



HAL
open science

candl: cosmic microwave background analysis with a differentiable likelihood

L. Balkenhol, C. Trendafilova, K. Benabed, S. Galli

► **To cite this version:**

L. Balkenhol, C. Trendafilova, K. Benabed, S. Galli. candl: cosmic microwave background analysis with a differentiable likelihood. *Astronomy and Astrophysics - A&A*, 2024, 686, pp.A10. 10.1051/0004-6361/202449432 . hal-04587836

HAL Id: hal-04587836

<https://hal.science/hal-04587836v1>

Submitted on 24 May 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

candl: cosmic microwave background analysis with a differentiable likelihood[★]

L. Balkenhol¹ , C. Trendafilova², K. Benabed¹, and S. Galli¹

¹ Sorbonne Université, CNRS, UMR 7095, Institut d'Astrophysique de Paris, 98 bis bd Arago, 75014 Paris, France
e-mail: lennart.balkenhol@iap.fr

² Center for AstroPhysical Surveys, National Center for Supercomputing Applications, University of Illinois Urbana-Champaign, Urbana, IL, 61801, USA

Received 31 January 2024 / Accepted 29 February 2024

ABSTRACT

We present `candl`, an automatically differentiable python likelihood for analysing cosmic microwave background power spectrum measurements. `candl` is powered by JAX, which makes it fast and easy to calculate derivatives of the likelihood. This facilitates, for example, robust Fisher matrices without finite-difference methods. We show the benefits of `candl` through a series of example calculations, covering forecasting, robustness tests, and gradient-based Markov chain Monte Carlo sampling. These also include optimising the band power bin width to minimise parameter errors of a realistic mock data set. Moreover, we calculate the correlation of parameter constraints from correlated and partially overlapping subsets of the SPT-3G 2018 TT/TE/EE data release. In a traditional analysis framework, these tasks are slow and require careful fine-tuning to obtain stable results. As such, a fully differentiable pipeline allows for a higher level of scrutiny; we argue that this is the paradigm shift required to leverage incoming data from ground-based experiments, which will significantly improve the cosmological parameter constraints from the *Planck* mission. `candl` comes with the latest primary and lensing power spectrum data from the South Pole Telescope and Atacama Cosmology Telescope collaborations and will be used as part of the upcoming SPT-3G TT/TE/EE and $\phi\phi$ data releases. Along with the core code, we release a series of auxiliary tools, which simplify common analysis tasks and interface the likelihood with other cosmological software. `candl` is pip-installable and publicly available on GitHub.

Key words. methods: data analysis – methods: statistical – cosmic background radiation

1. Introduction

Cosmic microwave background (CMB) observations are a cornerstone of modern cosmology. Measurements of the temperature and polarisation power spectrum of the CMB allow us to determine the free parameters of the Λ cold-dark matter (Λ CDM) model with high precision and tightly limit the space of possible extensions (Planck Collaboration VI 2020; Aiola et al. 2020; Balkenhol et al. 2023). While the *Planck* satellite has mapped the foreground-subdominant scales of the temperature power spectrum to the sample variance limit (Planck Collaboration V 2020), contemporary ground-based experiments, such as the South Pole Telescope (SPT; Carlstrom et al. 2011), the Atacama Cosmology Telescope (ACT; Kosowsky 2003), and the Simons Observatory (SO; Simons Observatory Collaboration 2019) complement the satellite data and map the CMB's polarisation anisotropies on intermediate and small angular scales with unseen precision. Data collected by each of these telescopes is expected to come close to or exceed the parameter constraints of the *Planck* data set (Simons Observatory Collaboration 2019) (based on the lensing data presented in Madhavacheril et al. 2024; Qu et al. 2024; Prabhu et al., in prep.).

This high-precision data demands a high level of scrutiny. Any potential claim of a detection of physics beyond the Standard Model can only be embraced with confidence by the

scientific community, if the analysis is demonstrably robust. The growing complexity of CMB analyses obliges a plethora of tests of alternate analysis choices to assert that conclusions remain unchanged. Moreover, complex forecasting tasks need to be carried out for the survey design and planning process of forthcoming CMB experiments, such as CMB-S4 (Abazajian et al. 2016). Ergo, fast, flexible, and robust tools for the analysis of CMB data are indispensable.

In this work, we present `candl`: an automatically differentiable, stand-alone python likelihood designed for the analysis of high-precision primary CMB (TT/TE/EE) and lensing power spectrum ($\phi\phi$) measurements. Calculating derivatives of the likelihood is easy and robust with `candl`; this ready access to Fisher matrices trivialised a plethora of common analysis tasks, including forecasting, propagating biases from band powers to parameters, and finding the maximum of the likelihood surface, which corresponds to the best-fit point. We show explicit examples of these tasks in this work, including optimising the band power bin width of a realistic mock data set to minimise parameter errors and calculating the correlation of parameter constraints from correlated and partially overlapping subsets of the SPT-3G 2018 TT/TE/EE data set (Balkenhol et al. 2023). `candl` interfaces seamlessly with established cosmological software, such as Cobaya (Torrado & Lewis 2021), MontePython (Brinckmann & Lesgourgues 2019; Audren et al. 2013), CAMB (Lewis et al. 2000), CLASS (Blas et al. 2011), and CosmoPower (Spurio Mancini et al. 2022). The likelihood code will

[★] `candl` is available at <https://github.com/Lbalkenhol/candl>

form a part of the analyses pipelines of the upcoming SPT-3G TT/TE/EE and $\phi\phi$ data releases.

`candl` likelihoods are differentiable thanks to JAX (Bradbury et al. 2018)¹. JAX is a python-library for high-performance array computations and is as such well-suited for CMB likelihoods. JAX offers just-in-time compilation, GPU optimisation, and automatic vectorisation. Crucially, JAX features an automatic differentiation algorithm, which can be used to calculate derivatives of complicated functions quickly and robustly. Due to the aforementioned advantages, JAX has seen increasing use in cosmology (Campagne et al. 2023; Piras & Spurio Mancini 2023; Kvasiuk & Münchmeyer 2024). We refer the reader to Campagne et al. (2023) for an excellent introduction to differentiable likelihoods and the wider potential JAX holds for the field of Cosmology. Campagne et al. (2023) touch on many of the techniques discussed in this work and pedagogically show the benefits of a differentiable framework. We understand `candl` as a related effort to JAX-COSMO (Campagne et al. 2023), designed specifically for the case of CMB data analysis and its particular requirements. Still, `candl` can be used without JAX for traditional inference.

This work is structured as follows. In Sect. 2 we discuss relevant background information. We provide an overview of the core likelihood code and additional aspects of the released material in Sect. 3. In Sect. 4 we highlight applications of the differentiability aspect of `candl`. We show how to get started using the code in Sect. 5 before concluding in Sect. 6.

2. Background

The likelihood \mathcal{L} is the probability of measuring the data band powers at hand given a certain model. In the simplest case, the likelihood is Gaussian:

$$-2 \log \mathcal{L}(\theta|\hat{D}) = \left\{ \hat{D} - \mathcal{T} \left[D^{\text{CMB}}(\theta), \theta \right] \right\}^T C^{-1} \left\{ \hat{D} - \mathcal{T} \left[D^{\text{CMB}}(\theta), \theta \right] \right\}, \quad (1)$$

where \hat{D} are the measured band powers with covariance C , θ are the model parameters (cosmological and nuisance), D^{CMB} is the CMB power spectrum, and \mathcal{T} represents the ‘data model’, that is the transformation of the CMB power spectra such that they can be compared to the data, for example foreground contamination, binning, calibration. Since the covariance matrix is constant we ignore the contribution of its determinant (Percival & Brown 2006; Hamimeche & Lewis 2008; Gerbino et al. 2020).

In practice, the calculation of Eq. (1) is typically split into two parts. First, the ‘theory code’ calculates $D^{\text{CMB}}(\theta)$, that is the CMB spectra for a given set of cosmological parameters; the theory code may be a Boltzmann solver such as CAMB (Lewis et al. 2000) or CLASS (Blas et al. 2011), or a surrogate emulator such as CosmoPower (Spurio Mancini et al. 2022), Capse.jl (Bonici et al. 2024), or the algorithm put forward by Günther (2023). Second, the ‘likelihood code’, or simply ‘likelihood’ for short, transforms the model spectra and completes the calculation. `candl` implements this part.

3. Overview

`candl` is a python-based, JAX-powered likelihood: its likelihoods are automatically differentiable, offering ready access

¹ <https://github.com/google/jax>

to robust derivatives and Fisher matrices. The code is self-contained, such that one can easily interact with a given data set and interface `candl` with other cosmology software smoothly. `candl` can be installed via `pip install candl-like`; detailed instructions are provided on the GitHub repository².

We have made the following public CMB data sets available with the release of `candl`:

1. SPT-3G 2018 TT/TE/EE (Balkenhol et al. 2023),
2. SPT-3G 2018 $\phi\phi$ (Pan et al. 2023),
3. ACT DR4 TT/TE/EE (Aiola et al. 2020; Choi et al. 2020)³,
4. ACT DR6 $\phi\phi$ (Madhavacheril et al. 2024; Qu et al. 2024).

The reimplementations of the above public data sets we provide match the log likelihood value of their corresponding official releases to within $\lesssim 10^{-4}\%$ given the same input CMB power spectra and nuisance parameter values. The upcoming SPT TT/TE/EE and $\phi\phi$ releases will use `candl` and we are committing to including other data sets as they become available to provide a unified framework to access and use popular CMB data.

Along with the core likelihood code, we supply a suite of auxiliary tools designed to help the user carry out common analysis tasks. Furthermore, we release a series of example notebooks that showcase different applications of the likelihood. These resources can be found on the GitHub repository and include:

- A showcase of the automatic-differentiation features of the likelihood, including the calculation of Fisher matrices and use of gradient-based minimisation and sampling algorithms.
- An interface between the likelihood code and the commonly-used Markov chain Monte Carlo (MCMC) packages Cobaya (Torrado & Lewis 2021) and MontePython (Brinckmann & Lesgourgues 2019; Audren et al. 2013), as well as an interface with the theory codes CAMB (Lewis et al. 2000), CLASS (Blas et al. 2011), CosmoPower (Spurio Mancini et al. 2022), CosmoPower-JAX (Piras & Spurio Mancini 2023), and Capse.jl (Bonici et al. 2024) independent of an MCMC sampler.
- Tools for the generation of mock data, the construction of a minimum-variance combination of the data from a multi-frequency version, and various inter-frequency consistency tests.

4. Applications of a differentiable inference pipeline

In this section, we highlight a range of applications enabled by the pairing of `candl` with a differentiable theory code. This creates a fully differentiable pipeline and provides easy access to derivatives of the likelihood with respect to cosmological and nuisance parameters, trivialising for example the calculation of robust Fisher matrices without the need for finite-difference methods.

As Campagne et al. (2023) discuss, the most common solution for a differentiable theory code is to use a differentiable emulator (e.g. Spurio Mancini et al. 2022; Piras & Spurio Mancini 2023; Bonici et al. 2024; Nygaard et al. 2023) of spectra computed with the Boltzmann solvers CAMB (Lewis et al. 2000) or CLASS (Blas et al. 2011). We note that fully differentiable Boltzmann solvers have recently been developed in JAX

² <https://github.com/Lbalkenhol/candl>

³ The frequency-compressed, foreground marginalised, ‘CMB-only’ likelihood.

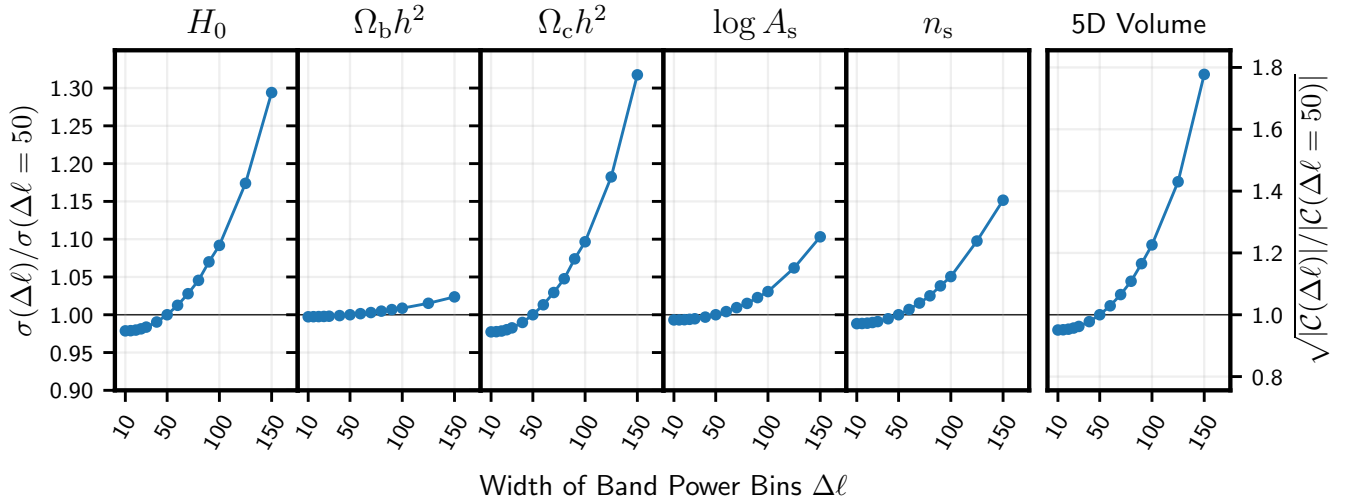


Fig. 1. Relative size of error bars of Λ CDM parameters ($\Omega_c h^2$, $\Omega_b h^2$, H_0 , n_s , $\log A_s$) for different bin sizes $\Delta\ell$ with respect to $\Delta\ell = 50$ for a mock CMB data set. In the rightmost panel, we show the determinant of the parameter covariance C , a proxy for the five-dimensional volume, using a different scale for the vertical axis. We note that the parameter that is affected the most by the broadening of bins is H_0 , which relies on an accurate measurement of the acoustic peak positions. Due to the quadratic relationship between $\Delta\ell$ and σ , bin sizes smaller than $\Delta\ell = 50$ yield a negligible improvement on marginalised parameter errors ($\leq 3\%$). Forecasts are fast and easy with `candl`, enabling thorough analysis optimisation.

(Hahn et al. 2023) and Julia (Ruiz-Zapatero et al. 2024) (BOLT.jl⁴, Li & Sullivan, in prep.). We anticipate a straightforward interface between `candl` and the JAX-based work of Hahn et al. (2023), similar to the one we have realised with CAMB and CLASS. Fundamentally, the log likelihood evaluation in `candl` is a function call, which requires a dictionary of parameter values as its argument; this function can be concatenated with the evaluation call of any Boltzmann solver with a minimal interface performing any necessary data format conversions. When coupling `candl` with JAX-based theory codes, the full advantages of the library remain available, whereas an inter-language interface (e.g. with Julia codes) may surrender these.

In the examples below, we use the CosmoPower CMB primary power spectrum emulators trained on high-accuracy CAMB spectra developed and used in Balkenhol et al. (2023) together with CosmoPower-JAX (Piras & Spurio Mancini 2023)⁵. In using emulators, one must be careful to match the accuracy of the emulator to the requirements set by the data set; we have encountered no such issues with the chosen models and data sets for the purpose of these examples.

4.1. Forecasting

CMB power spectrum measurements are typically binned into band powers for two reasons: first, in order to limit the computational cost of the likelihood evaluation by keeping the data vector small and second, because masking leads to mode-coupling, which prohibits the reconstruction of all individual modes. However, excessively coarse bins risk losing sensitivity to fine-grained features of the power spectrum, thus unnecessarily widening constraints on cosmological parameters. Therefore, we wish to optimise the bin width $\Delta\ell$ and find the point of diminishing returns⁶.

⁴ <https://github.com/xzackli/Bolt.jl>

⁵ These trained models are publicly available: <https://github.com/alessiospurio Mancini/cosmopower>

⁶ Practically, the largest array size feasible for computations is to be decided with the available hardware and the number of spectra in mind.

In this example, we consider a 150 GHz TT/TE/EE mock data set covering $\ell = 300 - 4000$. The covariance matrix contains realistic noise contributions and off-diagonal correlations based on the SPT-3G 2018 TT/TE/EE data set; we use a rectangular survey mask covering approximately 4% of the sky and similar noise curves though with a white noise level of $5\mu K$ arcmin. We use the foreground model from the SPT-3G 2018 TT/TE/EE data set⁷ and the same *Planck*-based prior on the optical depth to reionisation τ (Planck Collaboration VI 2020). For this data set, we consider 14 binning options in the range $\Delta\ell = 10 - 150$, allowing bins to exceed the maximum multipole moment if necessary. This has an insignificant effect on our results due to the negligible cosmological information at these scales and the noise in the data.

To accomplish this forecasting task, we instantiate a `candl` likelihood for each binning option and calculate the Fisher matrix (Heavens 2016) using CosmoPower-JAX and the supplied tools. We plot expected parameter errors on the cosmological parameters $\Omega_c h^2$, $\Omega_b h^2$, H_0 , n_s , and $\log A_s$ in Fig. 1. In addition, we plot the determinant of the parameter covariance matrix as a proxy for the five-dimensional volume. We observe a quadratic relationship between these metrics and the bin width, with little improvement for bins finer than $\Delta\ell = 50$: $\sigma(\Delta\ell = 50)/\sigma(\Delta\ell = 10) \leq 1.03$ for all parameters. This is in favour of the choices made by comparable data sets (cf. Dutcher et al. 2021; Balkenhol et al. 2023).

Given that the example likelihood has 25 total parameters (cosmological and nuisance) and deals with covariance matrices with up to $O(10^6)$ elements, this is no trivial task. Nevertheless, in the differentiable framework this exercise is easily completed in a few minutes on a laptop, and requires no special structuring of the code to minimise computing costs; this allows for a straightforward translation of ideas into code. The exercise above can easily be repeated for different covariance matrices based on different survey masks and noise levels, showing how `candl` can be used to optimise analysis choices.

⁷ We do not include the CIB-tSZ term due to its non-differentiable form.

4.2. Propagating band power biases to parameters

Next, we highlight how `candl` can be used to propagate a bias to band powers $\delta\hat{D}_\ell$ through to parameters $\delta\theta$. This computation involves the derivative of the model spectra with respect to parameters, $\partial D_\ell/\partial\theta$, and the Fisher matrix, F (see e.g. [Heavens 2016](#)):

$$\delta\theta = F^{-1} \frac{\partial D_\ell}{\partial\theta} C^{-1} \delta\hat{D}_\ell. \quad (2)$$

We use this formalism to reproduce the impact of a 5% rescaling of the TE power spectrum measurement of the ACT DR4 data set. As shown in [Aiola et al. \(2020\)](#), this increase in the TE power leads to a shift in the $\Omega_b h^2 - n_s$ plane. Using `candl`, we find a shift down in $\Omega_b h^2$ of 1.6σ and a shift up in n_s of 1.4σ . This appears consistent with what is shown in Fig. 14 of [Aiola et al. \(2020\)](#).

This result was reproduced within seconds and thanks to automatic differentiation did not require any tuning of step size parameters. In principle, one can therefore map out the underlying degeneracies in parameter space with ease by rescaling the TE spectrum incrementally. We note that this approach does not yield an updated parameter covariance, which the re-running of MCMC chains done in [Aiola et al. \(2020\)](#) produces, though given the comparatively small shift to the best-fit point, we expect that this effect is negligible⁸.

We note that [Mukherjee & Wandelt \(2018\)](#) have used this method to obtain parameter constraints from different patches of the *Planck* data. By propagating the difference between the data band powers and a model spectrum to parameters, one can calculate the best-fit point of the data using an anchor point. The accuracy of this method depends on the distance between the data best-fit and anchor points. Applications unlocked by a differentiable pipeline can be used to accomplish a variety of analysis tasks.

4.3. Parameter-level correlations of correlated band powers

We now use `candl` and `CosmoPower-JAX` to calculate the covariance of best-fit parameters of correlated data sets. We follow the prescription of [Kable et al. \(2020\)](#):

$$\langle(\bar{\theta}^X - \langle\bar{\theta}^X\rangle)(\bar{\theta}^Y - \langle\bar{\theta}^Y\rangle)\rangle = (M^X)^T C^{XY} M^Y, \quad (3)$$

where $\bar{\theta}^{X,Y}$ are the best-fit parameters of data sets X and Y , C^{XY} is the relevant block of the band power covariance matrix describing the correlation of X and Y and

$$M^X = (C^{XX})^{-1} \frac{\partial D_\ell}{\partial\theta} (F^{XX})^{-1}, \quad (4)$$

where C^{XX} is block the band power covariance matrix for data set X and F^{XX} the Fisher matrix when only considering data set X . The derivatives of the model spectra are typically evaluated at the best-fit point of the full data set ([Kable et al. 2020](#)).

We use the SPT-3G 2018 $TT/TE/EE$ data set and investigate the following subsets: TT , TE , and EE spectra, and large and small angular scale measurements (split at $\ell = 1500$)⁹. On a

⁸ Though if required, this effect can be approximated by rescaling the parameter covariance matrix obtained from the original MCMC run using the Hessian evaluated at the original and the offset parameter points.

⁹ For this example, we ignore the tSZ-CIB correlation term, as the functional form chosen in [Balkenhol et al. \(2023\)](#) is not differentiable. Given the tests presented in Sect. IVD therein we expect the impact of this to be negligible.

technical level, the principle task is the calculation of Fisher matrices for all subsets of interest as well as the calculation of the derivatives of the model spectra.

We plot the correlation between the cosmological parameters H_0 , $\Omega_b h^2$, $\Omega_c h^2$, n_s , and $\log A_s$ in Fig. 2. We see that there is little to no correlation between the constraints from different spectra (TT , TE , and EE); the highest correlation we observe is 0.14 for n_s between TT and TE . This is testament to the independent information CMB polarisation measurements hold ([Galli et al. 2014](#)). As expected, constraints from large and small angular scale data are independent of one another; in other words, the off-diagonal elements of each covariance block are small. While these two results may be expected, there exist non-negligible correlations between these two classes of subsets, which are unknown a priori and only revealed through this calculation.

Access to such correlation matrices facilitates a deeper understanding of the constraints a data set produces in a given model. A differentiable inference pipeline makes this exercise fast, reliable, and easy. We note that an alternative approach to calculating the above parameter correlations exists, in which one simulates mock band powers and find their best-fit values; this was done by [Kable et al. \(2020\)](#) to verify the above procedure. We note that this approach is also accelerated by `candl` thanks to access to gradient-based minimisers (see Sect. 4.4) or using the method of [Mukherjee & Wandelt \(2018\)](#) described in Sect. 4.2.

4.4. Smart exploration of the likelihood surface

Up until now, we have focussed on applications that are greatly simplified by a differentiable inference pipeline, but in principle with more time, computation power, and caution achievable using finite-difference methods. We now discuss gradient-based explorations of the likelihood surface, which only become feasible thanks to the easy access to derivatives `candl` facilitates through JAX's automatic differentiation algorithm.

A common analysis task is determining the minimum of the likelihood, that is the best-fit point of the data. Gradient-based minimisers allow for a quick descent to the best-fit point and simple algorithms often suffice for Gaussian model spaces when a good starting guess is available. This is typically the case for primary power spectrum analysis in Λ CDM, where data is rather constraining and results from previous analyses (e.g. *Planck*, [Planck Collaboration VI 2020](#)) can be used as a starting point.

As an example, we show the trajectory of two gradient-based minimisers for ACT DR4 $TT/TE/EE$ in Λ CDM in the left panel of Fig. 3. We show a simple Newton-Raphson algorithm (e.g. [Gill et al. 2019](#)) alongside the ADAM minimiser ([Kingma & Ba 2014](#)). The former is especially convenient as it requires the computation of the Hessian, which can then be repurposed as the proposal matrix for conventional MCMC chains and short-cut learning this matrix adaptively. For the ADAM minimiser, we interface the likelihood with `Optax`¹⁰ ([DeepMind et al. 2020](#)), a JAX-based library of minimisation tools that is widely used in machine-learning. This algorithm is stochastic and offers protections against getting stuck in local minima. In this simple example, the Newton-Raphson algorithm and ADAM minimiser perform 20 and 100 steps, respectively. The two algorithms agree to within $\Delta\chi^2 \lesssim 0.2$ and run within seconds. For reference, the BOBYQA algorithm ([Powell 2009](#); [Cartis et al. 2022, 2018a](#)) implemented in `Cobaya` typically requires around 250 steps for this problem.

¹⁰ <https://github.com/google-deeppmind/optax>

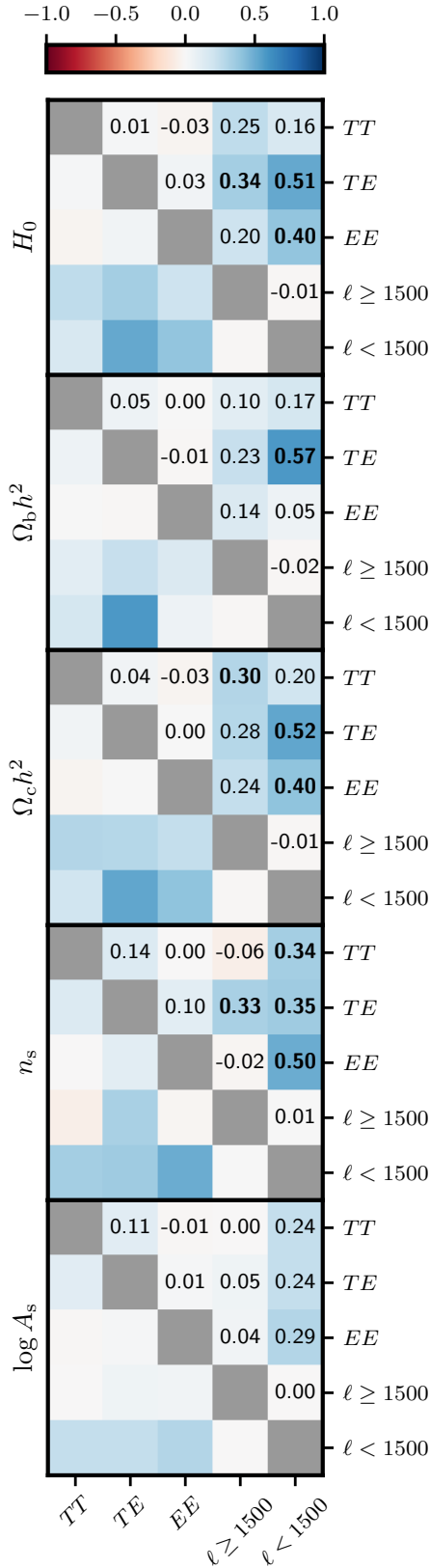


Fig. 2. Correlation between best-fit Λ CDM parameters of the TT , TE , EE , $\ell \geq 1500$, and $\ell < 1500$ data of SPT-3G 2018 $TT/TE/EE$ (values $\geq 30\%$ in bold). We find non-negligible correlation between spectrum types and scale cuts. `cand1` makes this calculation simple and robust, enabling a deeper understanding of parameter constraints.

In addition to determining the best-fit point of the data, we are often interested in the posterior distribution of parameters, which is typically obtained by building up MCMC samples. Here, we can use gradient-based MCMC sampling techniques, for example Hamiltonian Monte Carlo (Duane et al. 1987), Microcanonical Hamiltonian Monte Carlo (Robnik et al. 2022), or No-U-Turn (NUTS) sampling (Hoffman & Gelman 2011). In these frameworks, the likelihood is typically treated as a potential and the motion of a test mass is modelled. The result is a higher acceptance rate and a decreased correlation between sample points compared to the Metropolis-Hastings (MH) algorithm; this translates to a better performance in high-dimensional spaces and shorter MCMC chains.

In the right panel of Fig. 3, we show a comparison of the performance of MH and NUTS algorithms for ACT DR4 $TT/TE/EE$ in Λ CDM. The NUTS sampling was done with the aid of BlackJAX¹¹ (Cabezas et al. 2024), a straightforward sampling software written in JAX. Figure 3 contains an excerpt of H_0 samples of MH and NUTS chains; we see by eye that the correlation between neighbouring samples is indeed much lower for the gradient-based algorithm. We calculate the discrete Fourier transform of the H_0 samples for each of the two chains and plot the absolute square of the Fourier coefficients, which corresponds to the power spectrum $P(k)$, in the bottom panel. We apply the fitting procedure of Dunkley et al. (2005) to the power spectra. The fits highlight the suppression of small-scale power for the MH chain due to correlations between close-by point. In contrast, the NUTS samples resemble more closely the ideal case of uncorrelated white noise. The power spectrum at $k = 0$, P_0 , of NUTS samples is also lower compared to MH samples ($P_0^{\text{NUTS}}/P_0^{\text{MH}} = 0.9$), indicating a higher efficiency and faster convergence (Dunkley et al. 2005; Hajian 2007). Lastly, in the example above the proposal matrix of the MH algorithm is set to achieve a close-to optimal acceptance rate of ~ 0.25 , meaning that three out of four proposed steps are rejected (Roberts et al. 1997; Dunkley et al. 2005). By contrast, the optimal NUTS acceptance rate is ~ 0.6 (Hoffman & Gelman 2011) and the default target of contemporary implementations is often still higher, for example 0.8 for BlackJAX and NumPyro (Phan et al. 2019; Bingham et al. 2019; Cabezas et al. 2024), which is reflected in the example above.

While short chains like in the example above (both chains have around 5000 samples) run in circa one minute on appropriate hardware, for a full MCMC run one must weigh the advantages of different algorithms against their computational costs for the specific problem at hand. Important aspects to consider are: the size of the data vector¹², the number of likelihood evaluations per step, the correlation of samples, the dimensionality of the parameter space, and the available hardware. One particularly helpful metric in this assessment is the number of effective samples of a chain, which are the number of independent points with the same expression power (Whiteley et al. 2016). Running ACT DR4 Λ CDM chains on CPUs and GPUs, we find that NUTS sampling increases the number of effective samples produced per second by around a factor of four compared to MH sampling. Put differently, the gradient-based sampler only needs to run for a quarter of the time of the traditional algorithm to achieve the same performance. We remind

¹¹ <https://github.com/blackjax-devs/blackjax>

¹² We find that the likelihood evaluation time scales quadratically with the length of the data vector.

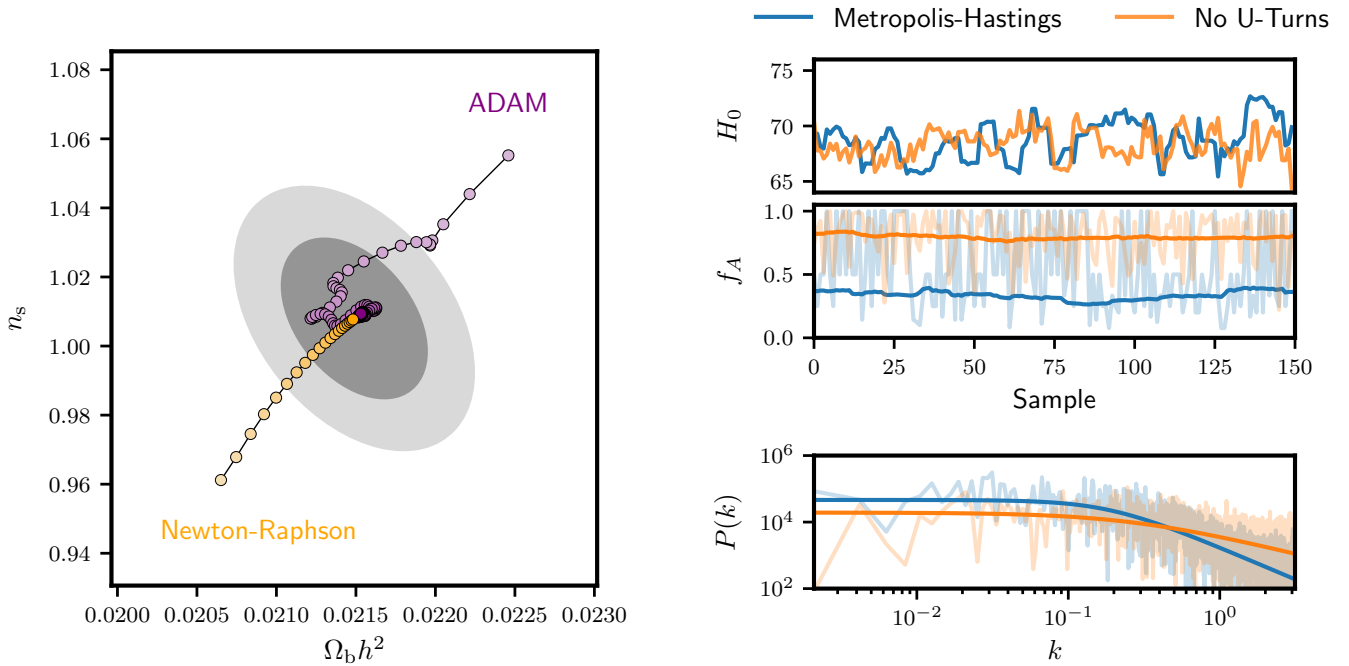


Fig. 3. Gradient-based explorations of the ACT DR4 likelihood surface in Λ CDM. Left: Newton-Raphson (orange) and ADAM (purple) minimisers use gradient information to find the best-fit point of the ACT DR4 data set (light to dark points). We highlight steps in the $n_s - \Omega_b h^2$ plane, though the minimisation is performed over all Λ CDM parameters and the calibration parameter of the data set. While the Newton-Raphson algorithm directly descends to the best-fit point, the ADAM minimiser follows a more complicated trajectory with some stochasticity to ensure it does not get trapped in local minima. The grey contours in the background show the 1 and 2σ confidence region obtained from the Hessian, which is a by-product of the Newton-Raphson algorithm. Right: comparison of Metropolis-Hastings (MH; blue) and No-U-Turns (orange) chains run using the ACT DR4 data set in Λ CDM. We focus on H_0 samples here, though the MCMC chains explore the full parameter space. The top and middle panels show the H_0 values and acceptance rate f_A (values in faint lines, rolling average over 50 samples in full colour) for an excerpt of 150 samples, respectively. The bottom panel shows the power spectrum $P(k)$ of the H_0 samples of the full-length chain (measured spectra in faint lines, parametric fit (Dunkley et al. 2005) in full colour). The NUTS samples are less correlated by eye than the MH samples, which is confirmed by the flatter power spectrum. Additionally, the acceptance rate of the gradient-based algorithm is substantially higher, with few rejected samples.

the reader that an example notebook covering the three gradient-based algorithms discussed in this section can be found on the GitHub repository.

4.5. Other applications

Finally, we briefly highlight other applications of `candl` not discussed so far. Running example calculations for these suggestions is beyond the scope of this work, though we wish to highlight them here as they stand to benefit from a differentiable likelihood.

A differentiable framework facilitates the construction of compressed likelihoods as was already demonstrated by Campagne et al. (2023) for the Massively Optimised Parameter Estimation and Data compression (Heavens et al. 2000, 2017; Zablocki & Dodelson 2016) algorithm. However, for CMB analysis the more common data compression strategy is to form a set of foreground-marginalised CMB-only band powers from the multifrequency data vector. This method was put forward by Dunkley et al. (2013) and has since been applied by the *Planck* and ACT collaborations (Planck Collaboration XI 2016; Louis et al. 2017; Planck Collaboration V 2020; Choi et al. 2020; Aiola et al. 2020). Crucially, the framework introduced by Dunkley et al. (2013) is independent of the cosmological model, such that no theory code (differentiable or otherwise) is required. One samples directly from the likelihood of the data given a CMB power spectrum and nuisance parameter values. Typically, this procedure involves alternating Gibbs and MH sampling to

build up samples of the CMB band powers and their covariance matrix.

This type of data compression is simplified by `candl`. Feeding the likelihood into a Newton-Raphson minimiser appears particularly helpful: the best-fit point yields the correct CMB-only band powers and as a by-product the Hessian provides a good approximation to the covariance matrix. Following on from this, one can then use a single gradient-based sampling algorithm, such as NUTS, to fully characterise the posterior of CMB-only band powers and obtain their covariance.

Furthermore, easy access to derivatives of the likelihood makes MCMC sampling with the use of the Laplace approximation feasible. In this approach, one fixes the nuisance parameters to their best-fit values, constructing a so-called profile likelihood. This reduces the dimensionality of the parameter space to only the parameters of interest, which in turn allows for a faster characterisation of the likelihood surface. However, one needs to account for the volume reduction as otherwise constraints on cosmological parameters would be artificially tight; this correction is done by adding a gradient-dependent contribution, the Laplace term, to the obtained χ^2 values. Thanks to `candl`, this correction can be computed robustly and sampling accelerated. We refer the reader to Hadzhiyska et al. (2023) for a detailed explanation of this method and examples in the cosmology context.

Another application `candl` seems well suited for is the reconstruction of the primordial power spectrum, $P(k)$, as well as the search for particular features within $P(k)$. In these types of

analyses, one typically adds a large number of additional parameters to the regular cosmological and nuisance parameters, which describe $P(k)$ through one formalism or another (Gauthier & Bucher 2012; Hazra et al. 2013; Handley et al. 2019; Planck Collaboration XXII 2014). For example, a common approach is to model $P(k)$ through N nodes connected by an interpolation scheme and marginalising over the amplitudes and positions of the nodes, hence adding $2N$ parameters. In this case, traditional MCMC approaches, such as MH sampling, are inefficient, since the required length of chains scales approximately linearly with dimensionality (Dunkley et al. 2005) and the posteriors tend to show complicated features, such as multimodality, which make it difficult to obtain robust constraints (Handley et al. 2019). Nested sampling can aid in these scenarios and is helpful for model comparison (Skilling 2006; Handley et al. 2015b,a), though a fully differentiable pipeline with `candl` opens the door to further options. Gauthier & Bucher (2012) have already applied a Newton-Raphson minimiser to obtain best-fit solutions in this context; we can expand on this by using gradient-based sampling methods, which scale better in higher dimensional spaces (e.g. Hoffman & Gelman 2011), to fully characterise the posterior.

5. Getting started

Lastly, we provide brief instructions for how to get started using `candl`. Apart from the final example, the operations below do not require the installation of JAX or a power spectrum emulator; we refer interested users to the tutorials and in-depth explanations on the GitHub repository.

To initialise a `candl` likelihood, it suffices to point the code to a data set file or use a short cut for a released data set. For example:

```
import candl
import candl.data
candl_like = candl.like(candl.data.SPT3G_2018_TTTEEE)
```

We can then access all aspects of the data set, such as the band powers, the covariance matrix, the window functions, as well as the data model through attributes of the likelihood instance.

If we have a set of CMB spectra at hand (`CMB_specs` below) and want to calculate the corresponding χ^2 value we can proceed with:

```
pars = {'Dl': CMB_specs,
        'nuisance_par': 1.0, ...}
chi2 = candl_like.chi_square(pars)
```

where in `pars` we pass the CMB spectrum we would like to test, as well as any required nuisance parameter values.

One of the most common use cases for CMB likelihoods is the use with popular sampling software, such as Cobaya. This is done via the interface module:

```
import candl.interface
from cobaya.run import run
cobaya_info = {'likelihood':
               candl.interface.
               get_cobaya_info_dict_for_like(candl_like), ...}
updated_cobaya_info = run(cobaya_info)
```

The function `get_cobaya_info_dict_for_like()` deals with the interface between the two pieces of software. The user can then select the theory code and sampler they wish and pass these to Cobaya as per usual (see the Cobaya documentation for further details¹³). A similarly light interface is supplied for

MontePython (Brinckmann & Lesgourgues 2019; Audren et al. 2013).

Lastly, to calculate Fisher matrices using JAX's automatic differentiation, we assume that we have a differentiable function at hand, `pars_to_theory_specs`, that moves from a dictionary of parameter values to CMB spectra. Helper functions exist in `candl` to generate such functions based on CosmoPower models, as well as non-differentiable functions based on CAMB, CLASS, Capse.jl, or any Cobaya.Theory subclass. Obtaining Fisher matrices is then straightforward with the `tools` module:

```
import candl.tools
fisher_mat, par_order = candl.tools.get_fisher_matrix(
    pars_to_theory_specs,
    candl_like,
    pars,
    return_par_order = True)
```

Here `fisher_mat` is the Fisher matrix in the order given by `par_order` and we use the same evaluation point (`pars`) as before.

6. Conclusions

In this work, we have presented `candl`: an automatically differentiable likelihood designed for the analysis of high-precision CMB data. Through the use of the JAX library, `candl` offers easy access to derivatives of the likelihood, trivialising in particular the calculation of robust Fisher matrices. `candl` is a python-based package that places emphasis on flexibility and efficiency, allowing the user to easily access all aspects of a data set. The code can easily be interfaced with other established software for instance for MCMC sampling.

We have paired `candl` with a differentiable theory code, CosmoPower-JAX, and performed a series of example calculations on real and simulated data, showing how the easy access to reliable Fisher matrices opens up new ways of analysing CMB data. In particular, we have calculated the evolution of parameter errors with respect to the width of band power bins for a mock data set and the correlation of parameter constraints from correlated and overlapping subsets of the SPT-3G 2018 TT/TE/EE data set. Moreover, we have showcased gradient-based explorations of the likelihood surface through the use of minimisers and samplers.

The use of differentiable methods in CMB data analysis allows for the construction of a more flexible and reliable pipeline: an increased number of consistency tests and variations on the baseline analysis. This is crucial for upcoming data sets, which have the potential to surpass *Planck*-precision on cosmological parameters and have the responsibility to deliver a robust analysis. `candl` delivers the necessary tools for this task.

We are eager to improve `candl` in the future. `candl` will be used as a part of the upcoming primary anisotropy and lensing analyses of the SPT-3G collaboration; the associated data sets, as well as those of other CMB experiments, will be made accessible through `candl` once public. We are exploring ways of optimising the code to increase its speed. We are also interested in interfacing the code with probabilistic programming languages, such as NumPyro (Phan et al. 2019; Bingham et al. 2019). Lastly, we anticipate the adoption of gradient-based methods by widespread cosmology software in the near future and will ensure our likelihood continues to interface with these neatly.

Acknowledgements. The authors thank Jo Lauritz for the logo design and are grateful for feedback on `candl` from Étienne Camphuis and the South Pole Telescope Collaboration. `candl` uses JAX (Bradbury et al. 2018) and the scientific python stack (Hunter 2007; Virtanen et al. 2001; van der Walt et al. 2011).

¹³ <https://cobaya.readthedocs.io/en/latest/index.html>

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 101001897). This work has received funding from the Centre National d'Etudes Spatiales and has made use of the Infinity Cluster hosted by the Institut d'Astrophysique de Paris. C.T. was supported by the Center for Astrophysical Surveys (CAPS) at the National Center for Supercomputing Applications (NCSA), University of Illinois Urbana-Champaign. This work made use of the Illinois Campus Cluster, a computing resource that is operated by the Illinois Campus Cluster Program (ICCP) in conjunction with the National Center for Supercomputing Applications (NCSA) and which is supported by funds from the University of Illinois at Urbana-Champaign.

References

- Abazajian, K. N., Adshead, P., Ahmed, Z., et al. 2016, arXiv e-prints [arXiv:1610.02743]
- Aiola, S., Calabrese, E., Maurin, L., et al. 2020, *J. Cosmol. Astropart. Phys.*, 2020, 047
- Audren, B., Lesgourgues, J., Benabed, K., & Prunet, S. 2013, *J. Cosmol. Astropart. Phys.*, 2013, 001
- Balkenhol, L., Dutcher, D., Spurio Mancini, A., et al. 2023, *Phys. Rev. D*, 108, 023510
- Bingham, E., Chen, J. P., Jankowiak, M., et al. 2019, *J. Mach. Learn. Res.*, 20, 28:1
- Blas, D., Lesgourgues, J., & Tram, T. 2011, *J. Cosmol. Astropart. Phys.*, 2011, 034
- Bonici, M., Baxter, E., Bianchini, F., & Ruiz-Zapatero, J. 2024, *Open J. Astrophys.*, 7, 10
- Bradbury, J., Frostig, R., Hawkins, P., et al. 2018, <http://github.com/google/jax>
- Brinckmann, T., & Lesgourgues, J. 2019, *Phys. Dark Universe*, 24, 100260
- Cabezas, A., Corenflos, A., Lao, J., et al. 2024, arXiv e-prints [arXiv:2402.10797]
- Campagne, J.-E., Lanusse, F., Zuntz, J., et al. 2023, *Open J. Astrophys.*, 6, 15
- Carlstrom, J. E., Ade, P. A. R., Aird, K. A., et al. 2011, *PASP*, 123, 568
- Cartis, C., Fiala, J., Marteau, B., & Roberts, L. 2018a, arXiv e-prints [arXiv:1804.00154]
- Cartis, C., Roberts, L., & Sheridan-Methven, O. 2022, *Optimization*, 71, 2343
- Choi, S. K., Hasselfield, M., Ho, S.-P. P., et al. 2020, *J. Cosmol. Astropart. Phys.*, 2020, 045
- DeepMind, Babuschkin, I., Baumli, K., et al. 2020, <http://github.com/google-deepmind>
- Duane, S., Kennedy, A. D., Pendleton, B. J., & Roweth, D. 1987, *Phys. Lett. B*, 195, 216
- Dunkley, J., Bucher, M., Ferreira, P. G., Moodley, K., & Skordis, C. 2005, *MNRAS*, 356, 925
- Dunkley, J., Calabrese, E., Sievers, J., et al. 2013, *J. Cosmol. Astropart. Phys.*, 7, 25
- Dutcher, D., Balkenhol, L., Ade, P. A. R., et al. 2021, *Phys. Rev. D*, 104, 022003
- Galli, S., Benabed, K., Bouchet, F., et al. 2014, *Phys. Rev. D*, 90, 063504
- Gauthier, C. & Bucher, M. 2012, *J. Cosmol. Astropart. Phys.*, 2012, 050
- Gerbino, M., Lattanzi, M., Migliaccio, M., et al. 2020, *Front. Phys.*, 8, 15
- Gill, P. E., Murray, W., & Wright, M. H. 2019, *Practical Optimization* (Philadelphia, PA: SIAM Publications Library)
- Günther, S. 2023, arXiv e-prints [arXiv:2307.01138]
- Hadzhiyska, B., Wolz, K., Azzoni, S., et al. 2023, *Open J. Astrophys.*, 6, 23
- Hahn, O., List, F., & Porqueres, N. 2023 JCAP, submitted [arXiv:2311.03291]
- Hajian, A. 2007, *Phys. Rev. D*, 75, 083525
- Hamimeche, S., & Lewis, A. 2008, *Phys. Rev. D*, 77, 103013
- Handley, W. J., Hobson, M. P., & Lasenby, A. N. 2015a, *MNRAS*, 450, L61
- Handley, W. J., Hobson, M. P., & Lasenby, A. N. 2015b, *MNRAS*, 453, 4384
- Handley, W. J., Lasenby, A. N., Peiris, H. V., & Hobson, M. P. 2019, *Phys. Rev. D*, 100, 103511
- Hazra, D. K., Shafieloo, A., & Souradeep, T. 2013, *Phys. Rev. D*, 87, 123528
- Heavens, A. 2016, *Entropy*, 18, 236
- Heavens, A. F., Jimenez, R., & Lahav, O. 2000, *MNRAS*, 317, 965
- Heavens, A. F., Sellentin, E., de Mijolla, D., & Vianello, A. 2017, *MNRAS*, 472, 4244
- Hoffman, M. D., & Gelman, A. 2011, arXiv e-prints [arXiv:1111.4246]
- Hunter, J. D. 2007, *Comput. Sci. Eng.*, 9, 90
- Kable, J. A., Addison, G. E., & Bennett, C. L. 2020, *ApJ*, 888, 26
- Kingma, D. P., & Ba, J. 2014, arXiv e-prints [arXiv:1412.6980]
- Kosowsky, A. 2003, in *The proceedings of the workshop on "The Cosmic Microwave Background and its Polarization"* New Astronomy Reviews, eds. S. Hanany, & K. A. Olive (Amsterdam: Elsevier)
- Kvasiuk, Y., & Münchmeyer, M. 2024, *Phys. Rev. D*, 109, 083515
- Lewis, A., Challinor, A., & Lasenby, A. 2000, *ApJ*, 538, 473
- Louis, T., Grace, E., Hasselfield, M., et al. 2017, *J. Cosmol. Astropart. Phys.*, 6, 031
- Madhavacheril, M. S., Qu, F. J., Sherwin, B. D., et al. 2024, *ApJ*, 962, 113
- Mukherjee, S., & Wandelt, B. D. 2018, *J. Cosmol. Astropart. Phys.*, 2018, 042
- Nygaard, A., Holm, E. B., Hannestad, S., & Tram, T. 2023, *J. Cosmol. Astropart. Phys.*, 2023, 025
- Pan, Z., Bianchini, F., Wu, W. L. K., et al. 2023, *Phys. Rev. D*, 108, 122005
- Percival, W. J., & Brown, M. L. 2006, *MNRAS*, 372, 1104
- Phan, D., Pradhan, N., & Jankowiak, M. 2019, arXiv e-prints [arXiv:1912.11554]
- Piras, D., & Spurio Mancini, A. 2023, *Open J. Astrophys.*, 6, 20
- Planck Collaboration VI. 2020, *A&A*, 641, A6
- Planck Collaboration V. 2020, *A&A*, 641, A5
- Planck Collaboration XI. 2016, *A&A*, 594, A11
- Planck Collaboration XXII. 2014, *A&A*, 571, A22
- Powell, M. 2009, *The BOBYQA Algorithm for Bound Constrained Optimization without Derivatives*, Tech. Rep. DAMTP, University of Cambridge
- Qu, F. J., Sherwin, B. D., Madhavacheril, M. S., et al. 2024, *ApJ*, 962, 112
- Roberts, G. O., Gelman, A., & Gilks, W. R. 1997, *Ann. Appl. Probab.*, 7, 110
- Robnik, J., De Luca, G. B., Silverstein, E., & Seljak, U. 2022, arXiv e-prints [arXiv:2212.08549]
- Ruiz-Zapatero, J., Alonso, D., García-García, C., et al. 2024, *Open J. Astrophys.*, 7, 11
- Simons Observatory Collaboration (Ade, P., et al.) 2019, *J. Cosmol. Astropart. Phys.*, 2019, 056
- Skilling, J. 2006, *Bayesian Anal.*, 1, 833
- Spurio Mancini, A., Piras, D., Alsing, J., Joachimi, B., & Hobson, M. P. 2022, *MNRAS*, 511, 1771
- Torrado, J., & Lewis, A. 2021, *J. Cosmol. Astropart. Phys.*, 2021, 057
- van der Walt, S., Colbert, S., & Varoquaux, G. 2011, *Comput. Sci. Eng.*, 13, 22
- Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, *Nat. Methods*, 17, 261
- Whiteley, N., Lee, A., & Heine, K. 2016, *Bernoulli*, 22, 494
- Zablocki, A., & Dodelson, S. 2016, *Phys. Rev. D*, 93, 083525