilje, G. Meylan, R. Nichol, K. Pedersen, V. Popa, R. Rebolo Lopez, H. W. Rix, H. Rottgering, W. Zeilinger, F. Grupp, P. Hudelot, R. Massey, M. Meneghetti, L. Miller, S. Paltani, S. Paulin-Henriksson, S. Pires, C. Saxton, T. Schrabback, G. Seidel, J. Walsh, N. Aghanim, L. Amendola, J. Bartlett, C. Baccigalupi, J. P. Beaulieu, K. Benabed, J. G. Cuby, D. Elbaz, P. Fosalba, G. Gavazzi, A. Helmi, I. Hook, M. Irwin, J. P. Kneib, M. Kunz, F. Mannucci, L. Moscardini, C. Tao, R. Teyssier, J. Weller, G. Zamorani, M. R. Zapatero Osorio, O. Boulade, J. J. Foumond, A. Di Giorgio, P. Guttridge, A. James, M. Kemp, J. Martignac, A. Spencer, D. Walton, T. Blümchen, C. Bonoli, F. Bortoletto, C. Cerna, L. Corcione, C. Fabron, K. Jahnke, S. Ligori, F. Madrid, L. Martin, G. Morgante, T. Pamplona, E. Prieto, M. Riva, R. Toledo, M. Trifoglio, F. Zerbi, F. Abdalla, M. Douspis, C. Grenet, S. Borgani, R. Bouwens, F. Courbin, J. M. Delouis, P. Dubath, A. Fontana, M. Frailis, A. Grazian, J. Koppenhöfer, O. Mansutti, M. Melchior, M. Mignoli, J. Mohr, C. Neissner, K. Noddle, M. Poncet, M. Scodeggio, S. Serrano, N. Shane, J. L. Starck, C. Surace, A. Taylor, G. Verdoes-Kleijn, C. Vuerli, O. R. Williams, A. Zacchei, B. Altieri, I. Escudero Sanz, R. Kohley, T. Oosterbroek, P. Astier, D. Bacon, S. Bardelli, C. Baugh, F. Bellagamba, C. Benoist, D. Bianchi, A. Biviano, E. Branchini, C. Carbone, V. Cardone, D. Clements, S. Colombi, C. Conselice, G. Cresci, N. Deacon, J. Dunlop, C. Fedeli, F. Fontanot, P. Franzetti, C. Giocoli, J. Garcia-Bellido, J. Gow, A. Heavens, P. Hewett, C. Heymans, A. Holland, Z. Huang, O. Ilbert, B. Joachimi, E. Jennins, E. Kerins, A. Kiessling, D. Kirk, R. Kotak, O. Krause, O. Lahav, F. van Leeuwen, J. Lesgourgues, M. Lombardi, M. Magliocchetti, K. Maguire, E. Majerotto, R. Maoli, F. Marulli, S. Maurogordato, H. McCracken, R. McLure, A. Melchiorri, A. Merson, M. Moresco, M. Nonino, P. Norberg, J. Peacock, R. Pello, M. Penny, V. Pettorino, C. Di Porto, L. Pozzetti, C. Quercellini, M. Radovich, A. Rassat, N. Roche, S. Ronayette, E. Rossetti, B. Sartoris, P. Schneider, E. Semboloni, S. Serjeant, F. Simpson, C. Skordis,

G. Smadja, S. Smartt, P. Spano, S. Spiro, M. Sullivan, A. Tilquin, R. Trotta, L. Verde, Y. Wang, G. Williger, G. Zhao, J. Zoubian, and E. Zucca, arXiv e-prints , arXiv:1110.3193 (2011), arXiv:1110.3193 [astro-ph.CO] .

[201] D. Spergel, N. Gehrels, C. Baltay, D. Bennett, J. Breckinridge, M. Donahue, A. Dressler, B. S. Gaudi, T. Greene, O. Guyon, C. Hirata, J. Kalirai, N. J. Kasdin, B. Macintosh, W. Moos, S. Perlmutter, M. Postman, B. Rauscher, J. Rhodes, Y. Wang, D. Weinberg, D. Benford, M. Hudson, W. S. Jeong, Y. Mellier, W. Traub, T. Yamada, P. Capak, J. Colbert, D. Masters, M. Penny, D. Savransky, D. Stern, N. Zimmerman, R. Barry, L. Bartusek, K. Carpenter, E. Cheng, D. Content, F. Dekens, R. Demers, K. Grady, C. Jackson, G. Kuan, J. Kruk, M. Melton, B. Nemati, B. Parvin, I. Poberezhskiy, C. Peddie, J. Ruffa, J. K. Wallace, A. Whipple, E. Wollack, and F. Zhao, arXiv e-prints , arXiv:1503.03757 (2015), arXiv:1503.03757 [astro-ph.IM] .

[202] P. Ade, J. Aguirre, Z. Ahmed, S. Aiola, A. Ali, D. Alonso, M. A. Alvarez, K. Arnold, P. Ashton, J. Austermann, H. Awan, C. Baccigalupi, T. Baildon, D. Barron, N. Battaglia, R. Battye, E. Baxter, A. Bazarko, J. A. Beall, R. Bean, D. Beck, S. Beckman, B. Beringue, F. Bianchini, S. Boada, D. Boettger, J. R. Bond, J. Borrill, M. L. Brown, S. M. Bruno, S. Bryan, E. Calabrese, V. Calafut, P. Calisse, J. Carron, A. Challinor, G. Chesmore, Y. Chinone, J. Chluba, H.-M. S. Cho, S. Choi, G. Coppi, N. F. Cothard, K. Coughlin, D. Crichton, K. D. Crowley, K. T. Crowley, A. Cukierman, J. M. D'Ewart, R. Dünner, T. de Haan, M. Devlin, S. Dicker, J. Didier, M. Dobbs, B. Dober, C. J. Duell, S. Duff, A. Duivenvoorden, J. Dunkley, J. Dusatko, J. Errard, G. Fabbian, S. Feeney, S. Ferraro, P. Fluxà, K. Freese, J. C. Frisch, A. Frolov, G. Fuller, B. Fuzia, N. Galitzki, P. A. Gallardo, J. Tomas Galvez Ghersi, J. Gao, E. Gawiser, M. Gerbino, V. Gluscevic, N. Goeckner-Wald, J. Golec, S. Gordon, M. Gralla, D. Green, A. Grigorian, J. Groh, C. Groppi, Y. Guan, J. E. Gudmundsson, D. Han, P. Hargrave, M. Hasegawa, M. Hasselfield, M. Hattori, V. Haynes, M. Hazumi, Y. He, E. Healy, S. W. Henderson, C. Hervias-Caimapo, C. A. Hill, J. C. Hill, G. Hilton, M. Hilton, A. D. Hincks, G. Hinshaw, R. Hložek, S. Ho, S.-P. P. Ho, L. Howe, Z. Huang, J. Hubmayr, K. Huffenberger, J. P. Hughes, A. Ijjas, M. Ikape, K. Irwin, A. H. Jaffe, B. Jain, O. Jeong, D. Kaneko, E. D. Karpel, N. Katayama, B. Keating, S. S. Kernasovskiy, R. Keskitalo, T. Kisner, K. Kiuchi, J. Klein, K. Knowles, B. Koopman, A. Kosowsky, N. Krachmalnicoff, S. E. Kuenstner, C.-L. Kuo, A. Kusaka, J. Lashner, A. Lee, E. Lee, D. Leon, J. S. Y. Leung, A. Lewis, Y. Li, Z. Li, M. Limon, E. Linder, C. Lopez-Caraballo,

T. Louis, L. Lowry, M. Lungu, M. Madhavacheril, D. Mak, F. Maldonado, H. Mani, B. Mates, F. Matsuda, L. Maurin, P. Mauskopf, A. May, N. McCallum, C. McKenney, J. McMahon, P. D. Meerburg, J. Meyers, A. Miller, M. Mirmelstein, K. Moodley, M. Munchmeyer, C. Munson, S. Naess, F. Nati, M. Navaroli, L. Newburgh, H. N. Nguyen, M. Niemack, H. Nishino, J. Orlowski-Scherer, L. Page, B. Partridge, J. Peloton, F. Perrotta, L. Piccirillo, G. Pisano, D. Poletti, R. Puddu, G. Puglisi, C. Raum, C. L. Reichardt, M. Remazeilles, Y. Rephaeli, D. Riechers, F. Rojas, A. Roy, S. Sadeh, Y. Sakurai, M. Salatino, M. Sathyanarayana Rao, E. Schaan, M. Schmittfull, N. Sehgal, J. Seibert, U. Seljak, B. Sherwin, M. Shimon, C. Sierra, J. Sievers, P. Sikhosana, M. Silva-Feaver, S. M. Simon, A. Sinclair, P. Siritanasak, K. Smith, S. R. Smith, D. Spergel, S. T. Staggs, G. Stein, J. R. Stevens, R. Stompor, A. Suzuki, O. Tajima, S. Takakura, G. Teply, D. B. Thomas, B. Thorne, R. Thornton, H. Trac, C. Tsai, C. Tucker, J. Ullom, S. Vagnozzi, A. van Engelen, J. Van Lanen, D. D. Van Winkle, E. M. Vavagiakis, C. Vergès, M. Vissers, K. Wagoner, S. Walker, J. Ward, B. Westbrook, N. Whitehorn, J. Williams, J. Williams, E. J. Wollack, Z. Xu, B. Yu, C. Yu, F. Zago, H. Zhang, N. Zhu, and Simons Observatory Collaboration, JCAP **2019**, 056 (2019), arXiv:1808.07445 [astro-ph.CO] .

[203] N. Sehgal, S. Aiola, Y. Akrami, K. Basu, M. Boylan-Kolchin, S. Bryan, S. Clesse, F.-Y. Cyr-Racine, L. Di Mascolo, S. Dicker, T. Essinger-Hileman, S. Ferraro, G. Fuller, D. Han, M. Hasselfield, G. Holder, B. Jain, B. R. Johnson, M. Johnson, P. Klaassen, M. Madhavacheril, P. Mauskopf, D. Meerburg, J. Meyers, T. Mroczkowski, M. Münchmeyer, S. K. Naess, D. Nagai, T. Namikawa, L. Newburgh, N. Nguyen, M. Niemack, B. D. Oppenheimer, E. Pierpaoli, E. Schaan, A. Slosar, D. Spergel, E. Switzer, A. van Engelen, and E. Wollack, in *Bulletin of the American Astronomical Society*, Vol. 51 (2019) p. 6, arXiv:1906.10134 [astro-ph.CO] .

[204] K. N. Abazajian, P. Adshead, Z. Ahmed, S. W. Allen, D. Alonso, K. S. Arnold, C. Baccigalupi, J. G. Bartlett, N. Battaglia, B. A. Benson, C. A. Bischoff, J. Borrill, V. Buza, E. Calabrese, R. Caldwell, J. E. Carlstrom, C. L. Chang, T. M. Crawford, F.-Y. Cyr-Racine, F. De Bernardis, T. de Haan, S. di Serego Alighieri, J. Dunkley, C. Dvorkin, J. Errard, G. Fabbian, S. Feeney, S. Ferraro, J. P. Filippini, R. Flauger, G. M. Fuller, V. Gluscevic, D. Green, D. Grin, E. Grohs, J. W. Henning, J. C. Hill, R. Hlozek, G. Holder, W. Holzapfel, W. Hu, K. M. Huffenberger, R. Keskitalo, L. Knox, A. Kosowsky, J. Kovac, E. D.

Kovetz, C.-L. Kuo, A. Kusaka, M. Le Jeune, A. T. Lee, M. Lilley, M. Loverde, M. S. Madhavacheril, A. Mantz, D. J. E. Marsh, J. McMahon, P. D. Meerburg, J. Meyers, A. D. Miller, J. B. Munoz, H. N. Nguyen, M. D. Niemack, M. Peloso, J. Peloton, L. Pogosian, C. Pryke, M. Raveri, C. L. Reichardt, G. Rocha, A. Rotti, E. Schaan, M. M. Schmittfull, D. Scott, N. Sehgal, S. Shandera, B. D. Sherwin, T. L. Smith, L. Sorbo, G. D. Starkman, K. T. Story, A. van Engelen, J. D. Vieira, S. Watson, N. Whitehorn, and W. L. Kimmy Wu, arXiv e-prints , arXiv:1610.02743 (2016), arXiv:1610.02743 [astro-ph.CO] .

[205] N. E. Chisari, D. Alonso, E. Krause, C. D. Leonard, P. Bull, J. Neveu, A. Villarreal, S. Singh, T. McClintock, J. Ellison, Z. Du, J. Zuntz, A. Mead, S. Joudaki, C. S. Lorenz, T. Tröster, J. Sanchez, F. Lanusse, M. Ishak, R. Hlozek, J. Blazek, J.-E. Campagne, H. Almoubayyed, T. Eifler, M. Kirby, D. Kirkby, S. Plaszczynski, A. Slosar, M. Vrastil, E. L. Wagoner, and LSST Dark Energy Science Collaboration, ApJS **242**, 2 (2019), arXiv:1812.05995 [astro-ph.CO] .

[206] C. D. Leonard, T. Ferreira, X. Fang, R. Reischke, N. Schoeneberg, T. Tröster, D. Alonso, J.-E. Campagne, F. Lanusse, A. Slosar, and M. Ishak, The Open Journal of Astrophysics **6**, 8 (2023), arXiv:2212.04291 [astro-ph.CO] .

[207] J.-E. Campagne, F. Lanusse, J. Zuntz, A. Boucaud, S. Casas, M. Karamanis, D. Kirkby, D. Lanzieri, A. Peel, and Y. Li, The Open Journal of Astrophysics **6**, 15 (2023), arXiv:2302.05163 [astro-ph.CO] .

[208] D. Piras and A. Spurio Mancini, arXiv e-prints , arXiv:2305.06347 (2023), arXiv:2305.06347 [astro-ph.CO] .

[209] A. Spurio Mancini, D. Piras, J. Alsing, B. Joachimi, and M. P. Hobson, MNRAS **511**, 1771 (2022), arXiv:2106.03846 [astro-ph.CO] .

[210] Z. Li and J. Sullivan, In prep, (2023).

[211] P. H. W. Hoffmann, arXiv e-prints , arXiv:1411.0583 (2014), arXiv:1411.0583 [math.NA] .

[212] E. R. Harrison, Phys. Rev. D **1**, 2726 (1970).

[213] Y. B. Zeldovich, MNRAS **160**, 1P (1972).

[214] J. Donald-McCann, F. Beutler, K. Koyama, and M. Karamanis, Mon. Not. Roy. Astron. Soc. **511**, 3768 (2022), arXiv:2109.15236 [astro-ph.CO] .

[215] M. Bonici, F. Bianchini, and J. Ruiz-Zapatero, arXiv e-prints , arXiv:2307.14339 (2023), arXiv:2307.14339 [astro-ph.CO] .

[216] A. Mootoovaloo, A. H. Jaffe, A. F. Heavens, and F. Leclercq, Astronomy and Computing **38**, 100508 (2022), arXiv:2105.02256 [astro-ph.CO] .

[217] C. C. Margossian and M. Betancourt, arXiv e-prints , arXiv:2112.14217 (2021), arXiv:2112.14217 [stat.CO] .

[218] A. J. Mead, S. Brieden, T. Tröster, and C. Heymans, MNRAS **502**, 1401 (2021), arXiv:2009.01858 [astro-ph.CO] .

[219] R. E. Angulo, M. Zennaro, S. Contreras, G. Aricò, M. Pellejero-Ibañez, and J. Stücker, MNRAS **507**, 5869 (2021), arXiv:2004.06245 [astro-ph.CO] .

[220] The Dark Energy Survey Collaboration, arXiv e-prints , astro-ph/0510346 (2005), arXiv:astro-ph/0510346 [astro-ph] .

[221] B. Hoyle, D. Gruen, G. M. Bernstein, M. M. Rau, J. De Vicente, W. G. Hartley, E. Gaztanaga, J. DeRose, M. A. Troxel, C. Davis, A. Alarcon, N. MacCrann, J. Prat, C. Sánchez, E. Sheldon, R. H. Wechsler, J. Asorey, M. R. Becker, C. Bonnett, A. Carnero Rosell, D. Carollo, M. Carrasco Kind, F. J. Castander, R. Cawthon, C. Chang, M. Childress, T. M. Davis, A. Drlica-Wagner, M. Gatti, K. Glazebrook, J. Gschwend, S. R. Hinton, J. K. Hoormann, A. G. Kim, A. King, K. Kuehn, G. Lewis, C. Lidman, H. Lin, E. Macaulay, M. A. G. Maia, P. Martini, D. Mudd, A. Möller, R. C. Nichol, R. L. C. Ogando, R. P. Rollins, A. Roodman, A. J. Ross, E. Rozo, E. S. Rykoff, S. Samuroff, I. Sevilla-Noarbe, R. Sharp, N. E. Sommer, B. E. Tucker, S. A. Uddin, T. N. Varga, P. Vielzeuf, F. Yuan, B. Zhang, T. M. C. Abbott, F. B. Abdalla, S. Allam, J. Annis, K. Bechtol, A. Benoit-Lévy, E. Bertin, D. Brooks, E. Buckley-Geer, D. L. Burke, M. T. Busha, D. Capozzi, J. Carretero, M. Crocce, C. B. D'Andrea, L. N. da Costa, D. L. DePoy, S. Desai, H. T. Diehl, P. Doel, T. F. Eifler, J. Estrada, A. E. Evrard, E. Fernandez, B. Flaugher, P. Fosalba, J. Frieman, J. García-Bellido, D. W. Gerdes, T. Giannantonio, D. A. Goldstein, R. A. Gruendl, G. Gutierrez, K. Honscheid, D. J. James, M. Jarvis, T. Jeltema, M. W. G. Johnson, M. D. Johnson, D. Kirk, E. Krause, S. Kuhlmann, N. Kuropatkin, O. Lahav, T. S. Li, M. Lima, M. March, J. L. Marshall, P. Melchior, F. Menanteau, R. Miquel, B. Nord, C. R. O'Neill, A. A. Plazas, A. K. Romer, M. Sako, E. Sanchez, B. Santiago, V. Scarpine, R. Schindler, M. Schubnell, M. Smith, R. C. Smith, M. Soares-Santos, F. Sobreira, E. Suchyta, M. E. C. Swanson, G. Tarle, D. Thomas, D. L. Tucker, V. Vikram, A. R. Walker, J. Weller, W. Wester, R. C. Wolf, B. Yanny, J. Zuntz, and DES Collaboration, MNRAS **478**, 592 (2018), arXiv:1708.01532 [astro-ph.CO] .

[222] H. Ge, K. Xu, and Z. Ghahramani, in *International Conference on Artificial Intelligence and Statistics, AISTATS 2018, 9-11 April 2018, Playa Blanca, Lanzarote, Canary Islands, Spain* (2018) pp. 1682–1690.

[223] A. Gelman, J. Carlin, H. Stern, D. Dunson, A. Vehtari, and D. Rubin, *Bayesian Data Analysis, Third Edition*, Chapman & Hall/CRC Texts in Statistical Science (Taylor & Francis, 2013).

[224] M. J. Betancourt, arXiv e-prints , arXiv:1212.4693 (2012), arXiv:1212.4693 [stat.ME] .

[225] G. Lavaux, J. Jasche, and F. Leclercq, arXiv e-prints , arXiv:1909.06396 (2019), arXiv:1909.06396 [astro-ph.CO] .

[226] A. E. Bayer, A. Banerjee, and U. Seljak, Phys. Rev. D **105**, 123510 (2022), arXiv:2108.04215 [astro-ph.CO] .

[227] B. Dai and U. Seljak, arXiv e-prints , arXiv:2306.04689 (2023), arXiv:2306.04689 [astro-ph.CO] .

[228] F. Lepori, S. Schulz, J. Adamek, and R. Durrer, JCAP **2023**, 036 (2023), arXiv:2209.10533 [astro-ph.CO] .

[229] S. L. Bridle, R. Crittenden, A. Melchiorri, M. P. Hobson, R. Kneissl, and A. N. Lasenby, MNRAS **335**, 1193 (2002), arXiv:astro-ph/0112114 [astro-ph] .

[230] B. Stölzner, B. Joachimi, A. Korn, H. Hildebrandt, and A. H. Wright, A&A **650**, A148 (2021), arXiv:2012.07707 [astro-ph.CO] .

[231] H. Hildebrandt, F. Köhlinger, J. L. van den Busch, B. Joachimi, C. Heymans, A. Kannawadi, A. H. Wright, M. Asgari, C. Blake, H. Hoekstra, S. Joudaki, K. Kuijken, L. Miller, C. B. Morrison, T. Tröster, A. Amon, M. Archidiacono, S. Brieden, A. Choi, J. T. A. de Jong, T. Erben, B. Giblin, A. Mead, J. A. Peacock, M. Radovich, P. Schneider, C. Sifón, and M. Tewes, A&A **633**, A69 (2020), arXiv:1812.06076 [astro-ph.CO] .

[232] T. Zhang, M. M. Rau, R. Mandelbaum, X. Li, and B. Moews, MNRAS **518**, 709 (2023), arXiv:2206.10169 [astro-ph.CO] .

[233] R. Reischke, arXiv e-prints , arXiv:2301.04085 (2023), arXiv:2301.04085 [astro-ph.CO] .

[234] J. P. Cordero, I. Harrison, R. P. Rollins, G. M. Bernstein, S. L. Bridle, A. Alarcon, O. Alves, A. Amon, F. Andrade-Oliveira, H. Camacho, A. Campos, A. Choi, J. DeRose, S. Dodelson, K. Eckert, T. F. Eifler, S. Everett, X. Fang, O. Friedrich, D. Gruen, R. A. Gruendl, W. G. Hartley, E. M. Huff, E. Krause, N. Kuropatkin, N. MacCrann, J. McCullough, J. Myles, S. Pandey, M. Raveri, R. Rosenfeld, E. S. Rykoff, C. Sánchez, J. Sánchez, I. Sevilla-Noarbe, E. Sheldon, M. Troxel, R. Wechsler, B. Yanny, B. Yin, Y. Zhang, M. Aguena, S. Allam, E. Bertin, D. Brooks, D. L. Burke, A. Carnero Rosell, M. Carrasco Kind, J. Carretero, F. J. Castander, R. Cawthon, M. Costanzi, L. da

Costa, M. E. da Silva Pereira, J. De Vicente, H. T. Diehl, J. Dietrich, P. Doel, J. Elvin-Poole, I. Ferrero, B. Flaugher, P. Fosalba, J. Frieman, J. Garcia-Bellido, D. Gerdes, J. Gschwend, G. Gutierrez, S. Hinton, D. L. Hollowood, K. Honscheid, B. Hoyle, D. James, K. Kuehn, O. Lahav, M. A. G. Maia, M. March, F. Menanteau, R. Miquel, R. Morgan, J. Muir, A. Palmese, F. Paz-Chinchon, A. Pieres, A. Plazas Malagón, E. Sánchez, V. Scarpine, S. Serrano, M. Smith, M. Soares-Santos, E. Suchyta, M. Swanson, G. Tarle, D. Thomas, C. To, T. N. Varga, and DES Collaboration, MNRAS **511**, 2170 (2022), arXiv:2109.09636 [astro-ph.CO] .

[235] I. Sevilla-Noarbe, K. Bechtol, M. Carrasco Kind, A. Carnero Rosell, M. R. Becker, A. Drlica-Wagner, R. A. Gruendl, E. S. Rykoff, E. Sheldon, B. Yanny, A. Alarcon, S. Allam, A. Amon, A. Benoit-Lévy, G. M. Bernstein, E. Bertin, D. L. Burke, J. Carretero, A. Choi, H. T. Diehl, S. Everett, B. Flaugher, E. Gaztanaga, J. Gschwend, I. Harrison, W. G. Hartley, B. Hoyle, M. Jarvis, M. D. Johnson, R. Kessler, R. Kron, N. Kuropatkin, B. Leistedt, T. S. Li, F. Menanteau, E. Morganson, R. L. C. Ogando, A. Palmese, F. Paz-Chinchón, A. Pieres, C. Pond, M. Rodriguez-Monroy, J. A. Smith, K. M. Stringer, M. A. Troxel, D. L. Tucker, J. de Vicente, W. Wester, Y. Zhang, T. M. C. Abbott, M. Aguena, J. Annis, S. Avila, S. Bhargava, S. L. Bridle, D. Brooks, D. Brout, F. J. Castander, R. Cawthon, C. Chang, C. Conselice, M. Costanzi, M. Crocce, L. N. da Costa, M. E. S. Pereira, T. M. Davis, S. Desai, J. P. Dietrich, P. Doel, K. Eckert, A. E. Evrard, I. Ferrero, P. Fosalba, J. García-Bellido, D. W. Gerdes, T. Giannantonio, D. Gruen, G. Gutierrez, S. R. Hinton, D. L. Hollowood, K. Honscheid, E. M. Huff, D. Huterer, D. J. James, T. Jeltema, K. Kuehn, O. Lahav, C. Lidman, M. Lima, H. Lin, M. A. G. Maia, J. L. Marshall, P. Martini, P. Melchior, R. Miquel, J. J. Mohr, R. Morgan, E. Neilsen, A. A. Plazas, A. K. Romer, A. Roodman, E. Sanchez, V. Scarpine, M. Schubnell, S. Serrano, M. Smith, E. Suchyta, G. Tarle, D. Thomas, C. To, T. N. Varga, R. H. Wechsler, J. Weller, R. D. Wilkinson, and DES Collaboration, ApJS **254**, 24 (2021), arXiv:2011.03407 [astro-ph.CO] .

[236] C. Sánchez and G. M. Bernstein, MNRAS **483**, 2801 (2019), arXiv:1807.11873 [astro-ph.CO] .

[237] D. Kodwani, D. Alonso, and P. Ferreira, The Open Journal of Astrophysics **2**, 3 (2019), arXiv:1811.11584 [astro-ph.CO] .

[238] B. Hadzhiyska, D. Alonso, A. Nicola, and A. Slosar, JCAP **2020**, 056 (2020), arXiv:2007.14989 [astro-ph.CO] .

[239] K. Kuijken, C. Heymans, H. Hildebrandt, R. Nakajima, T. Erben, J. T. A. de Jong, M. Viola, A. Choi, H. Hoekstra, L. Miller, E. van Uitert, A. Amon, C. Blake, M. Brouwer, A. Buddendiek, I. F. Conti, M. Eriksen, A. Grado, J. Harnois-Déraps, E. Helmich, R. Herbonnet, N. Irisarri, T. Kitching, D. Klaes, F. La Barbera, N. Napolitano, M. Radovich, P. Schneider, C. Sifón, G. Sikkema, P. Simon, A. Tudorica, E. Valentijn, G. Verdoes Kleijn, and L. van Waerbeke, Monthly Notices of the Royal Astronomical Society **454**, 3500 (2015), https://academic.oup.com/mnras/article-pdf/454/4/3500/18507938/stv2140.pdf .

[240] T. Hamana, M. Shirasaki, S. Miyazaki, C. Hikage, M. Oguri, S. More, R. Armstrong, A. Leauthaud, R. Mandelbaum, H. Miyatake, A. J. Nishizawa, M. Simet, M. Takada, H. Aihara, J. Bosch, Y. Komiyama, R. Lupton, H. Murayama, M. A. Strauss, and M. Tanaka, Publications of the Astronomical Society of Japan **72** (2020), 10.1093/pasj/psz138, 16, https://academic.oup.com/pasj/article-pdf/72/1/16/32408879/psz138.pdf .

[241] L. F. Secco, S. Samuroff, E. Krause, B. Jain, J. Blazek, M. Raveri, A. Campos, A. Amon, A. Chen, C. Doux, A. Choi, D. Gruen, G. M. Bernstein, C. Chang, J. DeRose, J. Myles, A. Ferté, P. Lemos, D. Huterer, J. Prat, M. A. Troxel, N. MacCrann, A. R. Liddle, T. Kacprzak, X. Fang, C. Sánchez, S. Pandey, S. Dodelson, P. Chintalapati, K. Hoffmann, A. Alarcon, O. Alves, F. Andrade-Oliveira, E. J. Baxter, K. Bechtol, M. R. Becker, A. Brandao-Souza, H. Camacho, A. Carnero Rosell, M. Carrasco Kind, R. Cawthon, J. P. Cordero, M. Crocce, C. Davis, E. Di Valentino, A. Drlica-Wagner, K. Eckert, T. F. Eifler, M. Elidaiana, F. Elsner, J. Elvin-Poole, S. Everett, P. Fosalba, O. Friedrich, M. Gatti, G. Giannini, R. A. Gruendl, I. Harrison, W. G. Hartley, K. Herner, H. Huang, E. M. Huff, M. Jarvis, N. Jeffrey, N. Kuropatkin, P. F. Leget, J. Muir, J. Mccullough, A. Navarro Alsina, Y. Omori, Y. Park, A. Porredon, R. Rollins, A. Roodman, R. Rosenfeld, A. J. Ross, E. S. Rykoff, J. Sanchez, I. Sevilla-Noarbe, E. S. Sheldon, T. Shin, A. Troja, I. Tutusaus, T. N. Varga, N. Weaverdyck, R. H. Wechsler, B. Yanny, B. Yin, Y. Zhang, J. Zuntz, T. M. C. Abbott, M. Aguena, S. Allam, J. Annis, D. Bacon, E. Bertin, S. Bhargava, S. L. Bridle, D. Brooks, E. Buckley-Geer, D. L. Burke, J. Carretero, M. Costanzi, L. N. da Costa, J. De Vicente, H. T. Diehl, J. P. Dietrich, P. Doel, I. Ferrero, B. Flaugher, J. Frieman, J. García-Bellido, E. Gaztanaga, D. W. Gerdes, T. Giannantonio, J. Gschwend, G. Gutierrez, S. R. Hinton, D. L. Hollowood, K. Honscheid, B. Hoyle, D. J. James, T. Jeltema, K. Kuehn,

O. Lahav, M. Lima, H. Lin, M. A. G. Maia, J. L. Marshall, P. Martini, P. Melchior, F. Menanteau, R. Miquel, J. J. Mohr, R. Morgan, R. L. C. Ogando, A. Palmese, F. Paz-Chinchón, D. Petravick, A. Pieres, A. A. Plazas Malagón, M. Rodriguez-Monroy, A. K. Romer, E. Sanchez, V. Scarpine, M. Schubnell, D. Scolnic, S. Serrano, M. Smith, M. Soares-Santos, E. Suchyta, M. E. C. Swanson, G. Tarle, D. Thomas, C. To, and DES Collaboration, Phys. Rev. D **105**, 023515 (2022), arXiv:2105.13544 [astro-ph.CO] .

[242] Z. Ma, W. Hu, and D. Huterer, ApJ **636**, 21 (2006), arXiv:astro-ph/0506614 [astro-ph] .

[243] The LSST Dark Energy Science Collaboration, R. Mandelbaum, T. Eifler, R. Hložek, T. Collett, E. Gawiser, D. Scolnic, D. Alonso, H. Awan, R. Biswas, J. Blazek, P. Burchat, N. E. Chisari, I. Dell'Antonio, S. Digel, J. Frieman, D. A. Goldstein, I. Hook, Ž. Ivezić, S. M. Kahn, S. Kamath, D. Kirkby, T. Kitching, E. Krause, P.-F. Leget, P. J. Marshall, J. Meyers, H. Miyatake, J. A. Newman, R. Nichol, E. Rykoff, F. J. Sanchez, A. Slosar, M. Sullivan, and M. A. Troxel, arXiv e-prints , arXiv:1809.01669 (2018), arXiv:1809.01669 [astro-ph.CO] .

[244] C. Bonnett, M. A. Troxel, W. Hartley, A. Amara, B. Leistedt, M. R. Becker, G. M. Bernstein, S. L. Bridle, C. Bruderer, M. T. Busha, M. Carrasco Kind, M. J. Childress, F. J. Castander, C. Chang, M. Crocce, T. M. Davis, T. F. Eifler, J. Frieman, C. Gangkofner, E. Gaztanaga, K. Glazebrook, D. Gruen, T. Kacprzak, A. King, J. Kwan, O. Lahav, G. Lewis, C. Lidman, H. Lin, N. MacCrann, R. Miquel, C. R. O'Neill, A. Palmese, H. V. Peiris, A. Refregier, E. Rozo, E. S. Rykoff, I. Sadeh, C. Sánchez, E. Sheldon, S. Uddin, R. H. Wechsler, J. Zuntz, T. Abbott, F. B. Abdalla, S. Allam, R. Armstrong, M. Banerji, A. H. Bauer, A. Benoit-Lévy, E. Bertin, D. Brooks, E. Buckley-Geer, D. L. Burke, D. Capozzi, A. Carnero Rosell, J. Carretero, C. E. Cunha, C. B. D'Andrea, L. N. da Costa, D. L. DePoy, S. Desai, H. T. Diehl, J. P. Dietrich, P. Doel, A. Fausti Neto, E. Fernandez, B. Flaugher, P. Fosalba, D. W. Gerdes, R. A. Gruendl, K. Honscheid, B. Jain, D. J. James, M. Jarvis, A. G. Kim, K. Kuehn, N. Kuropatkin, T. S. Li, M. Lima, M. A. G. Maia, M. March, J. L. Marshall, P. Martini, P. Melchior, C. J. Miller, E. Neilsen, R. C. Nichol, B. Nord, R. Ogando, A. A. Plazas, K. Reil, A. K. Romer, A. Roodman, M. Sako, E. Sanchez, B. Santiago, R. C. Smith, M. Soares-Santos, F. Sobreira, E. Suchyta, M. E. C. Swanson, G. Tarle, J. Thaler, D. Thomas, V. Vikram, A. R. Walker, and Dark Energy Survey Collaboration, Phys. Rev. D **94**, 042005 (2016), arXiv:1507.05909 [astro-ph.CO] .

[245] A. Nicola, C. García-García, D. Alonso, J. Dunkley, P. G. Ferreira, A. Slosar, and D. N. Spergel, JCAP **2021**, 067 (2021), arXiv:2010.09717 [astro-ph.CO] .

[246] C. García-García, D. Alonso, P. G. Ferreira, B. Hadzhiyska, A. Nicola, C. Sánchez, and A. Slosar, JCAP **2023**, 025 (2023), arXiv:2210.13434 [astro-ph.CO] .

[247] C. Laigle, H. J. McCracken, O. Ilbert, B. C. Hsieh, I. Davidzon, P. Capak, G. Hasinger, J. D. Silverman, C. Pichon, J. Coupon, H. Aussel, D. Le Borgne, K. Caputi, P. Cassata, Y. Y. Chang, F. Civano, J. Dunlop, J. Fynbo, J. S. Kartaltepe, A. Koekemoer, O. Le Fèvre, E. Le Floc'h, A. Leauthaud, S. Lilly, L. Lin, S. Marchesi, B. Milvang-Jensen, M. Salvato, D. B. Sanders, N. Scoville, V. Smolcic, M. Stockmann, Y. Taniguchi, L. Tasca, S. Toft, M. Vaccari, and J. Zabl, ApJS **224**, 24 (2016), arXiv:1604.02350 [astro-ph.GA] .

[248] A. Nicola *et al.*, In prep, (2023).

[249] D. Alonso, P. Bull, P. G. Ferreira, R. Maartens, and M. G. Santos, ApJ **814**, 145 (2015), arXiv:1505.07596 [astro-ph.CO] .

[250] J. Torrado and A. Lewis, "Cobaya: Bayesian analysis in cosmology," Astrophysics Source Code Library, record ascl:1910.019 (2019), ascl:1910.019 .

[251] J. Torrado and A. Lewis, JCAP **2021**, 057 (2021), arXiv:2005.05290 [astro-ph.IM] .

[252] D. Blas, J. Lesgourgues, and T. Tram, JCAP **2011**, 034 (2011), arXiv:1104.2933 [astro-ph.CO] .

[253] J. Elvin-Poole, M. Crocce, A. J. Ross, T. Giannantonio, E. Rozo, E. S. Rykoff, S. Avila, N. Banik, J. Blazek, S. L. Bridle, R. Cawthon, A. Drlica-Wagner, O. Friedrich, N. Kokron, E. Krause, N. MacCrann, J. Prat, C. Sánchez, L. F. Secco, I. Sevilla-Noarbe, M. A. Troxel, T. M. C. Abbott, F. B. Abdalla, S. Allam, J. Annis, J. Asorey, K. Bechtol, M. R. Becker, A. Benoit-Lévy, G. M. Bernstein, E. Bertin, D. Brooks, E. Buckley-Geer, D. L. Burke, A. Carnero Rosell, D. Carollo, M. Carrasco Kind, J. Carretero, F. J. Castander, C. E. Cunha, C. B. D'Andrea, L. N. da Costa, T. M. Davis, C. Davis, S. Desai, H. T. Diehl, J. P. Dietrich, S. Dodelson, P. Doel, T. F. Eifler, A. E. Evrard, E. Fernandez, B. Flaugher, P. Fosalba, J. Frieman, J. García-Bellido, E. Gaztanaga, D. W. Gerdes, K. Glazebrook, D. Gruen, R. A. Gruendl, J. Gschwend, G. Gutierrez, W. G. Hartley, S. R. Hinton, K. Honscheid, J. K. Hoormann, B. Jain, D. J. James, M. Jarvis, T. Jeltema, M. W. G. Johnson, M. D. Johnson, A. King, K. Kuehn, S. Kuhlmann, N. Kuropatkin, O. Lahav, G. Lewis, T. S. Li, C. Lidman, M. Lima, H. Lin, E. Macaulay, M. March, J. L. Marshall,

P. Martini, P. Melchior, F. Menanteau, R. Miquel, J. J. Mohr, A. Möller, R. C. Nichol, B. Nord, C. R. O'Neill, W. J. Percival, D. Petravick, A. A. Plazas, A. K. Romer, M. Sako, E. Sanchez, V. Scarpine, R. Schindler, M. Schubnell, E. Sheldon, M. Smith, R. C. Smith, M. Soares-Santos, F. Sobreira, N. E. Sommer, E. Suchyta, M. E. C. Swanson, G. Tarle, D. Thomas, B. E. Tucker, D. L. Tucker, S. A. Uddin, V. Vikram, A. R. Walker, R. H. Wechsler, J. Weller, W. Wester, R. C. Wolf, F. Yuan, B. Zhang, J. Zuntz, and DES Collaboration, Phys. Rev. D **98**, 042006 (2018), arXiv:1708.01536 [astro-ph.CO] .

[254] A. Lewis, Phys. Rev. D **87**, 103529 (2013), arXiv:1304.4473 [astro-ph.CO] .

[255] R. M. Neal, arXiv Mathematics e-prints , math/0502099 (2005), arXiv:math/0502099 [math.ST] .

[256] M. M. Rau, R. Dalal, T. Zhang, X. Li, A. J. Nishizawa, S. More, R. Mandelbaum, H. Miyatake, M. A. Strauss, and M. Takada, MNRAS **524**, 5109 (2023), arXiv:2211.16516 [astro-ph.CO] .

[257] W. K. Hastings, Biometrika **57**, 97 (1970).

[258] A. Nicola, D. Alonso, J. Sánchez, A. Slosar, H. Awan, A. Broussard, J. Dunkley, E. Gawiser, Z. Gomes, R. Mandelbaum, H. Miyatake, J. A. Newman, I. Sevilla-Noarbe, S. Skinner, and E. L. Wagoner, JCAP **2020**, 044 (2020), arXiv:1912.08209 [astro-ph.CO] .