THESIS FOR THE DEGREE OF LICENTIATE OF ENGINEERING

Efficient sampling of Bayesian posteriors and predictive distributions in χEFT

Isak Svensson

Department of Physics
Chalmers University of Technology
Gothenburg, Sweden, 2021

Efficient sampling of Bayesian posteriors and predictive distributions in χEFT

Isak Svensson

This thesis has been prepared using LaTeX.

Department of Physics
Chalmers University of Technology
SE-412 96 Gothenburg, Sweden
Phone: +46 (0)31 772 1000
www.chalmers.se

Cover illustration: Leapfrog trajectories through phase space recorded during an HMC sampling of a trimodal one-dimensional pdf. The total system energy is conserved along the trajectories. The horizontal and vertical axes represent position and momentum, respectively. The image is produced with a Python code written by the author.

Printed by Chalmers digitaltryck
Gothenburg, Sweden, 2021

Efficient sampling of Bayesian posteriors and predictive distributions in χEFT

Isak Svensson
Department of Physics
Chalmers University of Technology

Abstract

In this thesis I employ Bayesian statistics to quantify parametric and epistemic uncertainties in chiral effective field theories (χEFT) and propagate these forward to predictions of observables in low-energy nuclear physics. Two primary sources of uncertainty—experimental errors and the theoretical error induced by the truncation of the EFT at up to next-to-next-to-leading-order—are modelled and accounted for in the posterior distributions of the unknown low-energy constants (LECs) that govern interaction strengths in χEFT. These posteriors are computationally challenging to extract and I therefore introduce an advanced Markov chain Monte Carlo (MCMC) algorithm, known as Hamiltonian Monte Carlo, and investigate its performance. I compare its sampling efficiency to standard MCMC algorithms and find reductions in computation time by factors around 3-6 in the present work. I exploit the extracted posteriors to produce predictive distributions for neutron-proton and proton-proton scattering cross sections below and above the pion production threshold and check the consistency of the model predictions against empirical data and higher-order point estimates. I find that the predictive distributions provide reliable credibility intervals as long as the size of the truncation error is estimated from expansion coefficients at next-to-leading-order and above. The LEC posteriors are also central to uncertainty quantification in few- and manybody systems, and as part of a larger collaboration I explore constraints on three-nucleon forces imposed by light-nuclei observables.

Keywords: nuclear physics, two-nucleon scattering, chiral effective field theory, Bayesian parameter estimation, Markov chain Monte Carlo

Acknowledgements

I thank my supervisors, Andreas Ekström and Christian Forssén, for guiding me throughout this work. I further thank Martin Hoferichter and Jacobo Ruiz de Elvira for supplying the central values and full covariance matrices for the πN LECs. This work was supported by the European Research Council (ERC) under the European Unions Horizon 2020 research and innovation programme (Grant agreement No. 758027). The computations were enabled by resources provided by the Swedish National Infrastructure for Computing (SNIC) at Chalmers Centre for Computational Science and Engineering (C3SE), the National Supercomputer Centre (NSC) partially funded by the Swedish Research Council.

List of publications

This thesis is based on the following two papers and one published code.

Paper A
Fast and rigorous constraints on chiral three-nucleon forces from few-body observables
S. Wesolowski, I. Svensson, A. Ekström, C. Forssén, R.J. Furnstahl, J. A. Melendez, and D. R. Phillips
Accepted for publication in Phys. Rev. C
e-Print: arXiv:2104.04441

In this paper we explore constraints on chiral three-nucleon forces from bound-state observables in the few-nucleon sector. We include experimental and EFT truncation errors in our statistical model. Observables are calculated using fast eigenvector continuation emulators, whose errors are also taken into account.

Paper B
Bayesian parameter estimation in χEFT using Hamiltonian Monte Carlo
I. Svensson, A. Ekström, and C. Forssén
Manuscript in preparation

In this paper we sample posterior distributions of low-energy constants in chiral effective field theory using Hamiltonian Monte Carlo and investigate the improvement in sampling efficiency this algorithm brings compared to other MCMC algorithms. We adopt a statistical model for the EFT error that take into account the EFT truncation error and LEC uncertainties from experimental errors, and extract posterior predictive distributions for low-energy nuclear observables. We analyze the predictive power of the statistical model by comparing with empirical data and higher-order results.

Code
montepython
I. Svensson
Published under the GNU General Public Licence (GPL)
Available at `https://github.com/svisak/montepython.git`

In this Python code I implement the Hamiltonian Monte Carlo and Metropolis-Hastings sampling algorithms. The code also includes diagnostic tools.

Contents

Chapter 1

Introduction

Precision nuclear physics relies on realistic models of the nuclear interaction potential and predictions with quantified uncertainties. The wavefunction of an interacting A-nucleon system is governed by the non-relativistic A-nucleon Schrödinger equation

$$(\hat{H}_0 + \hat{V}_{2\mathrm{N}} + \hat{V}_{3\mathrm{N}} + \hat{V}_{4\mathrm{N}} + \ldots)\,|\Psi\rangle = E\,|\Psi\rangle\,, \tag{1.1}$$

where $\hat{H}_0$ denotes the kinetic energy operator of the free nucleons, $\hat{V}_{j\mathrm{N}}$ is a j-nucleon potential operator, and E is the eigenenergy of the system. Theoretical predictions for nuclear observables may be found by numerically solving Equation (1.1), a decades-old problem [1] for which there exists several ongoing research efforts using e.g. ab initio methods [2]. The challenge of solving the A-nucleon Schrödinger equation notwithstanding, we need a model for the nuclear potential

$$\hat{V} = \hat{V}_{2\mathrm{N}} + \hat{V}_{3\mathrm{N}} + \hat{V}_{4\mathrm{N}} + \ldots \tag{1.2}$$

which is the main focus of this work. Efforts to characterize the nucleon-nucleon (2N) potential $\hat{V}_{2\mathrm{N}}$ have been underway since the 1930's, with Yukawa achieving the first major breakthrough in 1935 when he proposed the existence of mesons: massive force carriers more than 200 times heavier than the electron [3]. Yukawa expressed pessimism regarding the result as no such particle had been observed, but his hypothesis was eventually proven correct by the discovery of pions (and, later, heavier mesons). However, high-precision nuclear potentials in this vein remained elusive [4].

The discovery of quantum chromodynamics (QCD), the theory of the strong interaction between quarks and gluons, provided a deeper understanding of the strong nuclear interaction. Unfortunately, QCD is nonperturbative

in the low-momentum region below the typical hadronic mass scale (~ 1 GeV) relevant for nuclear physics and it is as a result impossible to apply conventional perturbation theory in powers of the strong coupling constant. One can pursue QCD calculations on a finite space-time lattice using the lattice QCD (LQCD) method [5], but such calculations are still too computationally expensive for general use in nuclear theory [6]. Nuclear forces arise from residual quark interactions in a manner analogous to the molecular Van der Waals force, and it is possible that this comparatively weak effect may be perturbatively analyzed by generalizing so-called chiral perturbation theory (χPT) [7] to generate potentials for $A \geq 2$ systems.

Effective field theories (EFTs) [4, 8, 9] aim to bridge the gap between phenomenological models and costly LQCD calculations. The relevant degrees of freedom in nuclear EFTs are nucleons and (usually) pions, as opposed to quarks and gluons in QCD. Weinberg [10–12] proposed to construct a nuclear interaction potential from the most general effective Lagrangian $\mathcal{L}_{\text{eff}}$ that observes the symmetries of QCD, notably approximate chiral symmetry, and expanding $\mathcal{L}_{\text{eff}}$ in powers of external momenta p over the breakdown scale Λ_b of the theory. The expansion yields an infinite number of terms containing Feynman diagrams encoding different types of nucleon and pion interactions. For practical calculations, the expansion must be truncated at a finite chiral order, and it is therefore important to employ a suitable power counting scheme wherein the diagrams are ordered according to the importance of their contributions. The power counting scheme assigns an order $k \geq 0$ to each interaction: the dominant interactions comprise the leading order (LO), and higher-order corrections—suppressed by successively higher powers of the expansion parameter $Q = p/\Lambda_b$—are sorted as next-to-leading order (NLO), next-to-next-to-leading order (NNLO), and so on. This means that the magnitude of the error incurred by the truncation of the series is controlled, enabling us to extract reliable uncertainty estimates.

Figure 1.1 shows a hierarchy of nuclear forces in the most widely used chiral EFT (χEFT) arranged in a power counting scheme known as Weinberg power counting. At LO ($k = 0$), one 2N contact interaction and one leading pion-nucleon (πN) interaction appear: the short-range 2N contact interaction acts with orbital angular momentum $L = 0$ (S-waves) indicated by crossed solid lines, and the long-range one-pion exchange (1PE) indicated by a dashed line. In this scheme, no interactions contribute at order $k = 1$ [13], and consequently the order $k = 2$ is designated as NLO. The interactions at NLO include derivative 2N contacts acting in $L = 1$ (P-wave) and two-pion exchanges (2PE). At NNLO, three-nucleon (3N) interactions $\hat{V}_{\text{3N}}$ enter.

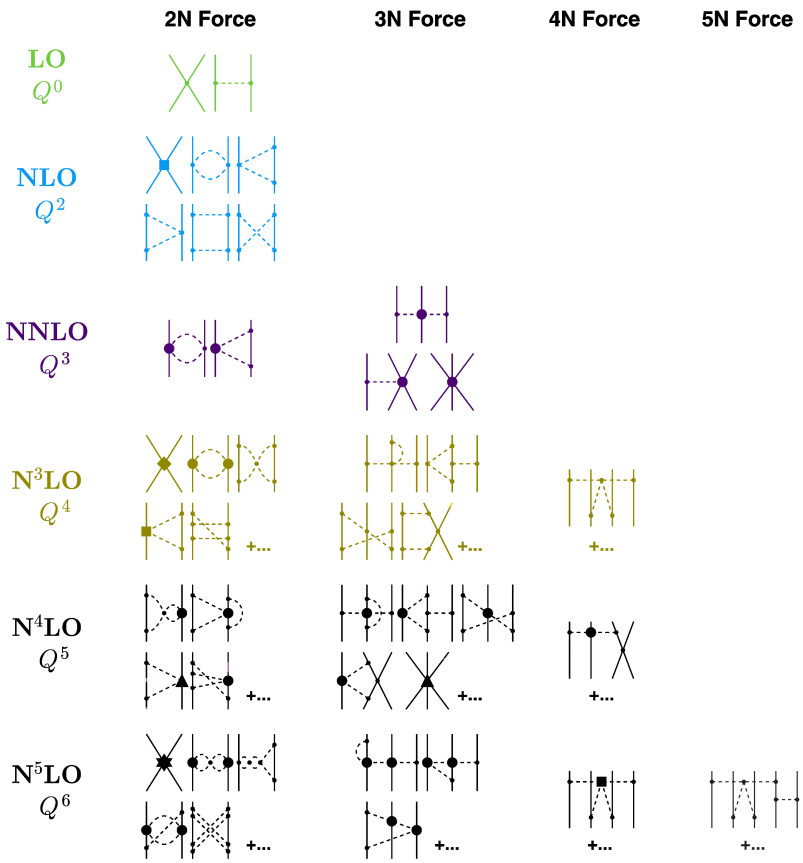

Figure 1.1: Hierarchy of nuclear forces in χEFT according to Weinberg power counting. Solid lines represent nucleons, and dashed lines represent pions. Figure adapted from Entem et al. [14].

χEFT, sometimes referred to as pionful EFT, employs nucleons and pions as effective degrees of freedom and is well suited for analyzing physics characterized by momenta comparable to the pion mass ($m_\pi \approx 140$ MeV) [9]. An exact determination of the location of the breakdown scale Λ_b has yet to be achieved. Nevertheless, χEFT has in many cases [2] proven to provide a successful model of the strong nuclear interaction, yet outstanding issues regarding appropriate power counting remain [15, 16]. I use the standard Weinberg power counting in this work.

Each new order in the χEFT expansion of the effective Lagrangian (from now on referred to as the chiral expansion) introduces a number of unknown parameters known as a low-energy constants (LECs) that govern the strength of the corresponding interactions. The LECs are thus divided into categories based on the type of interaction, e.g. "contact LECs" for short-range contact interactions and πN LECs for long-ranged π-exchanges. Power counting places an important expectation on the LECs in that they are supposed to be of "natural" size (i.e. of order 1) in appropriate units; a failure to fulfil this expectation is an indication that the EFT is not working as advertised. The LECs have to be inferred from data (usually experimental, but LQCD results can in principle be used). The LECs are consequently imbued with an uncertainty which will propagate to observable predictions. In addition there are truncation and numerical errors from the solution of the Schrödinger equation. Quantifying these uncertainties and their impact on predictions is therefore of fundamental importance for precision nuclear physics, and the fundamental motivation for the present work.

Historically, the most prevalent approach to LEC inference has been limited to maximum likelihood estimation of LEC values to reproduce experimental results with high accuracy, as was done in e.g. Ref. [17]. Subsequent efforts extended this approach within the frequentist statistical framework by extracting local estimates of the LEC covariances [18, 19]. In recent years, an alternative approach based on Bayesian inference methods has been developed as a tool for comprehensive error analysis in χEFT (see e.g. Refs. [20–25]). The Bayesian interpretation of probability enables us to infer a probability density function (pdf) for the LECs at a given chiral order. This pdf, known as a posterior in Bayesian lingo, can be conditional on a range of given information, such as the empirical data used for the parameter estimation and a statistical model for the EFT truncation error [21]. The Bayesian framework also allows (in fact, requires) us to encode a priori beliefs or assumptions in the so-called prior pdf. This gives us a straightforward way to incorporate for example the naturalness expectation of the LECs. The uncertainties encoded

in the LEC posteriors can be propagated to observable predictions to yield posterior predictive distributions (ppds) with an added model discrepancy term [26] for the energy-dependent EFT truncation error.

The number of LECs in realistic chiral potentials at order NLO and beyond are at least 10-15, meaning that the corresponding LEC posteriors have at least that many dimensions. Evaluating such multidimensional pdfs poses a significant computational challenge, and the only realistic strategy is to use Markov chain Monte Carlo (MCMC) sampling. However, standard MCMC methods typically yield strongly correlated samples from the posterior pdf when applied to high-dimensional problems. Sample correlations decrease the information contained in the MCMC chain, and the sampler must collect a correspondingly large number of samples to compensate. Minimizing sample correlations by advanced sampling methods can therefore be highly advantageous. One such method is Hamiltonian Monte Carlo (HMC) [27], originally developed by LQCD researchers [28].

In this thesis I use Bayesian analysis to define LEC posterior pdfs encoding parametric and epistemic uncertainties in χEFT at up to NNLO. I sample the relevant posteriors by developing a custom HMC implementation, and investigate its performance. I extract ppds of 2N elastic scattering cross sections and evaluate the reliability of the statistical model by comparisons with a large set of validation data reserved for this purpose.

Chapter 2

A statistical model for uncertainties in χEFT

The central theme of this thesis is to quantify uncertainties in χEFT predictions of nuclear observables—primarily 2N scattering observables—using Bayesian analysis. To achieve this goal it is necessary to adopt statistical models for the error terms. Several sources of uncertainty contribute to the total uncertainty of χEFT predictions; in this work I will make the assumption that the dominant contributions stem from (1) the parametric uncertainty of the LECs and (2) the truncation of the infinite chiral expansion at a finite order k. Further possible sources of uncertainty, which I will take to be negligible, include a numerical error when solving the Lippmann-Schwinger equation [29] by the standard matrix inversion method [30] used here. I will consider the matrix inversion error negligible in this thesis.

The objectives of this chapter are twofold: define posterior pdfs for the LECs that can be numerically evaluated to facilitate parameter estimation, and outline how to exploit those pdfs to define ppds for arbitrary 2N scattering observables. The ppds will combine the two dominant sources of uncertainty outlined in Chapter 1.

2.1 Calibration and validation data

Empirical data plays a crucial role for both parameter estimation and the subsequent validation of the statistical model. I will therefore begin this chapter by specifying what data I use.

I define two mutually exclusive sets of data. The first, called the calibration (or training) data $\mathcal{D}_{\text{cal}}$, is used for the LEC inference. The second, called the validation data $\mathcal{D}_{\text{val}}$, is used to verify the validity of the statistical model. The calibration data $\mathcal{D}_{\text{cal}}$ is slightly different between Papers A and B, and $\mathcal{D}_{\text{val}}$ is only used in Paper B. In both papers I use the 2013 Granada 2N scattering database [31, 32] which contains neutron-proton (np) and proton-proton (pp) elastic scattering cross sections gathered from a long list of scattering experiments spanning many decades, with some dating back to the 1950s. While the database covers laboratory scattering energies T_{lab} from 0 up to 350 MeV, I only use data below the pion production threshold (i.e. $T_{\text{lab}} \leq 290$ MeV) for the LEC inference. In Paper A all data in this range is used for inference, but in Paper B I withhold all data in the $80 \leq T_{\text{lab}} \leq 100$ MeV range for validation purposes in addition to one set [33] of total np cross sections covering the $33 \leq T_{\text{lab}} \leq 350$ MeV range. The validation data set $\mathcal{D}_{\text{val}}$ also contains all available data above the pion production threshold. In all, $\mathcal{D}_{\text{cal}}$ ($\mathcal{D}_{\text{val}}$) comprises 4366 (2018) data points in Paper B. This leaves ample data for both parameter inference and model checking.

A comprehensive summary of the distribution of data is shown in Table 2.1. The observables are denoted using the SAID convention [34], where e.g. total cross sections are denoted as SGT and unpolarized differential cross sections are denoted as DSG. When specificity is required I prepend "NP" or "PP" to indicate np or pp scattering, e.g. a total np cross section is denoted NPSGT. Total and differential cross sections comprise nearly half of the Granada database, and the remaining data are various types of spin observables. A significant portion of the spin observables correspond to the expectation values of the final spins along the direction normal to the scattering plane (polarization, denoted P) and the polarization of the incoming beam (denoted PB). Other prevalent spin observables correspond to the correlations betweeen incoming and outgoing spins along various combinations of axes (e.g. AXX, AYY).

Other types of data could straightforwardly be included in the calibration data. A prime candidate would be πN scattering data, but I have opted to let a prior that results from a separate data analysis constrain the πN LECs entering at NNLO (see Section 2.3.3 for details). Additional data types include nuclear binding energies and/or radii from e.g. light nuclei like ^{2}H, ^{3}H, or ^{3}He [35], or slightly heavier nuclei such as ^{16}O [36]. In principle, any low-energy observable may be used.

It is common for an experimental data set to be reported with a normalization constant which accounts for a common systematic uncertainty [32].

Table 2.1: The distribution of observables in the calibration ($\mathcal{D}_{\text{cal}}$) and validation ($\mathcal{D}_{\text{val}}$) data sets in Paper B. $\mathcal{D}_{\text{val}}$ is composed of all available data in the $80 \leq T_{\text{lab}} \leq 100$ and $290 < T_{\text{lab}} \leq 350$ MeV ranges, plus one set of NPSGT data [33] covering the $33 \leq T_{\text{lab}} \leq 350$ MeV range. All other available data constitutes $\mathcal{D}_{\text{cal}}$. See text for further details.

	Training data ($\mathcal{D}_{\text{cal}}$)			Validation data ($\mathcal{D}_{\text{val}}$)		
Obs	*np*	*pp*	Total (%)	*np*	*pp*	Total (%)
SGT	315	0	315 (7.2)	84	0	84 (4.2)
SGTL	11	0	11 (0.3)	4	0	4 (0.2)
SGTT	16	0	16 (0.4)	3	0	3 (0.1)
DSG	1221	756	1977 (45.2)	457	159	616 (30.5)
A	5	47	52 (1.2)	0	24	24 (1.2)
AP	0	5	5 (0.1)	0	0	0 (0.0)
AT	30	0	30 (0.7)	35	0	35 (1.7)
AXX	0	143	143 (3.2)	0	120	120 (5.9)
AYY	64	151	215 (4.8)	46	151	197 (9.8)
AZX	0	137	137 (3.1)	0	120	120 (5.9)
AZZ	45	39	84 (1.9)	27	10	37 (1.8)
CKP	0	1	1 (0.0)	0	1	1 (0.0)
D	13	56	69 (1.6)	14	42	56 (2.8)
D0SK	8	0	8 (0.2)	14	0	14 (0.7)
DT	39	0	39 (0.9)	39	0	39 (1.9)
MSKN	0	8	8 (0.2)	0	8	8 (0.4)
MSSN	0	8	8 (0.2)	0	8	8 (0.4)
NNKK	8	0	8 (0.2)	0	0	0 (0.0)
NSKN	12	0	12 (0.3)	13	0	13 (0.6)
NSSN	4	0	4 (0.1)	14	0	14 (0.7)
P	0	489	489 (11.2)	0	260	260 (12.9)
PB	590	0	590 (13.5)	253	0	253 (12.5)
PT	38	0	38 (0.9)	19	0	19 (0.9)
R	5	50	55 (1.3)	0	50	50 (2.5)
RP	0	22	22 (0.5)	0	5	5 (0.2)
RPT	1	0	1 (0.0)	1	0	1 (0.0)
RT	29	0	29 (0.7)	37	0	37 (1.8)
All	2454	1912	4366 (100.0)	1060	958	2018 (100.0)

Rather than attaching a corresponding normalization constant to the theoretically computed value and allowing it to vary, I fix the constants to the values reported in the Granada database. This approximation should not significantly affect the estimation of the LECs as it is unlikely that the normalization constants shift experimental results disproportionally in a certain direction when the database is viewed as a whole (see e.g. Ref. [37],) i.e. the floating normalization constants are assumed to be uncorrelated.

2.2 Linking truth, theory, and experiment

The true value $\mathcal{O}_{\text{true}}$ of an observable $\mathcal{O}$ may be regarded as the sum of a theoretical prediction $\mathcal{O}_{\text{theo}}$ and an unknown theoretical error $\delta\mathcal{O}_{\text{theo}}$, i.e.

$$\mathcal{O}_{\text{true}} = \mathcal{O}_{\text{theo}} + \delta\mathcal{O}_{\text{theo}}, \tag{2.1}$$

where in this work I make the assumption that $\delta\mathcal{O}_{\text{theo}}$ is dominated by the EFT truncation error. Further, an experimentally measured value $\mathcal{O}_{\text{exp}}$ of the same observable can be thought of as the sum of the observable's underlying true value and an experimental error:

$$\mathcal{O}_{\text{exp}} = \mathcal{O}_{\text{true}} + \delta\mathcal{O}_{\text{exp}}. \tag{2.2}$$

Substituting Equation (2.1) into Equation (2.2) yields

$$\mathcal{O}_{\text{exp}} = \mathcal{O}_{\text{theo}} + \delta\mathcal{O}_{\text{theo}} + \delta\mathcal{O}_{\text{exp}}, \tag{2.3}$$

thus forming a link between theory and experiment which I will use to define Bayesian pdfs for the LECs and ppds for observables.

2.3 Bayesian framework

Fundamental to all Bayesian statistics is Bayes' theorem (or Bayes' rule) which states

$$\text{pr}(A|B) = \frac{\text{pr}(B|A) \cdot \text{pr}(A)}{\text{pr}(B)}. \tag{2.4}$$

where A and B are propositions and $\text{pr}(B) \neq 0$. This simple formula is a direct consequence of the product rule in probability theory [38], yet it leads to interpretations of probability that are different in nature from the traditional interpretation of probability as a frequency of outcomes. Rather,

probability may be interpreted as the current state of knowledge and thus subject to change as more information becomes available. On a more prosaic note we may use Bayes' theorem to conveniently express a so-called posterior pdf for the χEFT LECs $\vec{\alpha}$ given some data D and information I as

$$\mathrm{pr}(\vec{\alpha}|D,I) = \frac{\mathrm{pr}(D|\vec{\alpha},I)\cdot\mathrm{pr}(\vec{\alpha}|I)}{\mathrm{pr}(D|I)}. \tag{2.5}$$

The information I includes all assumptions and knowledge on which the inference is based, e.g. a priori expectations for the parameters $\vec{\alpha}$, the model for the strong nuclear force, specific choices regarding model hyperparameters, and so on. The factors on the right hand side are the likelihood $\mathrm{pr}(D|\vec{\alpha},I)$, the prior $\mathrm{pr}(\vec{\alpha}|I)$, and the the marginal likelihood (or, succinctly, the evidence) $\mathrm{pr}(D|I)$, all of which are also pdfs. The marginal likelihood does not depend on $\vec{\alpha}$ and consequently does not influence the parameter estimation, and Equation (2.5) may therefore be conveniently written as

$$\mathrm{pr}(\vec{\alpha}|D,I) \propto \mathrm{pr}(D|\vec{\alpha},I)\cdot\mathrm{pr}(\vec{\alpha}|I). \tag{2.6}$$

Equation (2.6) provides the basis for the Bayesian parameter estimation of the LECs $\vec{\alpha}$. The posterior on the left hand side is the end product of the parameter estimation process. This multidimensional pdf encodes the full state of knowledge of the LECs conditional on D and I, and is of central importance for predicting nuclear observables.

2.3.1 Assigning a likelihood

The likelihood is central in both frequentist and Bayesian statistics. It measures the goodness of fit of a statistical model to the observed data. The traditional approach to fitting the LECs in χEFT is to define a χ^2 cost function [39]

$$\chi^2 = \sum_{i=1}^{N}\left(\frac{\mathcal{O}_{\mathrm{theo},i}(\vec{\alpha}) - \mathcal{O}_{\mathrm{exp},i}}{\sigma_{\mathrm{exp},i}}\right)^2, \tag{2.7}$$

where $\mathcal{O}_{\mathrm{exp},i}$ is the ith datum, $\mathcal{O}_{\mathrm{theo},i}(\vec{\alpha})$ is the corresponding computed observable value, their difference is called a residual, $\sigma_{\mathrm{exp},i}$ is the standard deviation of the experimental error, and N is the number of data points. A search for the LEC vector $\vec{\alpha}^*$ that minimizes this cost function is then performed, i.e.

$$\vec{\alpha}^* = \mathrm{argmin}(\chi^2) \tag{2.8}$$

This approach is equivalent to maximizing the χ^2 likelihood

$$\mathcal{L}_{\chi^2}(\vec{\alpha}) \propto \exp\left(-\frac{1}{2}\chi^2\right), \tag{2.9}$$

and is an example of maximum likelihood estimation (MLE). MLE is a staple of frequentist statistics and has been used in several studies to extract optimized LEC values $\vec{\alpha}^*$ (see e.g. Refs [17, 40, 41]) along with local approximations of LEC covariance matrices [18, 42] in order to propagate LEC uncertainties to predictions.

The likelihood (2.9) with χ^2 defined by Equation (2.7) assumes that the residuals are uncorrelated and follow a normal distribution. Furthermore, it assumes that the experimental error is the only source of uncertainty. We can introduce an independent theory error by replacing $\sigma_{\text{exp},i}$ in Equation (2.7) with σ_i, defined as [39]

$$\sigma_i^2 = \sigma_{\text{exp},i}^2 + \sigma_{\text{theo},i}^2 \tag{2.10}$$

where $\sigma_{\text{theo},i}$ is the standard deviation of an assigned theoretical (Gaussian-distributed) error. We can circumvent the restrictive assumption that the errors σ_i are uncorrelated by introducing a covariance matrix Σ, where

$$\Sigma = \Sigma_{\text{exp}} + \Sigma_{\text{theo}} \tag{2.11}$$

and fully uncorrelated errors correspond to the special case

$$\Sigma_{\text{exp}} + \Sigma_{\text{theo}} = \text{diag}(\sigma_{\text{exp},1}^2, \ldots, \sigma_{\text{exp},N}^2) + \text{diag}(\sigma_{\text{theo},1}^2, \ldots, \sigma_{\text{theo},N}^2). \tag{2.12}$$

By introducing the residual vector

$$\vec{r} = \vec{\mathcal{O}}_{\text{theo}} - \vec{\mathcal{O}}_{\text{exp}} \tag{2.13}$$

the corresponding likelihood can be written as [24]

$$\mathcal{L}(\vec{\alpha}) = \text{pr}(D|\vec{\alpha}, \Sigma_{\text{exp}}, \Sigma_{\text{theo}}, I) \propto \exp\left(-\frac{1}{2}\vec{r}^T \cdot \Sigma^{-1} \cdot \vec{r}\right) \tag{2.14}$$

The question is: how do we assign Σ_{theo}? In this thesis I follow the path laid down in Refs. [21, 23] and, in particular, Ref. [24]. As a starting point, the EFT expansion parameter Q introduced in Chapter 1 is written as

$$Q = \frac{\max(m_\pi, p)}{\Lambda_b} \tag{2.15}$$

where m_π is the pion mass, p is the center-of-mass momentum, and Λ_b is the EFT breakdown scale. In this work I fix $\Lambda_b = 600$ MeV as was done by Wesolowski et al. [24] and has been indicated as a reasonable value in a previous study [23]. The breakdown scale is generally expected to be of the order of the rho meson mass, $m_\rho \approx 775$ MeV. By introducing a dimensionful reference scale $\mathcal{O}_{\text{ref}}$ and dimensionless expansion coefficients c_n the infinite χEFT series for an observable may be rewritten as

$$\widetilde{\mathcal{O}} = \mathcal{O}_{\text{ref}} \sum_{n=0}^{\infty} c_n Q^n. \tag{2.16}$$

This series can then be split according to

$$\widetilde{\mathcal{O}} = \mathcal{O}^{(k)}_{\text{theo}} + \delta\mathcal{O}^{(k)}_{\text{theo}} = \mathcal{O}_{\text{ref}} \sum_{n=0}^{k} c_n Q^n + \mathcal{O}_{\text{ref}} \sum_{n=k+1}^{\infty} c_n Q^n \tag{2.17}$$

where the index k indicates the chiral order at which $\mathcal{O}_{\text{theo}}$ is computed, and $\delta\mathcal{O}^{(k)}_{\text{theo}}$ is the truncation error.

Several options exist for the choice of reference scale $\mathcal{O}_{\text{ref}}$, such as using experimental values or values computed using some simpler model. In this work I use values computed at LO, with preliminary LEC values $\vec{\alpha}_{\text{prel}}$ obtained from an MLE fit to phase shifts, for SGT and DSG observables. Melendez et al. argue in Ref. [25] that it is appropriate to set the reference value for spin observables, whose values are in the range $[-1, 1]$, to 1 in order to avoid potential issues when the value is close to 0. This procedure is not ideal in my experience as a value of 1 is an extreme value for a spin observable. I instead compute an average value of the LO results based on $\vec{\alpha}_{\text{prel}}$ and use this value ($\overline{\mathcal{O}}^{\text{spin}}_{\text{LO}} = 0.15$) as the reference scale.

The last piece of information needed to characterize the truncation error $\delta\mathcal{O}^{(k)}_{\text{theo}}$ are the values of the expansion coefficients c (where I drop the subscript n to indicate that I am not specifying the chiral order). Importantly, the c's are expected to be of order 1, as each new term in the EFT expansion should be suppressed by approximately a factor Q. As noted in Chapter 1, the contribution at order $n = 1$ is zero in Weinberg power counting, so $c_1 \equiv 0$. The coefficients at order $n \neq 1$ may be estimated from consecutive orders as e.g.

$$c_n = \frac{\mathcal{O}^{(n)}_{\text{theo}}(\vec{\alpha}_{\text{prel}}) - \mathcal{O}^{(n-1)}_{\text{theo}}(\vec{\alpha}_{\text{prel}})}{\mathcal{O}_{\text{ref}} \cdot Q^n}, \tag{2.18}$$

where the preliminary LEC values $\vec{\alpha}_{\text{prel}}$ are employed to compute approximate observable values. However, the theoretical values at $n > k$ are typically not

available, as k is the order at which the prediction is made; if a higher order was available, we should use that order to make the prediction. It is thus necessary to estimate the characteristic size of c_n for $n > k$. All expansion coefficients are therefore assumed to be drawn from the same underlying normal distribution,

$$\mathrm{pr}(c|\bar{c}) = \mathcal{N}(0, \bar{c}^2), \tag{2.19}$$

where $\bar{c}$ is the typical size of the expansion coefficients (i.e. we expect $\bar{c} \approx 1$). Leaving $\bar{c}$ undetermined for the moment, Equation (2.19) yields (see Appendix B)

$$\mathrm{pr}(\delta\mathcal{O}^{(k)}_{\mathrm{theo}}|\bar{c}, Q) = \mathcal{N}\left(0, (\sigma^{(k)}_{\mathrm{theo}})^2\right) \tag{2.20}$$

where

$$(\sigma^{(k)}_{\mathrm{theo}})^2 = \bar{c}^2\mathcal{O}^2_{\mathrm{ref}}\frac{Q^{2(k+1)}}{1-Q^2}. \tag{2.21}$$

Under the simplifying assumption that the c's are completely uncorrelated the theory covariance matrix Σ_{theo} becomes

$$(\Sigma_{\mathrm{theo}})_{ij} = \bar{c}^2(\mathcal{O}^{(i)}_{\mathrm{ref}})^2\frac{Q_i^{2(k+1)}}{1-Q_i^2}\delta_{ij}. \tag{2.22}$$

I now turn my attention to the assignment of $\bar{c}$. An example of computed expansion coefficients $\vec{c}_3$, i.e. at order $n = 3$ (NNLO), for the np differential cross section is shown in Figure 2.1. The displayed behavior of these particular expansion coefficients is fairly representative of the disposition of the c_n's in general. The first thing to note is that, encouragingly, the c's are indeed of order 1. Of further note is that the values of c show a clear, but complicated, correlation structure. In a recent study [25], this correlation structure was modelled using Gaussian processes. In this thesis I have adopted the method described in Ref. [24]: I estimate the standard deviation of expansion coefficients by calculating the root-mean-square (RMS) value of a set of computed c_n values. I use a relatively sparse grid of energies T_{lab} and scattering angles θ to minimize the influence of correlations, and work with the four observables (SGT, DSG, P, and PB) which together account for the majority of the scattering data in the database (see Section 2.1). The grid, indicated in Figure 2.1 by the displayed energies and the two vertical dashed lines indicating the angles, is $T_{\mathrm{lab}} = 20, 70, 120, 170, 220, 270$ MeV and $\theta = 50°, 150°$. The RMS value $\bar{c}$ is computed using all orders up to order k, i.e.

$$\bar{c}(k) = \sqrt{\frac{1}{N_{\mathrm{obs}}N_{\mathrm{ord}}}\sum_{n=0}^{k}\sum_{i=0}^{N_{\mathrm{obs}}}c_{n,i}^2}, \quad n \neq 1, \tag{2.23}$$

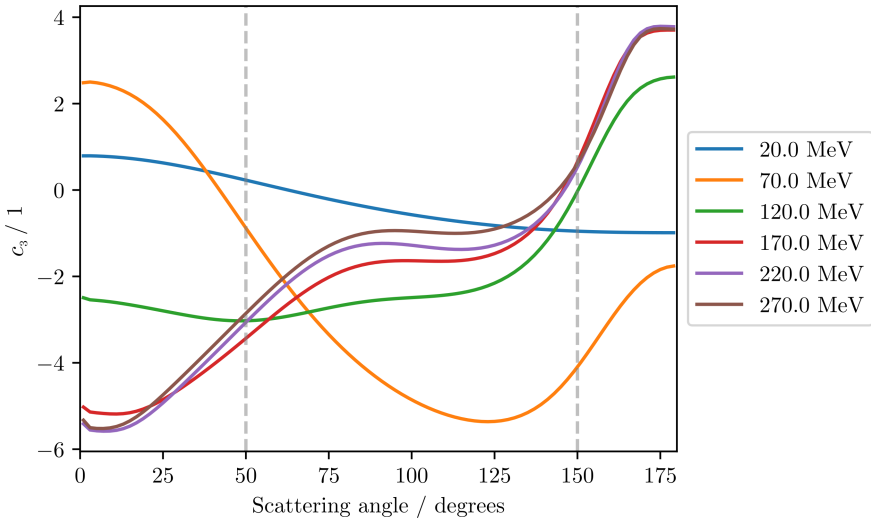

Figure 2.1: Computed expansion coefficients $\vec{c}_3$ for the np differential cross section at NNLO. Each curve was calculated at the laboratory energy T_{lab} indicated in the legend. The dashed vertical lines indicate the scattering angles θ used in the assignment of $\bar{c}$.

where N_{obs} is the number of different observables (defined by the grid outlined above) and N_{ord} is the number of chiral orders up to and including order k (e.g. $k = 2$ and $N_{\text{ord}} = 2$ at NLO). Before I apply Equation (2.23) I remove outlier c values which would otherwise disproportionally influence the estimation of $\bar{c}$; I define outlier c values as those that lie more than three times the inter-quartile range outside the upper or lower quartiles. As we will see in Chapter 5, the inclusion of the LO expansion coefficients $\vec{c}_0$ turned out to be inadvisable as these are uninformed by order-by-order differences and in our case underestimate $\bar{c}$. I will therefore present results based on alternative prescriptions for $\bar{c}$ as well. The computed values of $\bar{c}$ are presented in Table 2.2.

2.3.2 A prior for the contact LECs

The prior we place on the 2N contact LECs is a simple multivariate Gaussian with independent variables of standard deviation $\bar{\alpha} = 5$. This encodes that

Table 2.2: Results from the $\bar{c}$ analysis. $\bar{c}_n$ denotes the root-mean-squared value of the expansion coefficients at a particular order n, while $\bar{c}(k)$ also includes lower order coefficients. The number of outliers compared to the total number of kinematic points is shown in the outlier columns; see text for how outliers are defined.

Order	k	$\bar{c}_n$	$\bar{c}(k)$	Outliers $\bar{c}(k)$
LO	0	1.17	1.17	1/54
NLO	2	4.95	2.08	15/108
NNLO	3	2.84	2.72	8/162

the LECs are expected to be of natural size and penalizes large (unnatural) LEC values. All LECs at LO and NLO are contact LECs, so the prior at these orders is

$$\mathrm{pr}(\vec{\alpha}|I) \propto \exp\left(-\frac{1}{2}\vec{\alpha}^T \cdot \Sigma_{\mathrm{prior}}^{-1} \cdot \vec{\alpha}\right) \tag{2.24}$$

where

$$(\Sigma_{\mathrm{prior}})_{ij} = (\Sigma_{\mathrm{prior,2N}})_{ij} = \bar{\alpha}^2 \delta_{ij} \tag{2.25}$$

As there are two relevant LECs at LO,

$$\vec{\alpha}_{\mathrm{LO}} = \left(\widetilde{C}_{1S0}, \widetilde{C}_{3S1}\right), \tag{2.26}$$

and ten at NLO,

$$\begin{aligned}\vec{\alpha}_{\mathrm{NLO}} = \Big(&\widetilde{C}_{1S0}^{\mathrm{np}}, \widetilde{C}_{1S0}^{\mathrm{pp}}, \widetilde{C}_{3S1}, C_{1S0}, C_{3P0}, \\ &C_{1P1}, C_{3P1}, C_{3S1}, C_{3S1-3D1}, C_{3P2}\Big),\end{aligned} \tag{2.27}$$

the priors at LO and NLO will be two- and ten-dimensional, respectively.

2.3.3 NNLO prior

The contact LECs at NNLO are the same as at NLO, as odd-k 2N contact terms are zero due to parity [18]. Three new πN LECs do enter at NNLO, though, called c_1, c_3, c_4.[1] These LECs govern the strength of πN interactions as well as the two-pion-exchange in 3N interactions [43]. In principle, πN scattering data could be included in the likelihood to aid this inference, but I instead exploit the power of the Bayesian framework and place a prior on

[1] Note that c_1, c_3, and c_4 are not the same as the expansion coefficients c_n introduced earlier. This clash of notation has unfortunately become standard in the field.

the πN LECs from a previous analysis. This prior, based on a so-called Roy-Steiner (R-S) analysis [44], imposes tighter constraints than plain scattering data would. The central values and covariance matrix are gathered from Ref. [45]. The full NNLO prior will thus be a 13-dimensional Gaussian with one block of the covariance matrix encoding tight constraints on the three πN LECs about the central values from the R-S analysis, and one block encoding the same permissive, uncorrelated prior from NLO for the contact LECs:

$$\mathrm{pr}(\vec{\alpha}|I) \propto \exp\left(-\frac{1}{2}(\vec{\alpha}-\vec{\mu})^T \cdot \Sigma_{\mathrm{prior}}^{-1} \cdot (\vec{\alpha}-\vec{\mu})\right) \tag{2.28}$$

$$\Sigma_{\mathrm{prior}} = \begin{bmatrix} \Sigma_{\mathrm{prior},\pi\mathrm{N}} & 0 \\ 0 & \Sigma_{\mathrm{prior,2N}} \end{bmatrix} \tag{2.29}$$

where $\vec{\mu}$ is a vector with the central values from the R-S analysis for the πN LECs and zeros for the contact LECs.

2.3.4 Beyond the 2N system

Venturing beyond two-nucleon interactions introduces at least two new LECs, called c_D and c_E, which govern the strength of leading contact 3N interactions. These parameters enter at NNLO in χEFT (assuming Weinberg power counting) and influences predictions of all atomic nuclei except the deuteron. Like any LEC, they are unknown and have to be inferred from data. Doing so with rigorous statistical constraints is the focus of Paper A. My main contribution to the study was to provide a tightly constrained Bayesian prior for the 2N contact LECs to facilitate MCMC sampling over the full posterior including the 3N LECs, thus handling the 2N LECs in a similar fashion to how I treat the πN LECs in Paper B.

The approach to obtain this pdf is to set up a likelihood similar to that described in Section 2.3.1 and perform a maximum likelihood estimation along with a local estimation of the LEC covariance matrix using second derivatives. The likelihood is essentially the same as in Equation (2.14) but with a couple of simplifications. Perhaps the most important simplification is that we used $\bar{c} = 1$ in the definition of the covariance matrix (2.22) for the truncation error. Next the πN LECs were fixed to the central values from the Roy-Steiner analysis and not allowed to vary. Finally we used $\mathcal{O}_{\mathrm{ref}} = \mathcal{O}_{\mathrm{exp}}$ rather than using LO calculated observable values. The charge-dependent LO contact LEC $\widetilde{C}_{1S0}^{nn}$ is also included. This LEC is unconstrained by 2N scattering data and an empirical $^1\mathrm{S}_0$ nn scattering length $a_{nn} = -18.95 \pm 0.40$ fm [46, 47]

and effective range $r_{nn} = 2.75 \pm 0.11$ fm [48] are included in the calibration data in order to constrain it.

The set of LECs $\vec{\alpha}^*_{\text{2N}}$ that maximizes the likelihood was found by minimizing the negative log likelihood in two stages. The first stage entailed performing a broad search of the parameter space using the first-order Levenberg-Marquardt optimization algorithm. From the resulting array of candidate optima I picked the one that resulted in the maximum likelihood value and used that as a starting point for a second-order Newton method, which resulted in a further increase of the likelihood value.

I extracted an approximate LEC covariance matrix using the method outlined in Ref. [18]. The method relies on access to second-order derivatives of the negative log likelihood. These derivatives were extracted to near machine precision using so-called automatic differentiation (AD) [49], which will also play a major role for the MCMC sampling detailed in the next chapter.

2.4 Predictive distributions

Bayesian posteriors of the form (2.6) may be used to propagate parameter uncertainties to predictions of observables, resulting in predictive distributions. Such predictive distributions can have either intrinsic or instrumental value, or both, depending on the observable in question. Intrinsic value emerges when the observable is difficult or impossible to empirically measure and theoretical predictions contribute information that is otherwise unattainable. An example of this is the proton-proton fusion cross section in the energy domain relevant to main sequence stars [50]. Instrumental value, on the other hand, enables us to check whether our model is useful by comparing model predictions with data. In this thesis my primary concern regarding predictive distributions is to check the consistency of model predictions with validation data.

I model the true value of a scattering observable as the sum of the predicted value at chiral order k and a corresponding truncation error according to

$$\widetilde{\mathcal{O}}^{(k)} = \mathcal{O}^{(k)}_{\text{theo}} + \delta\mathcal{O}^{(k)}_{\text{theo}}, \tag{2.30}$$

where I have added the index k to $\widetilde{\mathcal{O}}$ to indicate the order at which the prediction is made. Both terms on the RHS, and consequently $\widetilde{\mathcal{O}}$, are stochastic variables described by pdfs. Assuming that we can draw samples from the pdfs of both $\mathcal{O}^{(k)}_{\text{theo}}$ and $\delta\mathcal{O}^{(k)}_{\text{theo}}$ we can also sample the pdf of $\widetilde{\mathcal{O}}$, $\text{pr}(\widetilde{\mathcal{O}}|D, k, I)$, by adding the independent samples of $\mathcal{O}^{(k)}_{\text{theo}}$ and $\delta\mathcal{O}^{(k)}_{\text{theo}}$. To derive a pdf

$\mathrm{pr}(\mathcal{O}^{(k)}_{\mathrm{theo}}|D,I)$ for $\mathcal{O}^{(k)}_{\mathrm{theo}}$ we begin by marginalizing in the LECs $\vec{\alpha}$, i.e. [51]

$$\mathrm{pr}(\mathcal{O}^{(k)}_{\mathrm{theo}}|D,I) = \int \mathrm{pr}(\mathcal{O}^{(k)}_{\mathrm{theo}},\vec{\alpha}|D,I)d\vec{\alpha}. \tag{2.31}$$

Using the product rule we rewrite this equation as

$$\mathrm{pr}(\mathcal{O}^{(k)}_{\mathrm{theo}}|D,I) = \int \mathrm{pr}(\mathcal{O}^{(k)}_{\mathrm{theo}}|\vec{\alpha},D,I)\mathrm{pr}(\vec{\alpha}|D,I)d\vec{\alpha} \tag{2.32}$$

and then utilize the conditional independence between $\mathcal{O}^{(k)}_{\mathrm{theo}}$ and D given $\vec{\alpha}$ to arrive at

$$\mathrm{pr}(\mathcal{O}^{(k)}_{\mathrm{theo}}|D,I) = \int \mathrm{pr}(\mathcal{O}^{(k)}_{\mathrm{theo}}|\vec{\alpha},I)\mathrm{pr}(\vec{\alpha}|D,I)d\vec{\alpha}. \tag{2.33}$$

This expression is easy to sample assuming that we have an MCMC representation of the LEC posterior $\mathrm{pr}(\vec{\alpha}|D,I)$. All that is needed now to draw a sample from the pdf of the true value $\widetilde{\mathcal{O}}$ (2.30) is to add a sample of the truncation error. This is easy as we have an analytical expression for the truncation error pdf in Equation (2.20).

Chapter 3

Hamiltonian Monte Carlo

Straightforward function evaluation of multidimensional pdfs, such as the N-dimensional posterior $\text{pr}(\vec{\alpha}|D, I)$ in Equation (2.5) where $\vec{\alpha} = (\alpha_1, \ldots, \alpha_N)$, on a grid of $\vec{\alpha}$-values quickly becomes intractable as the number of parameters increases. This is because the number of needed gridpoints grows exponentially. Consider, for example, a 13-dimensional pdf like the NNLO posterior in this work. Even if we limit ourselves to just four evaluations in each direction, we end up with $4^{13} \approx 7 \cdot 10^7$ evaluations. It is furthermore difficult to select a suitable grid, and many of the selected points are hence likely to provide a negligible contribution to an expectation value $\langle f(\vec{\alpha}) \rangle$ with regards to the posterior for some function f of the LECs. What is required, then, is a method that

1. minimizes the number of posterior evaluations, and

2. samples the *relevant* locations of the parameter space.

These requirements are fulfilled by Markov chain Monte Carlo (MCMC) sampling, which samples the posterior by performing a guided random walk through the parameter space.

In this chapter I first present an overview of MCMC methods and common issues encountered by practitioners. I then introduce the Hamiltonian Monte Carlo (HMC) algorithm along with the most important details of the specific Python implementation I have developed. The package, christened `montepython`, is available at `https://github.com/svisak/montepython.git`. A practical overview and a toy example is presented in Appendix A.

3.1 The Metropolis-Hastings algorithm

The Metropolis-Hastings (M-H) algorithm [52, 53] samples a pdf $\text{pr}(\vec{\alpha})$ by producing a Markov chain whose stationary (or equilibrium) distribution is $\text{pr}(\vec{\alpha})$. Given a current sample $\vec{\alpha}$, the algorithm proposes a new sample $\vec{\alpha}'$ from a proposal distribution $q(\vec{\alpha}'|\vec{\alpha})$. The new sample is accepted with a probability a given by

$$a = \min(1, r) \tag{3.1}$$

where r is the Hastings ratio

$$r = \frac{\text{pr}(\vec{\alpha}')q(\vec{\alpha}'|\vec{\alpha})}{\text{pr}(\vec{\alpha})q(\vec{\alpha}|\vec{\alpha}')} \tag{3.2}$$

The next sample in the chain will be $\vec{\alpha}'$ if the update is accepted, or $\vec{\alpha}$ if it is rejected. The proposal distribution $q(\vec{\alpha}'|\vec{\alpha})$ is typically a normal distribution $\mathcal{N}(\vec{\alpha}, \text{diag}(\sigma^2))$ with some step size σ, but many other options exist. The step size σ needs to be tuned so that the acceptance rate $\bar{a}$, i.e. the fraction of accepted proposed samples, is around 0.25 [54] in order to explore the parameter space efficiently. The M-H algorithm is a so-called single-walker algorithm, i.e. the algorithm is only aware of a single location in the parameter space at any given time.

3.2 The affine invariant ensemble sampler

Software packages implementing improved variants the original M-H MCMC algorithm are currently the go-to tools for sampling parameter posteriors in χEFT. A significant improvement to the original M-H algorithm called the affine invariant ensemble sampler was introduced in 2010 [55]. A Python implementation of this algorithm, known as `emcee` [56], is a workhorse of modern Bayesian inference, in particular in the astrophysics community. The primary reasons for its popularity are its high performance and easy tuning. Only 1-2 hyperparameters need to be tuned, and the algorithm usually performs well without careful tuning. In this thesis I will usually use the name `emcee` to denote both the general algorithm and the specific implementation. `emcee`, in contrast to M-H, relies on multiple walkers and is thereby aware of multiple locations of the parameter space at once. This ensemble of walkers is evolved by using a proposal distribution for every walker based on the positions of the other walkers in the ensemble.

To sample an N-dimensional pdf $\text{pr}(\vec{\alpha})$ using `emcee`, the user first chooses the number of walkers n_w to use (which should ideally be high, $n_w \sim 100$ [56])

and initializes them to some small volume in the parameter space. The algorithm then updates the position $\vec{\alpha}_k$ of each walker k in sequence. The update is performed by randomly choosing a walker $j \neq k$, drawing a random variable Z from a distribution $g(z|b)$ defined as

$$g(z|b) \propto \begin{cases} \frac{1}{\sqrt{z}} & \text{if } z \in \left[\frac{1}{b}, b\right] \\ 0 & \text{otherwise} \end{cases} \tag{3.3}$$

where b is a tunable hyperparameter, and calculating

$$\vec{\alpha}'_k = \vec{\alpha}_j + Z(\vec{\alpha}_k - \vec{\alpha}_j) \tag{3.4}$$

where $\vec{\alpha}'_k$ is the proposed next position for walker k [56]. The new position $\vec{\alpha}'_k$ is then accepted with probability a calculated using a criterion similar to the Metropolis rejection criterion (3.1):

$$a = \min\left(1, Z^{N-1}\frac{\mathrm{pr}(\vec{\alpha}'_k)}{\mathrm{pr}(\vec{\alpha}_k)}\right). \tag{3.5}$$

A strong point of this algorithm is hinted at by the name *affine invariant* ensemble sampler: the algorithm is insensitive to affine transformations of the target pdf, i.e. it performs well even if the target pdf is highly skewed by correlations between the sampled parameters. Most MCMC methods would need tuning to perform well for such distributions.

3.3 Practical MCMC challenges

The primary drawback of the M-H algorithm, and most of its derivatives, is that the samples it produces are strongly correlated in most practical applications. Correlations between samples decrease the amount of information contained in the MCMC chain, requiring longer chains than would otherwise be necessary. The problem is generally exacerbated when the dimensionality of the sampled distribution $\mathrm{pr}(\vec{\alpha})$ increases.

To see why correlated MCMC samples are problematic it is instructive to first consider a situation where the samples are uncorrelated. According to the central limit theorem, the sampling variance of a parameter expectation value $\langle\alpha_i\rangle$ is given by

$$\mathrm{Var}\left[\langle\alpha_i\rangle\right] = \frac{\mathrm{Var}\left[\alpha_i\right]}{N} \tag{3.6}$$

where N is the length of the MCMC chain. If instead the samples of α_i are correlated, it can be shown that the sampling variance is given by [57]

$$\mathrm{Var}\left[\langle\alpha_i\rangle\right] = \tau\frac{\mathrm{Var}\left[\alpha_i\right]}{N} \tag{3.7}$$

where in the limit $N \to \infty$

$$\tau = 1 + 2\sum_{h=1}^{N}\rho_i(h). \tag{3.8}$$

Here $\rho_i(h)$ is the so-called autocorrelation function and measures the correlation between samples separated by h samples. The variable h is usually called the lag, and Equation (3.6) is a special case of Equation (3.7) with $\rho_i(h) = 0$ for all lags h.

τ is known as the integrated autocorrelation time and is greater than 1 in most practical MCMC applications. This increases the sampling variance and the MCMC practitioner must increase N to compensate for the larger sampling error, thus motivating us to define the effective sample size (ESS) as

$$\mathrm{ESS} = \frac{N}{\tau}. \tag{3.9}$$

A large value for τ thus greatly increases the required length of the MCMC chain, usually at great computational cost. Designing MCMC algorithms that yield small τ values is therefore highly desirable, even if the cost of each sample rises dramatically. Consider, for example two MCMC samplers A and B. Sampler A yields $\tau_A \approx 100$ for a particular problem while Sampler B yields $\tau_B \approx 1$, and the computational cost per sample is C_A and C_B, respectively. As long as $C_B < 100C_A$, Sampler B will be more efficient.

Of note is that the definitions of τ and ESS allow for the tantalizing possibility of effective sample sizes that are greater than the actual number of samples N if $\tau < 1$, which can occur if the samples are anticorrelated. In such a situation, sometimes known as antithetical sampling [58], Equation (3.9) becomes an advantage rather than a burden. This effect will become manifest in Chapter 4. It is important to keep in mind that the integrated autocorrelation time will in general be different for each expectation value one computes using the MCMC chain, i.e. one should not only compute τ for $\mathrm{Var}\left[\langle\alpha_i\rangle\right]$ but also for $\mathrm{Var}\left[\langle f(\vec{\alpha})\rangle\right]$ where f is some function of $\vec{\alpha}$. Some care is required for computing τ to avoid a signal-to-noise problem [57]; we use a Fourier method implemented in `emcee`.

Most MCMC algorithms have sampling hyperparameters that can be tuned to each specific problem in order to increase their sampling efficiency. For example, the proposal distribution in the M-H algorithm is typically a normal distribution of the same dimensionality as the target distribution. The proposal's covariance matrix is in the simplest case an identity matrix multiplied by a scalar, but performance gains can usually be achieved by using a diagonal matrix with different step sizes for each parameter, or in the most general case a matrix with off-diagonal elements to account for correlations between the target parameters. In the latter case there are thus $O(n^2)$ tunable hyperparameters for an n-dimensional problem, and for maximum performance each of these need to be tuned correctly. Regardless of tuning, I find that both the M-H algorithm and `emcee` produce chains with strong autocorrelations when employed to sample LEC posteriors (see also Chapter 4).

Another practical challenge is multimodal target distributions. In infinite time all properly constructed MCMC algorithms will explore the entire parameter space and converge to the true target distribution, but in practice MCMC will frequently fail to move between highly probable areas that are separated by areas of very low probability density. This is a formidable problem without easy solutions, but methods to alleviate it exist. One example is parallel tempering methods [59, 60], wherein the target distribution is flattened to make mode transitions more probable. A special sampling algorithm known as nested sampling [61] can also be employed to deal with multimodal distributions. Nested sampling is actually designed for computing the Bayesian evidence (the denominator in Bayes' theorem (2.5), useful for model selection [62]) but yields posterior samples as a by-product and is inherently well-suited to deal with multimodality.

As mentioned in the introduction to this chapter, the motivation for MCMC is to combat the curse of dimensionality which very quickly makes function evaluation on a grid intractable. Unfortunately, MCMC also suffers from the same curse, albeit to a lesser degree. To see why, it is important to realize that the M-H algorithm performs a (guided) random walk through the parameter space. Each new proposed sample is drawn by taking a random step from the current sample. As the dimensionality of the problem increases, the fraction of interesting (i.e. probable) space decreases compared to the total parameter space. Obviously, walking randomly through a parameter space of mostly uninteresting regions generally results in proposed samples that are unlikely to be accepted, thus causing low acceptance rates $\bar{a}$. The step size must be decreased to combat this behavior and keep $\bar{a}$ reasonable, but reducing the step size unfortunately increases the autocorrelation. The end result is a small

effective sample size and long runtimes. All MCMC algorithms suffer from this problem to a greater or lesser extent, but one algorithm—Hamiltonian Monte Carlo—is particularly adept at minimizing random walk behavior and circumventing the curse of dimensionality.

3.4 Hamiltonian Monte Carlo

In 1987, physicists working with lattice QCD simulations introduced a new MCMC algorithm based on Hamiltonian dynamics to combat the problems of correlated samples and poor scaling with increased target dimensionality [28]. They dubbed the algorithm Hybrid Monte Carlo but it has subsequently become known as Hamiltonian Monte Carlo (HMC) [27].

HMC is based on the Metropolis-Hastings algorithm, but the method for proposing a new sample is radically modified. Rather than randomly drawing a perturbation of the current sample, the pdf to be sampled is treated as a potential energy surface. The current sample of the chain is regarded as the location of a particle, which is given a momentum that is randomly drawn from a predefined distribution. The algorithm then simulates the motion of the particle by solving the Hamiltonian dynamics for a finite length of time, and then computes the Hastings ratio for the joint pdf of the final position and momentum variables and accepts or rejects the proposed sample in accordance with Equation (3.1). The momentum variable is then discarded. The proposed sample is likely to be accepted even though the distance from the current sample is great. With suitable choices of simulation length and other tuning parameters, the result is a new sample of the target pdf that is uncorrelated with the previous sample. Due to this, the length of the MCMC chain produced by HMC can be drastically reduced compared to chains with a high degree of correlation as less correlated samples are equivalent to a shorter integrated autocorrelation time and reduced sampling error. The strongest benefit of using HMC is that the algorithm is capable of producing uncorrelated samples even when the target pdf is high-dimensional, as the Hamiltonian dynamics make it less susceptible to the curse of dimensionality than other MCMC algorithms.

Simulating Hamiltonian dynamics is computationally expensive compared to the method of randomly proposing new samples used in M-H, and the reduction in overall chain length must be sufficiently large to warrant the increased cost per sample. In particular, the algorithm requires multiple evaluations of the target pdf's gradient $\vec{\nabla}\mathrm{pr}(\vec{\alpha})$ in the course of proposing new samples. A mere lack of access to the gradient information may be a showstopper for

HMC depending on the implementation of the target pdf, for example if the pdf is evaluated using third-party software that does not provide gradient information. Evaluating the gradient is ideally done through AD unless an analytical expression for the gradient is available. AD generally incurs a factor ~ 2 overhead compared to just computing the value $\mathrm{pr}(\vec{\alpha})$ of the target pdf [63]. I employ AD to perform the HMC samplings in this work.

In the following subsections I present the HMC algorithm in detail. For the purpose of the analysis presented in this thesis, I have written a custom HMC implementation (`montepython`) using Python [64] and NumPy [65] in lieu of using a standard package such as Stan [66]. Specific implementation choices will be presented as appropriate. The implementation presented here is largely based on Ref. [27]. A conceptual introduction to HMC can be found in Ref. [67].

3.4.1 From probability to energy

The link between an energy function $E(\cdot)$ and a pdf $\mathrm{pr}(\cdot)$ is provided by the canonical (Boltzmann) distribution from statistical mechanics,

$$\mathrm{pr}(\cdot) = \frac{1}{Z} \exp\left(\frac{-E(\cdot)}{T}\right), \tag{3.10}$$

where Z is a normalization constant and T is the temperature of the system. The energy function E will in the case of HMC be the Hamiltonian H (note that H is a technical MCMC construct that is physically unrelated to any quantum mechanical Hamiltonian). H is the sum of the potential and kinetic energies,

$$H(\vec{\alpha}, \vec{p}) = U(\vec{\alpha}) + K(\vec{p}), \tag{3.11}$$

where $\vec{\alpha}$—the parameters whose pdf we wish to sample—can be thought of as the position of the MCMC particle and $\vec{p}$ its momentum. The normalization constant Z can be set to any convenient positive constant as it is independent from $\vec{\alpha}$ and $\vec{p}$ and does not impact the parameter estimation, and it will later be cancelled out in the Metropolis rejection criterion anyway (see Equation (3.23)). In the following, we let $Z = 1$. The joint pdf of $\vec{\alpha}$ and $\vec{p}$, $\mathrm{pr}(\vec{\alpha}, \vec{p})$, can thus be expressed as

$$\mathrm{pr}(\vec{\alpha}, \vec{p}) = \exp\left(\frac{-H(\vec{\alpha}, \vec{p})}{T}\right) \tag{3.12}$$

$$= \exp\left(\frac{-U(\vec{\alpha})}{T}\right) \exp\left(\frac{-K(\vec{p})}{T}\right). \tag{3.13}$$

The joint pdf $\text{pr}(\vec{\alpha}, \vec{p})$ can be regarded as the product of two canonical distributions, one for $\vec{\alpha}$ and one for $\vec{p}$, and we identify the form of $U(\vec{\alpha})$ as

$$U(\vec{\alpha}) = -\ln \text{pr}(\vec{\alpha}). \tag{3.14}$$

with $T = 1$. In other words, $U(\vec{\alpha})$ is the *negative log posterior*.

The form of the kinetic energy function $K(\vec{p})$ is up to the implementer of the algorithm and the choice can have a powerful impact on the overall performance of the sampler. Since many target distributions encountered in practice are approximately Gaussian, it is common practice to employ a multivariate normal distribution for $\vec{p}$:

$$\text{pr}(\vec{p}) \propto \exp\left(-\frac{1}{2}\vec{p}^{\,T}\mathcal{M}^{-1}\vec{p}\right). \tag{3.15}$$

Strictly speaking, the pdf for $\vec{p}$ is conditional on $\vec{\alpha}$ since

$$\text{pr}(\vec{\alpha}, \vec{p}) = \text{pr}(\vec{\alpha})\text{pr}(\vec{p}|\vec{\alpha}), \tag{3.16}$$

but I omit the conditional as $\vec{\alpha}$ and $\vec{p}$ are conditionally independent by design. The result is a quadratic form for $K(\vec{p})$,

$$K(\vec{p}) = \frac{1}{2}\vec{p}^{\,T}\mathcal{M}^{-1}\vec{p}, \tag{3.17}$$

where $\mathcal{M}$ is a problem-specific, user-chosen, positive-definite symmetric matrix, frequently called the mass matrix in a continuation of the physical analogy. $\mathcal{M}$ effectively sets the scale for how the sampler will traverse the parameter space in the various directions, and consequently influences the efficiency of the sampler. My implementation uses Equation (3.17) as the form of the kinetic energy.

Rather than sampling the posterior $\text{pr}(\vec{\alpha})$ directly, the HMC algorithm actually samples the joint pdf $\text{pr}(\vec{\alpha}, \vec{p})$ and then marginalizes over the auxiliary momentum, i.e.

$$\text{pr}(\vec{\alpha}) = \int \text{pr}(\vec{\alpha}, \vec{p}) d\vec{p}, \tag{3.18}$$

to obtain the sought posterior for the parameters.

3.4.2 Hamiltonian dynamics

A particle moving in a d-dimensional space governed by Hamiltonian dynamics moves through the $2d$-dimensional phase space $(\vec{\alpha}, \vec{p})$ on a hypersurface of

constant energy. The time evolution of the system is described by Hamilton's equations

$$\frac{d\alpha_i}{dt} = \frac{\partial H}{\partial p_i} \tag{3.19}$$

$$\frac{dp_i}{dt} = -\frac{\partial H}{\partial \alpha_i} \tag{3.20}$$

where $i = 1 \dots d$. The form of the Hamiltonian, Equation (3.11), allows us to rewrite Hamilton's equations as

$$\frac{d\alpha_i}{dt} = \frac{\partial K}{\partial p_i} = (\mathcal{M}^{-1}p)_i \tag{3.21}$$

$$\frac{dp_i}{dt} = -\frac{\partial U}{\partial \alpha_i}. \tag{3.22}$$

Note that $\partial U/\partial \alpha_i$ is a partial derivative of the negative log posterior. We must therefore be able to evaluate this partial derivative in order to simulate Hamiltonian dynamics and, by extension, use HMC sampling.

We may now move the particle to another position by solving Equations (3.21) and (3.22) for a finite period of time. The prescription for evaluating the potential energy, Equation (3.14), associates high probability with low potential energy. Intuitively, then, we expect the MCMC particle to tend towards areas of high probability density, while it can still reach low density areas given high momentum.

3.4.3 Advancing the HMC sampler

With the tools in place for gliding along constant-energy surfaces and for associating energies with probabilities, the stage is set for an MCMC sampler based on Hamiltonian dynamics. Assuming we have a parameter sample $\vec{\alpha}$, we proceed to draw a random momentum $\vec{p}$ from Equation (3.15). We then solve Equations (3.21) and (3.22) starting from $(\vec{\alpha}, \vec{p})$ for a period of time and end up with new position and momentum vectors. The new momentum is subsequently negated for reasons that will be made clear momentarily, and the result is a proposed sample $(\vec{\alpha}', \vec{p}')$. For a hypothetical exact solution to Hamilton's equations the joint probability $\mathrm{pr}(\vec{\alpha}', \vec{p}')$ for the proposed sample would be exactly equal to $\mathrm{pr}(\vec{\alpha}, \vec{p})$ due to the conservation of energy. We could then simply accept the sample and discard the auxiliary momentum $\vec{p}'$, store the new parameter sample $\vec{\alpha}'$, and repeat the process starting from the new sample.

In practice it is necessary to solve Equations (3.21) and (3.22) numerically. The unavoidable numerical error will cause the total energy $H(\vec{\alpha}, \vec{p})$ to only be approximately conserved, leading to a difference between the probabilities $\mathrm{pr}(\vec{\alpha}, \vec{p})$ and $\mathrm{pr}(\vec{\alpha}', \vec{p}')$. This difference would, if uncorrected, cause the Markov chain to converge to a different stationary distribution than the sought $\mathrm{pr}(\vec{\alpha})$. The solution is to vet the proposed sample with the Metropolis rejection criterion. We compute the Hastings ratio, Equation (3.2), for the joint probabilities:

$$r = \frac{\mathrm{pr}(\vec{\alpha}', \vec{p}')}{\mathrm{pr}(\vec{\alpha}, \vec{p})} = \exp(-H(\vec{\alpha}', \vec{p}') + H(\vec{\alpha}, \vec{p})). \tag{3.23}$$

The proposal distribution $q(\cdot)$ in Equation (3.2) is in this case symmetrical, i.e.

$$\frac{q(\vec{\alpha}', \vec{p}' | \vec{\alpha}, \vec{p})}{q(\vec{\alpha}, \vec{p} | \vec{\alpha}', \vec{p}')} = 1, \tag{3.24}$$

because we previously negated the momentum so we do not need to include that factor. The new sample is then accepted with the probability $a = \min(1, r)$ given by Equation (3.1). The Metropolis rejection criterion requires the so-called detailed balance condition to be satisfied, meaning that

$$\mathrm{pr}(s)T(s, s') = \mathrm{pr}(s')T(s', s) \tag{3.25}$$

where $s = (\vec{\alpha}, \vec{p})$ and $T(s, s')$ is the transition probability to move from s to s'. The MCMC update must be reversible and volume-preserving for (3.25) to hold, a condition that is upheld by HMC since Hamiltonian dynamics are time-reversible and volume preserving. The method for solving Hamilton's equations must be chosen such that these properties are observed.

An illustration of how HMC navigates the joint position-momentum parameter space is shown in Figure 3.1, using a one-dimensional (1D) standard normal distribution $\mathcal{N}(0, 1)$ as an example posterior. In the upper plot the position α in the 1D parameter space is shown on the x axis and the momentum is shown on the y axis. The algorithm was started in $\alpha_0 = 0$ and randomly drew a momentum $p_0 \approx 0.18$ from a standard normal distribution. It then evolved α and p using Hamilton's equations for a fixed period of time. Because the sampler was started at the maximum a posteriori (MAP) point (i.e. the peak of the sampled distribution), the momentum immediately started decreasing—remember that high probability translates to low potential energy. After some time the HMC particle lost all momentum and started falling towards the MAP point again, ending up at the opposite side of the

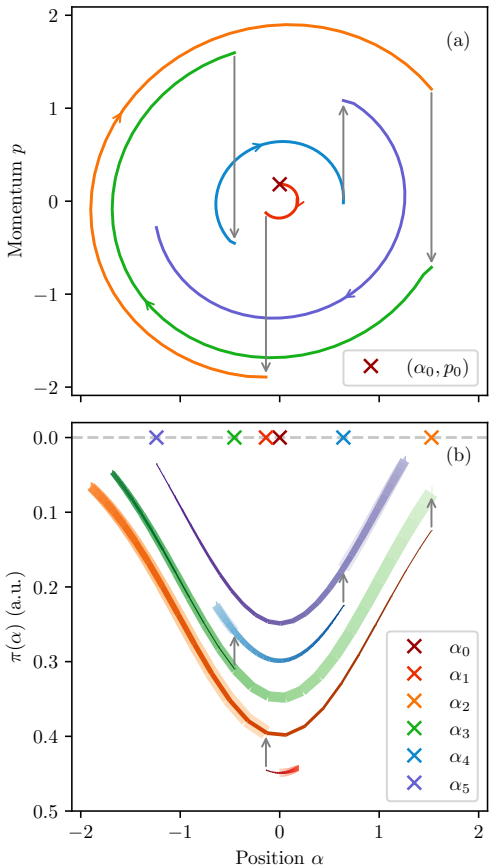

Figure 3.1: A detailed look at how HMC navigates through the parameter space when sampling a one-dimensional standard normal distribution $\pi(\alpha) = \mathcal{N}(0, 1)$. (a) five particle trajectories through the joint position-momentum space. The color coding follows the rainbow, i.e. the first trajectory is red, the second orange, and so on. The gray arrows indicate (random) resamplings of the momentum. The dark red cross indicates the starting position with $\alpha_0 = 0$ and a random momentum. The end points of each trajectory are the MCMC samples; intermediary states are discarded. (b) the same trajectories and color coding as above, but with the corresponding pdf value on the y axis. The y axis is inverted to visualize how HMC operates: the maximum of the pdf corresponds to the lowest potential energy and the particle tends to "fall" towards this position. The wide and pale parts of the lines are the beginnings of the trajectories, while the narrow and saturated parts are the ends. Note that the curves have vertical offsets to aid visualization. The accepted samples of α are shown as crosses.

MAP point. At this point the simulation stopped, and the proposed sample was vetted (and accepted) by the Metropolis rejection criterion. A new, much higher momentum was randomly drawn and the process repeated. The resampled momenta are indicated by gray arrows; note that the arrows are always vertical, i.e. the resampling of momentum does not affect the position. In the lower figure the pdf value is on the y axis and we get a visual representation of the pdf as a potential energy surface: the HMC particle tends to fall towards the MAP point, but can climb high given sufficient momentum, thus exploring the tails of the distribution. Note that the y axis is inverted in order to reflect the inner workings of HMC. Attentive readers may have noted that the momentum is *not* negated at the end of each trajectory; the reason for this is that the pdf for the momentum, Equation (3.15), is symmetrical with $\text{pr}(\vec{p}) = \text{pr}(-\vec{p})$ which means that the negation of the momentum does not need to be done in practice because Equation (3.24) holds true without the negation. This is obviously not true if $\text{pr}(\vec{p}) \neq \text{pr}(-\vec{p})$.

3.4.4 Leapfrogging the parameter space

Many algorithms exist for solving a set of differential equations like (3.21)-(3.22), but the HMC algorithm places important requirements on the solver: the numerical error must not be accumulative, and the solver must be time-reversible and volume preserving. These requirements disqualify several well-known methods, such as e.g. the Euler or Runge-Kutta methods. If the error was accumulative, the length of the particle trajectory would be limited because a large accumulated error would result in an unacceptably low acceptance rate. Worse, if the method is not reversible and volume-preserving, the Markov chain could converge to the wrong stationary distribution as the detailed balance condition is unsatisfied. To get around these issues HMC uses a special type of integrators called *symplectic integrators*. The local discretization error of a symplectic integrator is equally likely to be positive or negative in each step of the integration as long as the step size (i.e. discrete time step) ϵ is below some threshold value at which the energy diverges. The result is that the total energy is conserved to a high degreee for any arbitrarily long trajectory as long as ϵ is smaller than the threshold value, and the acceptance probability a remains high even for extremely long trajectories. The appropriate step size needs to be determined by trial-and-error in practice.

The standard symplectic integrator used in HMC is known as the leapfrog method, a second-order method similar to the velocity Verlet method [68]. Central to its function as a symplectic integrator is to take half-steps $\epsilon/2$ for advancing the momentum, and one iteration is completed as follows.

1. Starting from a position-momentum pair $(\vec{\alpha}_0, \vec{p}_0)$, the momentum is advanced by a half-step

$$\vec{p}_{\epsilon/2} = \vec{p}_0 - \frac{\epsilon}{2}\vec{\nabla}U(\vec{\alpha}_0). \tag{3.26}$$

2. The position is then advanced by a full step

$$\vec{\alpha}_\epsilon = \vec{\alpha}_0 + \epsilon\vec{\nabla}K(\vec{p}_{\epsilon/2}). \tag{3.27}$$

3. Finally, the momentum is advanced by another half-step

$$\vec{p}_\epsilon = \vec{p}_{\epsilon/2} - \frac{\epsilon}{2}\vec{\nabla}U(\vec{\alpha}_\epsilon). \tag{3.28}$$

More than one iteration is usually performed, in which case steps 3 and 1 may be combined except for the first and last iterations. The calculation proceeds as follows for $t = 0, \epsilon, \ldots, \epsilon L$ where L is the number of iterations.

1. First the momentum is advanced by a half-step

$$\vec{p}_{\epsilon/2} = \vec{p}_0 - \frac{\epsilon}{2}\vec{\nabla}U(\vec{\alpha}_0) \tag{3.29}$$

2. Then

$$\vec{\alpha}_{t+\epsilon} = \vec{\alpha}_t + \epsilon\vec{\nabla}K(\vec{p}_{t+\epsilon/2}) \tag{3.30}$$

$$\vec{p}_{t+3/2\epsilon} = \vec{p}_{t+\epsilon/2} - \epsilon\vec{\nabla}U(\vec{\alpha}_{t+\epsilon}) \tag{3.31}$$

   is repeated for $0 \leq t \leq \epsilon(L-2)$.

3. Finally the position and momentum at the end of the trajectory is computed as

$$\vec{\alpha}_{\epsilon L} = \vec{\alpha}_{\epsilon(L-1)} + \epsilon\vec{\nabla}K(\vec{p}_{\epsilon(2L-1)/2}) \tag{3.32}$$

$$\vec{p}_{\epsilon L} = \vec{p}_{\epsilon(2L-1)/2} - \frac{\epsilon}{2}\vec{\nabla}U(\vec{\alpha}_{\epsilon L}). \tag{3.33}$$

The proposed sample $(\vec{\alpha}_{\epsilon L}, -\vec{p}_{\epsilon L})$ is then vetted through the Metropolis update as discussed in Section 3.4.3. Figure 3.2 shows an example of how the leapfrog integrator navigates the phase space; notice that each trajectory begins and ends with a half-step in the momentum.

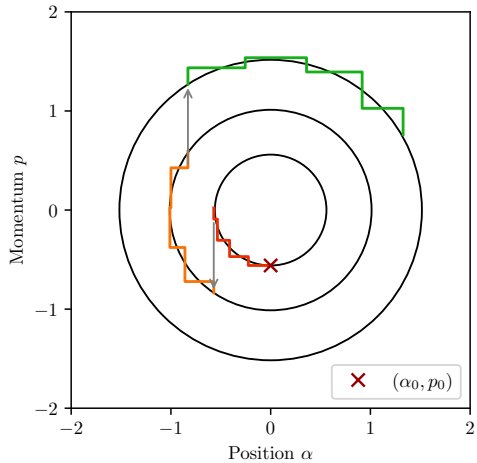

Figure 3.2: Symplectic leapfrog integration using $L = 4$ iterations visualized during three HMC samples of an $\mathcal{N}(0, 1)$ distribution. The color coding is the same as in Figure 3.1 (a). The black circles are the constant energy curves that the integrator is approximating.

The leapfrog hyperparameters ϵ and L need to be chosen by the user in the standard HMC algorithm I have implemented here. The step size ϵ directly affects the acceptance rate and should be tuned to be close to, but not greater than, the threshold value for which the total energy can no longer be conserved. A nice property of HMC is that the acceptance rate drops precipitously if ϵ is greater than the threshold value; the lack of quiet failures makes it easy to diagnose a too-large ϵ. The step size should not be chosen too small, either, as this would result in wasteful computational effort. Tuning ϵ such that the acceptance rate becomes around 90-95% is a good compromise between stability and efficiency in my experience. The upper bound for ϵ is generally determined by the most constrained parameter in the posterior $\mathrm{pr}(\vec{\alpha})$ that is being sampled by HMC.

The number of leapfrog iterations ("steps") L can drastically influence the performance of the sampler, and it should neither be chosen too high nor too low. A too-low choice partly defeats the purpose of using HMC in the first place, as it would result in a random-walk-like behaviour with highly correlated samples. In contrast, a too-high choice of L would waste valuable CPU cycles without improving performance. Note that choosing a very small

step size ϵ needs to be compensated for by increasing L in order to avoid random walks.

Neither ϵ nor L are fixed in `montepython`. Rather, the user picks nominal values and random values ϵ^* and L^* are drawn from predefined probability distributions prior to each invocation of the leapfrog solver. The parameters are drawn from

$$\epsilon^* \sim \mathcal{U}\left(\frac{1}{2}\epsilon, \frac{3}{2}\epsilon\right) \tag{3.34}$$

$$L^* \sim \mathcal{U}\left\{\frac{1}{2}L, \frac{3}{2}L\right\}, \quad \frac{1}{2}L \geq 1. \tag{3.35}$$

The reasons for randomly perturbing the leapfrog parameters are threefold. First, varying the trajectory length ϵL can decrease correlations between samples. Second, a fixed trajectory length may result in oscillatory behavior if ϵL happens to approximately match some periodic feature of the target distribution. To see this, revisit Figure 3.1 and imagine that the trajectories were about 1.5 times longer: the initial and final states would be almost identical each time. Third, the target pdf may have smaller sections where the gradient is very steep so that the nominal ϵ is too large to resolve features in that section of the parameter space, and by drawing ϵ from a distribution we may occasionally draw a small enough value for the sampler to be able to efficiently explore this region.

Extensions of HMC whose purpose is to relieve the user from the burden of tuning ϵ and L exist. The choice of ϵ may be automated by letting the computer try different values and adjusting based on acceptance rates of small trial runs. The state-of-the-art No-U-Turn Sampler (NUTS), which is based on HMC, automates the choice of L using heuristic rules for when continuing a trajectory no longer increases the performance of the sampler. Both of these improvements are described in Ref. [69].

3.4.5 Tuning in to the target

An array of tunable HMC parameters have been introduced in the previous subsections: the step size ϵ, the number of leapfrog iterations L, and the mass matrix $\mathcal{M}$. A drawback of HMC is the need to carefully tune these hyperparameters to each target distribution, or risk poor performance. However, as mentioned in Section 3.4.4, the tuning of ϵ and L may be automated, and as we shall see the same holds true for $\mathcal{M}$, paving the way for a fully automated setup. `montepython`, in its current state of development, uses a manual tun-

ing procedure and in this section I will describe the process of tuning these parameters to achieve efficient sampling.

The keen reader may have noticed that ϵ, L, and $\mathcal{M}$ are interlinked and changing one may force us to change one (or both) of the others. The link between ϵ and $\mathcal{M}$ can be seen by inspecting Equation (3.27): ϵ multiplies $\vec{\nabla}K(\vec{p})$, which in turn depends on $\mathcal{M}$. Both ϵ and $\mathcal{M}$ relate to the topography of the sampled posterior. The link between L and $\mathcal{M}$ is similar to that between ϵ and L, in that an unsuitable choice of $\mathcal{M}$ forces L to increase in order to maintain short autocorrelation times in all directions. The tuning procedure is iterative to an extent because of these relationships, and there is no fixed order in which to tune the parameters. It is, however, often best to start by estimating a ballpark value for ϵ.

Leapfrog step size and number of iterations

For tuning ϵ we exploit that the acceptance rate is independent of L (a consequence of using a symplectic integrator) and use a small number of leapfrog steps, $L \approx 3$, to save time. We need an initial guess of ϵ; this guess may in some cases be completely blind, or it may be educated based on previous experience of similar target distributions. For example, choosing $\epsilon \approx 0.5$ as a first guess is reasonable if the target distribution is a Gaussian of width 1. We then run the sampler and collect a few tens of samples, and monitor $\bar{a}$. If $\bar{a}$ is low (typically 0) we decrease ϵ by an order of magnitude and try again. If $\bar{a}$ is 100% we instead increase ϵ and try again. We stop once we see an acceptance rate of around 60-90%. The step size may need to be modified later if the acceptance rate changes during production runs.

Careful tuning of L is rather pointless until $\mathcal{M}$ is locked in since the number of leapfrog steps is intimately linked to the choice of the mass matrix. We have found that choosing $L \approx$ 10-20 yields excellent performance for approximately Gaussian distributions with around a dozen or so parameters, assuming that the mass matrix (and ϵ) is well chosen. To quickly assess the choice of L—and the overall performance of the sampler—it is useful to study trace plots of each individual parameter. The trace plots will reveal no significant structures (i.e. autocorrelations) if the sampler is performing well. Autocorrelations may be quenched by increasing L, at the obvious expense of increased computational effort, or by improving the mass matrix. The latter alternative should take precedence. Two example traceplots are shown in Figure 3.3, where I have sampled a 1D standard normal distribution with two different values of L while keeping other parameters fixed. I have deliberately chosen a too-low value for the step size ϵ in order to accentuate the impact of L. The left

panel, with $L = 3$, show clear correlations between subsequent samples. In fact, the sampler has barely visited positive parameter values over the course of the 100 samples; this sampler is not performing well. The right panel on the other hand, with $L = 25$, performs much better: there are few, if any, correlations between samples and the sampler moves quickly across the typical set of parameter values.

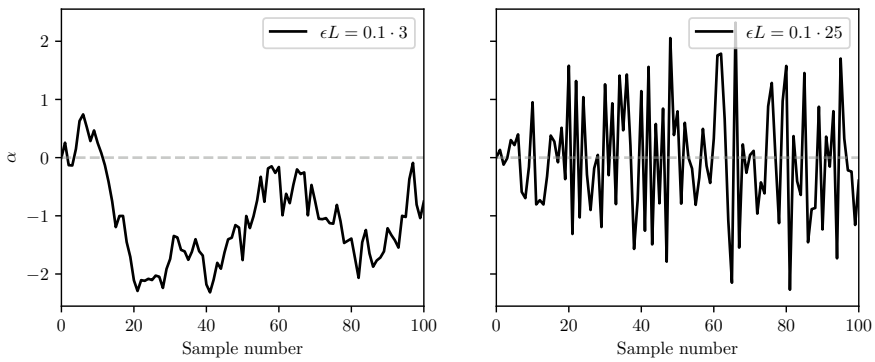

Figure 3.3: Parameter traces for two HMC samplings of a 1D $\mathcal{N}(0, 1)$ distribution. *Left:* $L = 3$. *Right:* $L = 25$. Note the difference in correlation structure between the two chains.

Mass matrix

Both ϵ and L are scalar values with no distinction for each individual parameter α_i and thus cannot be used to compensate for differences in scale between parameters. Like M-H, and unlike `emcee`, HMC is sensitive to such differences of scale and we need a way to account for them. This is the role of the mass matrix.

As an example, picture a two-dimensional Gaussian target distribution with identical variances $\sigma_1^2 = \sigma_2^2 = 1$ for the two parameters, as shown in Figure 3.4 (a). Assuming we are using an identity mass matrix, $\mathcal{M} = I$, the sampler will explore both parameters with equal efficiency, travelling a distance of order ϵL in both directions. But what if we have $\sigma_1^2 \neq \sigma_2^2$, e.g.

$$\Sigma = \text{diag}(\sigma_1^2, \sigma_2^2) = \text{diag}(1, 100), \tag{3.36}$$

where Σ is the covariance matrix of the target distribution, as in Figure 3.4 (b)? The sampler will still travel a distance of about ϵL in both directions. The

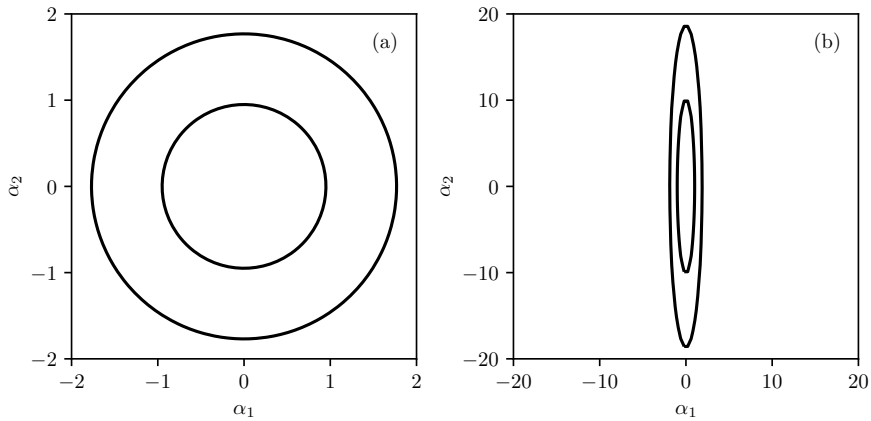

Figure 3.4: Two-dimensional Gaussian pdfs $\mathcal{N}(0, \Sigma)$. The indicated contours enclose 39% and 86% of the probability mass, respectively. (a) $\Sigma = \text{diag}(1, 1)$. (b) $\Sigma = \text{diag}(1, 100)$.

result is that about ten times more samples are needed to explore the second parameter, i.e. we will have correlated samples of the second parameter.

The solution to the issue of disparate scales is, in a continuation of the physical analogy, to decrease the mass of the MCMC particle in the less constrained direction, thus increasing its velocity. A more suitable mass matrix for the second example above will be

$$\mathcal{M} = \text{diag}(m_1^2, m_2^2) = \text{diag}(1, 0.1^2) \tag{3.37}$$

where we have decreased m_2 by the same factor that σ_2 increased. Note that $\mathcal{M} = \Sigma^{-1}$. $\mathcal{M}$ will in the most general case have non-zero off-diagonal elements to reflect correlations between the target parameters α_i.

I have found that constructing a mass matrix that captures the most important features of the target distribution is critical to the performance of HMC. An improper choice of $\mathcal{M}$ may degrade the performance by several orders of magnitude. I have tested four different methods for constructing $\mathcal{M}$, two of which—(a) and (d)—are completely general, and two—(b) and (c)—that are specific to sampling χEFT LEC posteriors:

(a) $\mathcal{M} = I$. This is the least informed mass matrix, but the easiest to construct. It is the default mass matrix in `montepython`.

(b) Exploiting the naturalness expectation of the LECs, i.e. the expectation that the LECs should be $O(1)$ in suitable units of the breakdown scale of χEFT. The contact LECs at LO-NNLO are expected to be of size [4]

$$\left|\widetilde{C}_i\right| \sim \frac{4\pi}{F_\pi^2}, \quad |C_i| \sim \frac{4\pi}{F_\pi^2 \Lambda_b^2} \tag{3.38}$$

where $F_\pi \simeq 92$ MeV is the pion decay constant. Inserting $\Lambda_b = 600$ MeV (see Section 2.3.1) yields

$$\left|\widetilde{C}_i\right| \sim 0.15\ 10^4 \cdot \text{GeV}^{-2}, \quad |C_i| \sim 0.4\ 10^4 \cdot \text{GeV}^{-4}. \tag{3.39}$$

A diagonal "naturalness" mass matrix can be constructed given the assumption that the relative sizes of $\widetilde{C}_i$ and C_i also hold for the corresponding standard deviations.

(c) Exploiting published LEC uncertainties from a previous study, e.g. Ref. [19] or [18], to construct a diagonal mass matrix.

(d) Performing a preliminary sampling with a mass matrix based on the method in (c) and estimating the covariance matrix $\Sigma_{\vec{\alpha}}$ of the parameters based on the resulting Markov chain. The mass matrix is then set to the inverse of the estimated covariance matrix, i.e. $\mathcal{M} = \Sigma_{\vec{\alpha}}^{-1}$ [67].

I have found that only Method (d) yields high performance in practice. The mass matrices used to produce the posteriors presented in the next chapter were constructed using Method (d). Around 500 preliminary samples were necessary, although I used 2,000 at LO and NLO due to an abundance of samples. Note that the initial sampling required in Method (d) does not have to be performed using HMC at all; indeed, it may be preferable to use a tuning-insensitive MCMC sampler such as `emcee` for the preliminary sampling.

It may be necessary to revisit the tuning of ϵ after the mass matrix has been updated. In some cases it is necessary to perform another full iteration of tuning if the sampler shows signs of sub-par performance.

Chapter 4

Posteriors, efficiency, and convergence

In this chapter I present the extracted LEC posteriors at LO, NLO, and NNLO of χEFT from inference using 2N scattering data and assess the efficiency of HMC. I also address the question of MCMC convergence, as is necessary to judge the reliability of the results.

4.1 Sampling procedure

As noted in the previous chapter, the sampling performance of HMC is sensitive to the topography of the target pdf and hyperparameter tuning is usually required in order to achieve high performance. Here I will summarize the procedure I have used to produce the results presented in this chapter. The procedure is essentially the same for each chiral order, with a minimal simplification at the lowest order. It should readily generalize to an extended analysis at higher orders and also to different fields of research.

A starting point $\vec{\alpha}_0$ that is reasonably close to the mode of the pdf is a valuable base for the tuning procedure. Such a point is now fairly easy to acquire in the context of χEFT due to numerous previous studies, see e.g. [17–19, 70]. In the absence of such explicit prior information one may try to locate the mode using e.g. optimization or history matching [71]. I have used the preliminary LEC values $\vec{\alpha}_{\text{prel}}$, see Section 2.3.1, as starting points. It should be noted that HMC itself is rather adept at finding regions of high probability density as the HMC "particle" will tend towards the mode (or one

Table 4.1: Detailed statistics of the HMC chains during the tuning and sampling phases. n_{tune} is the number of samples gathered during the tuning procedure. M is the number of parallel chains. n_{total} refers to the total number of samples across all M chains. The HMC parameters ϵ and L denote the step length and total number of steps taken with the leapfrog algorithm to integrate Hamilton's equations for each HMC step, and $\bar{a}$ is the average acceptance rate during the production runs.

Order	n_{tune}	M	n_{total}	ϵ	L	$\bar{a}$
LO	2,000	3	50,063	0.1	8	99%
NLO	2,000	10	57,134	0.09	20	99%
NNLO	591	3	10,155	0.08	20	99%

of the modes in the case of a multimodal pdf), so while tools such as history matching are certainly helpful, they are not mandatory.

Once a region of interest has been identified it is time to tune the three hyperparameters of the HMC sampler: the step size ϵ, the number of leapfrog steps L, and the mass matrix $\mathcal{M}$. I did this in two stages, where the purpose of the first stage is to acquire the mass matrix and the second stage is for fine-tuning ϵ and L. The objective of the first stage is to approximate a covariance matrix of the sampled parameters; this covariance matrix is then inverted to yield the mass matrix. I achieve this by performing an initial HMC sampling with a basic mass matrix constructed following Method (c) from Section 3.4.5. In the second stage I adjust ϵ (if necessary) to yield an acceptance rate $\bar{a}$ of approximately 90% and perform short sampling runs to find the smallest L that yields uncorrelated samples as judged from visual inspection of traceplots. This procedure uses HMC for the preliminary samplings, but as noted in the previous chapter, it may be advantageous to employ another sampling algorithm for this purpose.

Summary sampling statistics are shown in Table 4.1, including the number of samples used during tuning, number of parallel chains, total number of samples, and acceptance rate. In general, at least $M = 3$ chains produced with identical sampling hyperparameters—but different starting points—are needed to assess convergence. I used $M = 10$ at NLO in order to see if this yielded any appreciable advantage beyond faster collection of samples (it did not). I have discarded the first samples of each chain in order to avoid bias from the choice of starting point; this procedure is called "burn-in". The number of burn-in samples varies between chains, but is of the order of 100 samples.

4.2 Extracted LEC posteriors

4.2.1 LO

As we shall see in Chapter 5, the LO chiral potential using Weinberg power counting often yields insufficiently accurate predictions to be of practical value, yet sampling its parameters has high instrumental value for a couple of reasons. First of all there is generally a dearth of information regarding the EFT convergence pattern and the leading order provides a highly influential piece of information for characterizing the EFT truncation error. Second, this sampling is very manageable due to its low number of parameters (two) and relatively low computational cost, which is about 2.5 s per likelihood evaluation in our implementation on modern CPUs. The LO posterior thus bridges the gap between toy models and the truly challenging higher-order posteriors.

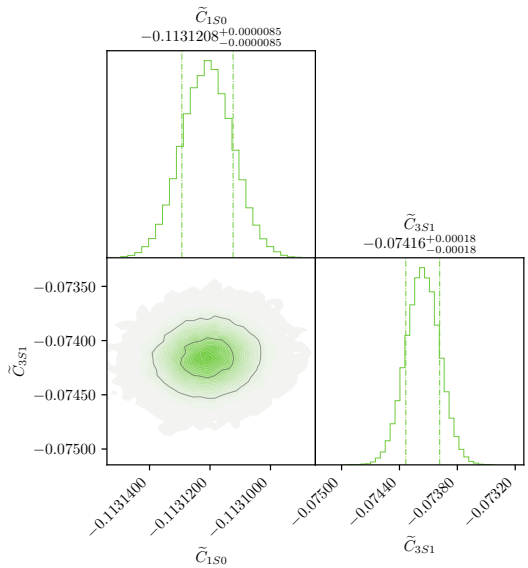

Figure 4.1: LO posterior sampled with HMC. The sampling consists of three independent MCMC chains totalling 50,063 samples with the first ten samples of each chain discarded as burn-in. The LECs are shown in units of $10^4 \cdot \mathrm{GeV}^{-2}$. The inner (outer) gray contour line encloses 39% (86%) of the probability mass. The dot-dashed vertical lines indicate a 68% equal-tailed credibility interval (see Chapter 5). White areas indicate zero counts.

The HMC sampling at LO was performed following the recipe outlined earlier. A short preliminary sampling was performed using an identity mass matrix. A parameter covariance matrix was extracted from the chain using the `cov` function in NumPy [65], and the mass matrix was set to the inverse of the extracted covariance matrix. The two leapfrog parameters ϵ and L (see Table 4.1) were chosen by inspecting acceptance rates and trace plots, followed by production runs using $M = 3$ parallel chains totalling about $50,000$ samples. This rigor is not strictly necessary at LO because the posterior is not particularly challenging; the identity mass matrix already performed quite adequately in this case. I chose to optimize the sampling performance anyway as it provided valuable hands-on experience in a relatively controlled setting.

The extracted LO posterior pdf is shown in Figure 4.1[1]. The LECs, especially $\widetilde{C}_{1S0}$, are well-determined with almost vanishingly small uncertainties and are not significantly correlated. Quantitatively, this result closely resembles central values and uncertainties reported in previous (frequentist) studies such as Ref [18], which is unsurprising because the likelihoods used are very similar and the prior used here is essentially uniform in the region of the likelihood peak. However, the interpretation is fundamentally different. Figure 4.1 shows an extracted probability distribution for the *parameters*, whereas Ref. [18] reports the maximum likelihood estimator of the *data*, with covariances estimated from the local gradients at the MLE.

4.2.2 NLO

The NLO potential in χEFT introduces seven $k = 2$ renormalizing contact LECs in partial waves with orbital angular momentum $l \leq 2$. Furthermore, the LO $^1\mathrm{S}_0$ LEC is split into three LECs—$\widetilde{C}^{nn}_{1S0}, \widetilde{C}^{np}_{1S0}$, and $\widetilde{C}^{np}_{1S0}$—to account for charge-independence breaking effects seen in e.g. $^1\mathrm{S}_0$ 2N scattering lengths stemming from differences in the up and down quark masses and electromagnetic interactions between quarks [4]. The $\widetilde{C}^{nn}_{1S0}$ LEC can be estimated from e.g. scattering lengths extracted from low-energy scattering cross sections. I have omitted this LEC from my samplings, bringing the total number of LECs to estimate at NLO to ten. This is a far more challenging problem than the LO sampling, and proper tuning is required in order to keep the computational cost within reasonable limits. The tuning procedure is the same as at LO, except that the mass matrix used in the preliminary sampling was informed by parameter uncertainties from a previous study [19].

[1] The software package used to produce this and subsequent so-called corner plots is available at `https://github.com/svisak/prettyplease.git`

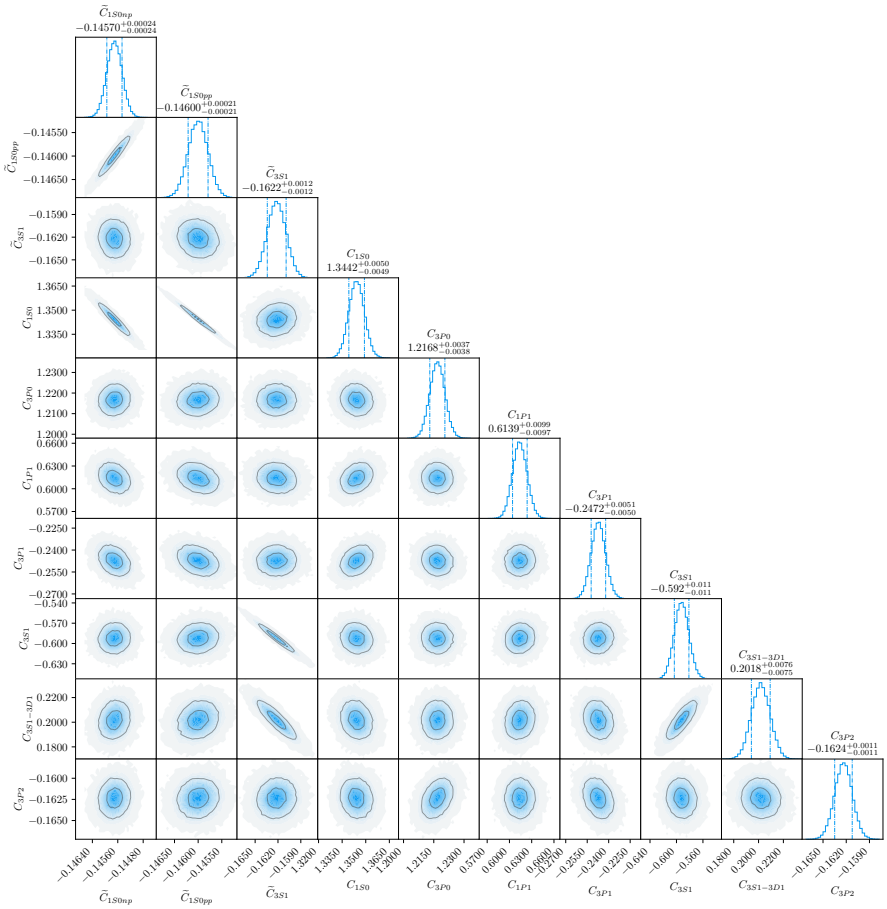

Figure 4.2: NLO posterior sampled with HMC. The LECs are shown in units of $10^4 \cdot \mathrm{GeV}^{-2}$ for the LO LECs and $10^4 \cdot \mathrm{GeV}^{-4}$ for the NLO LECs. The inner (outer) gray contour line encloses 39% (86%) of the probability mass. The dot-dashed vertical lines indicate a 68% credibility interval.

The extracted NLO posterior pdf is shown in Figure 4.1 and the standout feature of this pdf compared to the LO pdf—beyond the increased number of parameters—is that several pairs of parameters now show strong correlations or anticorrelations (indicated by tilted 2D marginal distributions) hinting at possible parameter redundancies. For example, the isospin symmetry breaking LECs discussed above are quite strongly correlated and display nearly identical distributions. The LO LECs in the $^1\mathrm{S}_0$ and $^3\mathrm{S}_1$ partial waves are also anticorrelated with their higher-order counterparts. Beyond these correlations, most parameters are notably uncorrelated. In general the NLO LECs are less constrained than the LO LECs in terms of the widths of the credibility intervals.

The computational cost of a single evaluation of the NLO likelihood is only modestly higher than at LO, with each evaluation taking around 4 s compared to the 2.5 s at LO. The increased sampling difficulty stems largely from the increased dimensionality of the posterior.

4.2.3 NNLO

The three πN LECs c_1, c_3, c_4 entering at NNLO bring the number of parameters to be estimated to thirteen. This posterior presents a more difficult sampling problem than the NLO posterior. The reason for this is not primarily the greater dimensionality of the sampled pdf, but rather a significant increase in the cost per sample which is around three to four times higher at NNLO compared to NLO. This increase originates from the increased number of diagrams that need to be calculated. The specific cost increase is of course specific to the implementation I am using.

The NNLO posterior, shown in Figure 4.3, again reveals a strong correlation and overall similarity between the isospin breaking LECs. However, the anticorrelations between the LO and NLO LECs in the S-waves have vanished. This is likely a result of allowing the πN LECs, which act in all angular momentum channels, to vary; as we shall see shortly, the strong $\widetilde{C}_{1S0}$–C_{1S0} anticorrelations return if the πN LECs are fixed. In general, more contact LECs show correlations at NNLO than at NLO, but the correlations are rather weak. The πN LECs show some notable correlations with the 2N LECs. In particular, c_1 correlates strongly with the two $\widetilde{C}_{1S0}$ LECs. c_3, on the other hand, correlates with virtually all NLO contact LECs. The LECs c_1 and c_3, which appear in combination in the central potential term [4], show little internal correlation. The correlations between the remaining πN LEC, c_4, and the other parameters are weak.

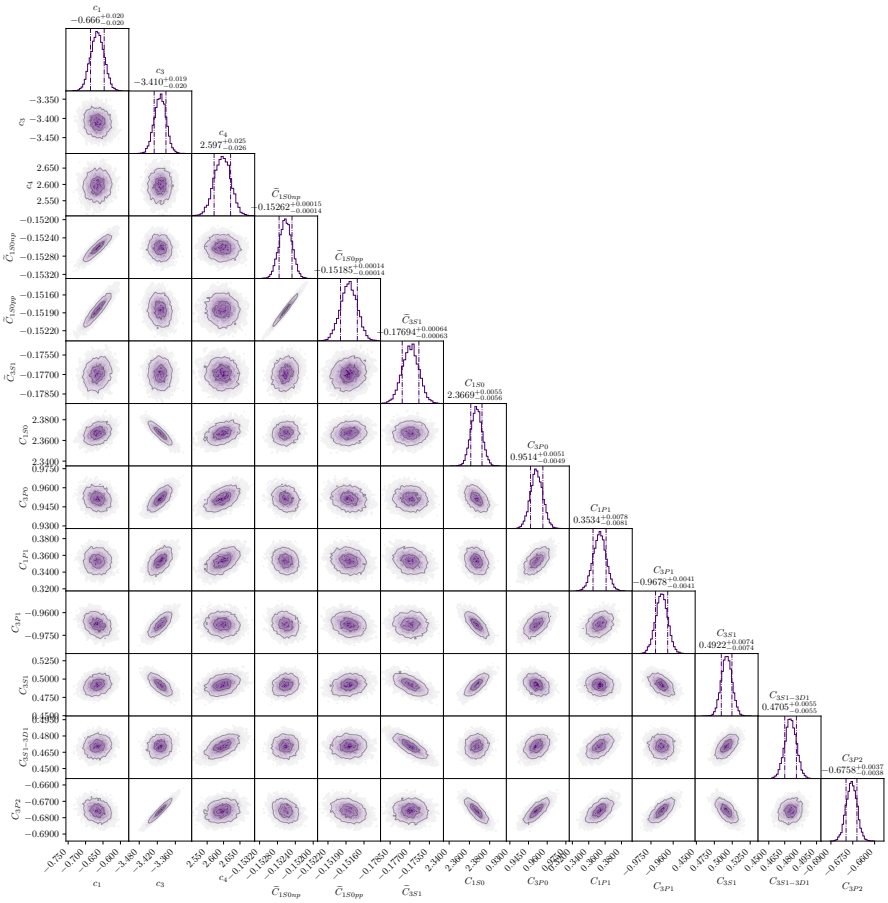

Figure 4.3: NNLO posterior sampled with HMC. The LECs are shown in units of $10^4 \cdot \mathrm{GeV}^{-2}$ for the LO contact LECs, $10^4 \cdot \mathrm{GeV}^{-4}$ for the NLO contact LECs, and GeV^{-1} for the πN LECs. The inner (outer) gray contour line encloses 39% (86%) of the probability mass. The dot-dashed vertical lines indicate a 68% credibility interval.

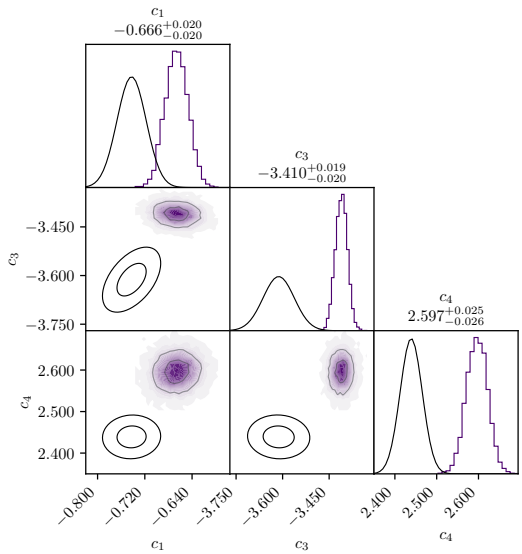

Figure 4.4: Prior and posterior pdfs for the πN LECs c_1, c_3, and c_4, indicated with black-line ellipses and colored (purple jagged) regions, respectively. The inner (outer) black ellipses enclose 39% (86%) of the prior probability mass, and the jagged gray lines do the same for the posterior probability mass. The posteriors were obtained using HMC. The LECs are shown in units of GeV^{-1}. See text and Figure 4.3 for further details.

Are there any notable differences between the prior and posterior distributions for the πN LECs? Remember that the prior for the πN LECs comes from an R-S analysis of πN scattering data. It is therefore interesting to study how the 2N data updates our knowledge. Figure 4.4 shows both the prior for the πN LECs and the corresponding marginal posteriors. The shape of the distribution has largely remained the same in the transition from prior to posterior, except that the c_3 marginal distribution has narrowed somewhat. However, the central values have shifted by about 5-10 % in the direction of less attraction on the potential level and the pdfs do not show any significant overlap except for a small interval of the LEC c_1. This is indicative of tension between the 2N and πN sectors and hints at possible underestimation of the uncertainties for either the πN LECs or the 2N LECs, or both. Ideally, the difference should be reconciled and the pdfs should show a considerable

overlap; further research is needed to shed more light on this result.

4.2.4 Naive maximum likelihood versus Bayesian parameter estimation

As I have performed two related, but conceptually very different, parameter estimations of the 2N LECs at NNLO, it is instructive to compare the two results. In Paper A, I extracted a tightly constraining prior for the 2N LECs by finding an MLE with local covariance estimation on a simplified version of the NNLO likelihood that was later used in Paper B. Important differences between the two likelihoods include that we, somewhat naively, set the value of $\bar{c}$ to unity in Paper A while $\bar{c}$ was estimated from order-by-order differences in Paper B (see Table 2.2), and that the πN LECs were fixed in Paper A but allowed to vary in Paper B.

The resulting pdfs are jointly shown in Figure 4.5. Here I have only included the LECs that were allowed to vary in both studies, i.e. I have omitted the $\widetilde{C}_{1S0}^{nn}$ LEC from the Paper A prior and the πN LECs from Paper B. The HMC sampled posterior is shown in purple as before, while the 39% and 86% contours extracted from the MLE curvature are shown in black.

Two differences stand out. The majority of the LECs have notably different central values between the two pdfs. This is a result of the different values of $\bar{c}$, an effect I have also encountered when performing HMC samplings of the NLO and NNLO posterior with different values for $\bar{c}$. However, the most notable qualitative difference is in the correlation structure between the $\widetilde{C}_{1S0}$ and C_{1S0} parameters. While the HMC sampled NNLO posterior shows no discernible correlation between these parameters, the MLE-based prior shows strong anticorrelations similar to those found in the HMC-sampled NLO posterior. It is likely that allowing the πN LECs to vary softens these correlations and that the strong anticorrelations seen in the Paper A prior would vanish had the πN LECs not been kept fixed.

4.3 HMC performance evaluation

Having presented the extracted posteriors at three chiral orders, two questions arise:

1. Is it actually beneficial to use HMC?
2. Can the results be trusted not to change as we continue to collect more samples?

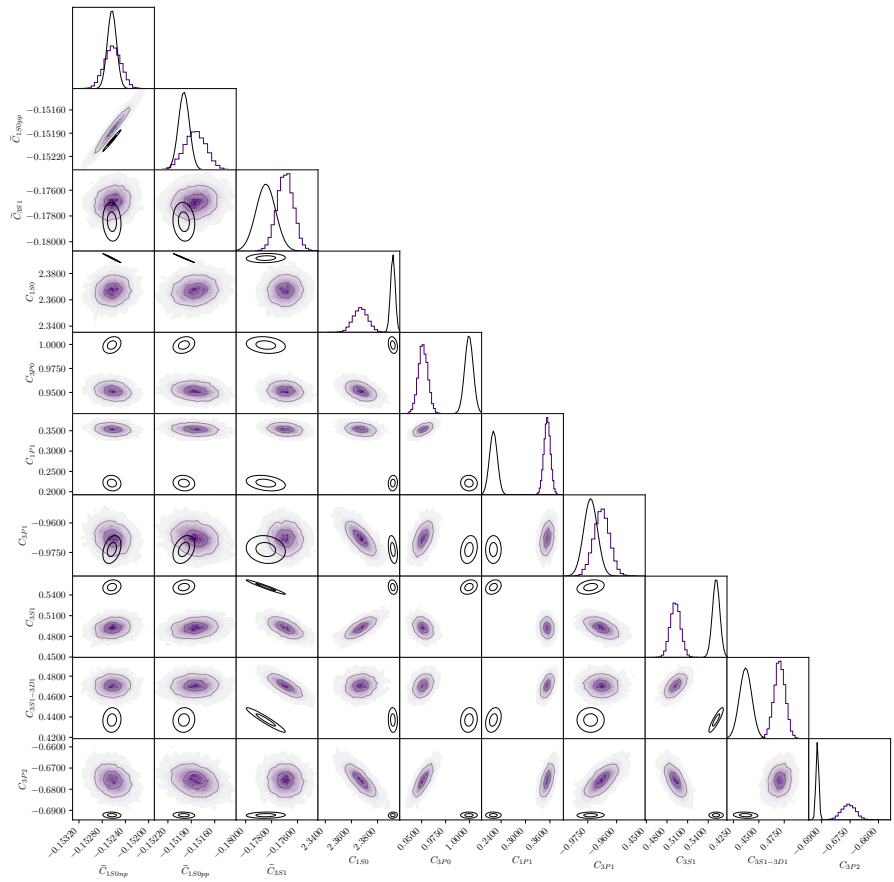

Figure 4.5: Side-by-side comparison between the prior from Paper A (black contour lines) and the posterior in Paper B (filled purple contours) at NNLO. Only parameters estimated in both projects are included.

The remainder of this chapter will be dedicated to answering these questions. The first is a major motivation for Paper B, whereas the second should always be addressed in any application of MCMC.

4.3.1 Efficiency

The efficiency of an MCMC sampler can be boiled down to a simple number: the number of effective samples—see Equation (3.9)—per unit time. An increase in efficiency enables us to either explore more aspects of an inference problem, or attack problems that are otherwise out of reach. Two factors primarily affect the efficiency and need to be balanced appropriately: the computational cost of a single MCMC sample and the integrated autocorrelation time τ. Here I will compare the efficiency of HMC and `emcee`, which take different approaches to maximizing efficiency. While `emcee` opts to gather many correlated samples quickly, HMC aims to produce costly samples with minimal correlations.

It is very common in Bayesian inference that the evaluation of the likelihood functions accounts for the vast majority of the computational cost. This is also the case here, and I therefore judge the sampling efficiency based on the number of likelihood evaluations per effective sample, $N_{\mathcal{L}}/\text{ESS}$. As the gradient extraction (carried out using AD) required by HMC incurs an additional overhead I will also take this into account when judging the efficiency. The measured overhead factors AD-cost are 1.1, 1.24, and 1.43 at LO, NLO, and NNLO, respectively, and I use these figures in my efficiency assessment. I also account for the cost of the tuning samples required by HMC, but it is worth noting that this cost is partially offset by a reduction in the number of necessary burn-in samples. Thus, the relative efficiency $\mathcal{S}$ of HMC compared to e.g. `emcee` can be calculated as

$$\mathcal{S} = \frac{[N_{\mathcal{L}}/\text{ESS}]_{\texttt{emcee}}}{[N_{\mathcal{L}}/\text{ESS}]_{\text{HMC}} \times [\text{AD-cost}]}. \tag{4.1}$$

There is an immediate advantage for `emcee` here, because while `emcee` only needs one likelihood evaluation per sample, HMC requires (on average) L evaluations per sample, where L is the number of leapfrog steps. Recall that I use $L_{\text{LO}} = 8$ and $L_{\text{NLO}} = L_{\text{NNLO}} = 20$. To be competitive, HMC must counter this by achieving integrated autocorrelation times τ_{HMC} that are at least a factor $L \times$ AD-cost smaller than the `emcee` counterpart $\tau_{\texttt{emcee}}$, i.e.

$$\tau_{\text{HMC}}(L \times \text{AD-cost}) \leq \tau_{\texttt{emcee}}. \tag{4.2}$$

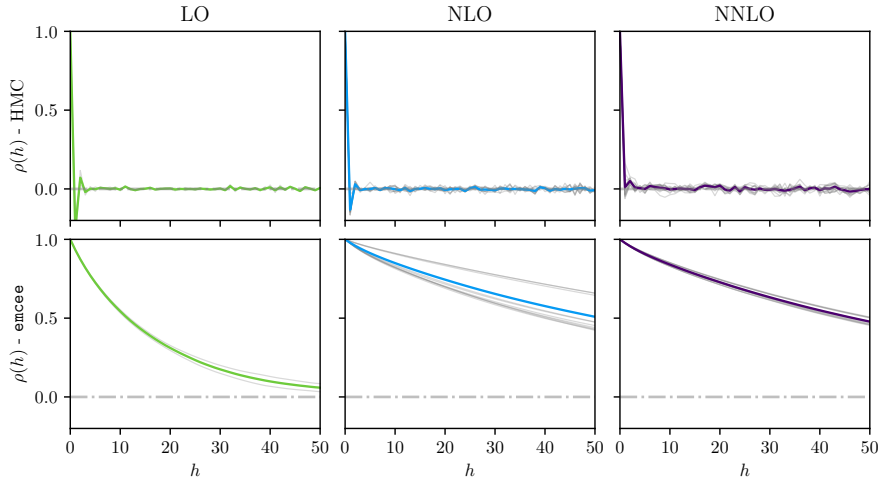

Figure 4.6: Autocorrelation functions of the MCMC chains at LO-NNLO. The chains produced with HMC are shown in the first row and chains produced with `emcee` are shown in the second row. The gray lines show autocorrelations in individual parameters whereas the colored lines show the average over all parameters. The results are averaged over all walkers in the plots showing the `emcee` autocorrelations. See Table 4.1 for details about each sampling.

The expression for τ, Equation (3.8), reveals that the autocorrelation function $\rho(h)$ must be kept small for all lags $h \neq 0$ in order to achieve a short integrated autocorrelation time. In Figure 4.6 I show $\rho(h)$ for both HMC and `emcee` samplings at all three chiral orders. All autocorrelation functions were computed using `emcee`'s built-in function `emcee.autocorr.function_1d`, which uses an efficient algorithm based on the fast Fourier transform. It is apparent that the `emcee` samples are much more correlated than the HMC samples across all three orders, but especially at NLO and NNLO. A closer look at the first two panels in the top row reveals that subsequent HMC samples at LO and NLO are slightly *anti*correlated; as discussed in Section 3.3, this means that we have $\tau < 1$ and that the ESS will actually be greater than the length of the chain.

Table 4.2 shows vast reductions in the integrated autocorrelation time with HMC compared to `emcee`. These results are for the parameter samples; the results will generally be different for functions of these samples. The estimate of τ for the `emcee` sampling at NNLO is likely underestimated, for reasons

Table 4.2: Integrated autocorrelation times τ achieved with HMC and `emcee` at LO, NLO, and NNLO.

Order	τ_{HMC}	$\tau_{\texttt{emcee}}$
LO	0.58	34
NLO	0.78	154
NNLO	1.2	> 111

that will be discussed in the next section. As indicated by Figure 4.6, we see $\tau < 1$ for the HMC samplings at LO and NLO; at LO we have $\tau = 0.58$ which will result in an ESS that is nearly twice as large as the length of the MCMC chain. Using Equation (4.1), the results from Table 4.2, and recorded information regarding the exact number of likelihood evaluations during the tuning and burn-in stages of the samplings, I can estimate the HMC speedups $\mathcal{S}$ at each chiral order to

$$\begin{aligned} \mathcal{S}_{\text{LO}} &= 5.9, \\ \mathcal{S}_{\text{NLO}} &= 5.2, \\ \mathcal{S}_{\text{NNLO}} &> 3.3. \end{aligned}$$

These speedups enable us to efficiently extract LEC posteriors and make more precise observable predictions than previously possible. Employing HMC sampling is indeed beneficial.

4.3.2 Convergence

Efficient sampling is nice—and, in many real world scenarios, required—but a fundamentally more important question is whether or not the result can be trusted not to change as more samples are added. In the world of MCMC the question of trust can be summarized by the crisp word convergence, and every MCMC study should include an assessment of convergence. Unfortunately it is not possible to prove convergence in most realistic settings, and the best we can do is to look for—and, if all is well, fail to find—signs of non-convergence. For notational simplicity I will use the word convergence throughout this section, but please keep in mind that when I say that a chain has converged, what I actually mean is that I have tried and failed to find evidence of non-convergence.

An array of convergence tests have been developed over the years, see e.g. Ref. [72] for an overview of methods. Some convergence measures, notably the

standard Gelman-Rubin diagnostic $\hat{R}$ [73, 74] and its derivatives, are premised on running multiple identically-prepared chains and looking for systematic differences, while others can be applied to individual chains. In this thesis I apply the $\hat{R}$ criterion and a single-chain diagnostic premised on monitoring the evolution of the integrated autocorrelation time τ [57]. I will discuss $\hat{R}$ first.

$\hat{R}$ is based on comparing variances both within and between MCMC chains that have different starting points but are otherwise identical. The method proceeds in four steps, which are repeated for each parameter (LEC) α_i:

1. Compute the between-chain variance B.
2. Compute the within-chain variance W.
3. Estimate the variance V of the marginal posterior for α_i.
4. Compare V and W to estimate the potential variance reduction.

Assuming that we have M chains of length N, B and W for the ith LEC can be calculated as

$$B = \frac{N}{M-1} \sum_{m=1}^{M} \left(\bar{\alpha}_i^{(m)} - \bar{\alpha}_i \right)^2 \tag{4.3}$$

$$W = \frac{1}{M} \sum_{m=1}^{M} \frac{1}{N-1} \sum_{n=1}^{N} \left(\alpha_i^{(nm)} - \alpha_i^{(m)} \right)^2 . \tag{4.4}$$

where the mean of all chains, $\bar{\alpha}_i$, is given by

$$\bar{\alpha}_i = \frac{1}{M} \sum_{m=1}^{M} \bar{\alpha_i}^{(m)}, \tag{4.5}$$

the within-chain mean for the mth chain, $\bar{\alpha_i}^{(m)}$, is given by

$$\bar{\alpha}_i^{(m)} = \frac{1}{N} \sum_{n=1}^{N} \alpha_i^{(nm)}, \tag{4.6}$$

and $\alpha_i^{(nm)}$ is the nth sample of the ith LEC in the mth MCMC chain. For finite chain lengths N, B will overestimate the marginal variance provided that the chains are initialized at overdispersed locations, i.e. with greater variability than the true posterior. Similarly, W will underestimate the marginal variance

because the full posterior has not been explored. A weighted average of B and W can therefore be used to estimate the variance of the marginal posterior for the LEC α_i according to

$$\mathrm{Var}^+[\alpha_i] = \frac{N-1}{N}W + \frac{1}{N}B \tag{4.7}$$

The plus sign indicates that this quantity overestimates the posterior variance provided that the M chains are initialized at overdispersed locations. Applying finite-N corrections leads to a Student's t distribution for α_i with variance estimated by

$$V = \mathrm{Var}^+[\alpha_i] + \frac{B}{MN}. \tag{4.8}$$

The Gelman-Rubin measure then expresses the potential scale (i.e. variance) reduction by forming the ratio

$$\hat{R} = \sqrt{\frac{V}{W}}, \tag{4.9}$$

which approaches 1 as $N \to \infty$. The Gelman-Rubin diagnostic declares convergence once $\hat{R}$ is sufficiently close to 1. Any threshold will necessarily be arbitrary; a commonly used value is $\hat{R} < 1.01$ [51].

Figure 4.7 shows the evolution of $\hat{R}$ for the HMC samplings (with burn-in removed) at LO ($M = 3$), NLO ($M = 10$), and NNLO ($M = 3$). Clearly, $\hat{R} < 1.01$ is achieved virtually as soon as the samplings have started and a stricter criterion of $\hat{R} < 1.001$ is also achieved. In short, no signs of non-convergence can be detected here, and there is also little reason to continue sampling for the purpose of reducing the variance. The $\hat{R}$ values for the `emcee` samplings are slightly larger, but still within the $\hat{R} < 1.01$ threshold.

Convergence may also be assessed by studying the evolution of τ over time. Accurately determining the integrated autocorrelation time requires a chain that is significantly longer than τ. If this is not the case, τ will generally be underestimated (with a corresponding overestimation of the ESS). This is the reason why $\tau_{\texttt{emcee}}$ at NNLO is reported as *greater than* 111 in Table 4.2; this chain's τ estimate has not yet stabilized. A τ-based convergence criterion asserts that the chain length N should be some multiple (typically a few tens of times) greater than the τ estimate; examples include $N \geq 10\tau$ [56] and $N \geq 50\tau$ [75]. I employ the latter criterion here.

Figure 4.8 shows the evolution of τ for both HMC and `emcee` at the three chiral orders, along with the convergence criterion $N \geq 50\tau$. Here I have only

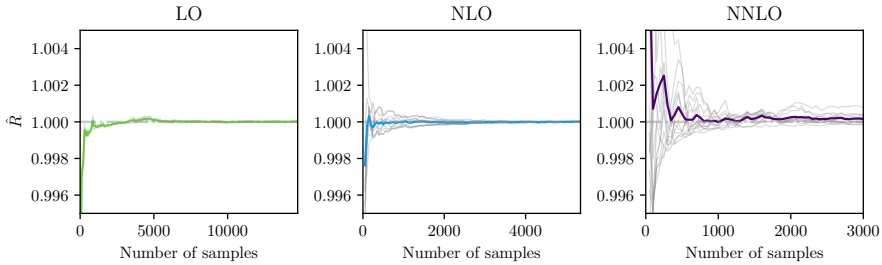

Figure 4.7: The Gelman-Rubin convergence diagnostics $\hat{R}$ corresponding to the HMC sampled chains at LO, NLO, and NNLO. $\hat{R}$ for each individual parameter is shown in gray, while the mean $\hat{R}$ is shown in green, blue, and purple, respectively.

plotted the results for a single chain for each case, but the results are similar for the identical chains used for computing $\hat{R}$. The HMC chains achieve convergence within a couple of hundred samples in each case and remain relatively stable at around $\tau = 1$, although it is noteworthy that τ_{NNLO} increases somewhat until around $N = 1000$. As for the `emcee` chains, it is obvious that τ is initially underestimated; the curves continue to rise deep into the samplings, and only at LO can convergence comfortably be declared. The NLO sampling manages to just cross the convergence threshold, while the NNLO sampling fails to achieve convergence although its τ estimate appears to be on the verge of stabilizing.

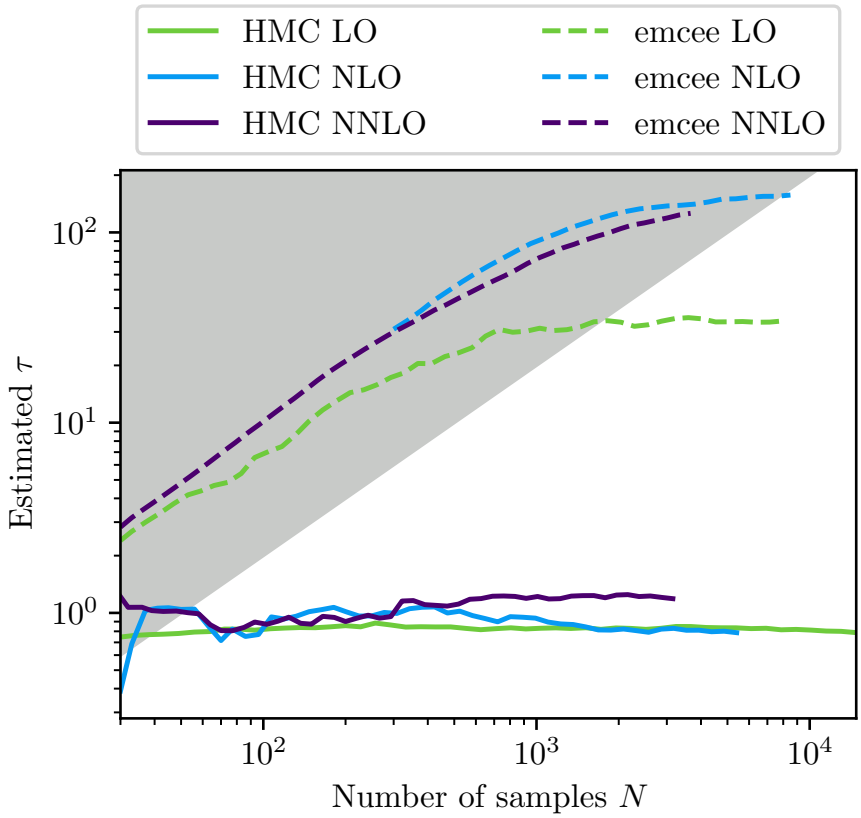

Figure 4.8: τ vs number of samples. The gray area indicates the "non-convergence zone", i.e. $N \geq 50\tau$. The `emcee` results are averaged over all walkers, i.e. to get the true number of collected samples the indicated number N should be multiplied by the number of walkers, which is 10 at LO, 50 at NLO, and 65 and NNLO.

Chapter 5

Model checking

We now have a statistical model of the strong nuclear interaction that may be used to produce posterior predictive distributions—ppds—for low-energy scattering observables. Like all models, it is wrong [76] due to approximations, omissions and incomplete understanding of the underlying physics. The important question is not, however, if the model is right or wrong, but rather if it is useful? In this chapter I investigate this question by extracting predictive distributions of 2N elastic scattering cross sections, checking the results against empirical data and higher-order results, and analyzing the quality of the statistical model for the EFT uncertainties.

5.1 Predictive distributions and credibility intervals

The procedure for sampling a predictive distribution is explained in Section 2.4. As a brief reminder, I model the true value $\widetilde{\mathcal{O}}$ of an observable as

$$\widetilde{\mathcal{O}}^{(k)} = \mathcal{O}^{(k)}_{\text{theo}} + \delta\mathcal{O}^{(k)}_{\text{theo}} \tag{5.1}$$

where $\mathcal{O}^{(k)}_{\text{theo}}$ is the observable value computed at chiral order k and $\delta\mathcal{O}^{(k)}_{\text{theo}}$ is the corresponding truncation error. The index k on $\widetilde{\mathcal{O}}^{(k)}$ indicates that the prediction is made at order k. Both terms on the right hand side are uncertain: $\mathcal{O}_{\text{theo}}$ due to parameter uncertainty, and $\delta\mathcal{O}_{\text{theo}}$ due to the truncation error being unknown. Sampling the pdfs of these stochastic variables is straightforward and—relative to the sampling of the LEC posteriors—undemanding

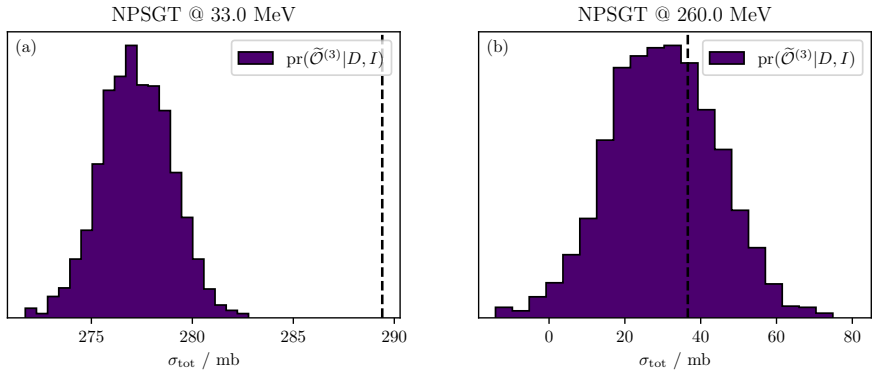

Figure 5.1: Extracted NNLO predictive distributions $\mathrm{pr}(\widetilde{\mathcal{O}}^{(3)}|\mathcal{D}_{\rm cal}, I)$ for the np total cross section at (a) $T_{\rm lab} = 33$ MeV and (b) $T_{\rm lab} = 260$ MeV. $N = 2000$ samples were used in both cases. The black dashed lines indicate experimental results gathered from Ref. [33].

thanks to the short (but information-packed) HMC chains and the closed-form expression (2.20) for the truncation error pdf.

Applying the procedure above for a given observable type, energy, and (if applicable) scattering angle yields a one-dimensional ppd, $\mathrm{pr}(\widetilde{\mathcal{O}}^{(k)}|D, I)$, wherein all available information that in principle summarizes all we know is contained. Variables that can be chosen freely are often called control variables; the primary control variables here will be the scattering (lab) energy $T_{\rm lab}$ and the scattering angle θ. Two NNLO examples $\mathrm{pr}(\widetilde{\mathcal{O}}^{(3)}|\mathcal{D}_{\rm cal}, I)$ are shown in Figure 5.1, where I have predicted the total np scattering cross section $\sigma_{\rm tot}$ (NPSGT) at (a) $T_{\rm lab} = 33$ MeV and (b) $T_{\rm lab} = 260$ MeV; these cross sections are both present in the validation data set $\mathcal{D}_{\rm val}$. The pdf in (b) is much wider than in (a) due to the increased truncation error at the higher energy, in accordance with the EFT expectation that $\delta\mathcal{O}_{\rm theo}$ increases with $T_{\rm lab}$. Note that the pdf in (b) stretches into negative values of $\sigma_{\rm tot}$; this is obviously unphysical and a more advanced statistical model for the truncation error should not allow it.

It is instructive to compare the contributions from the two terms in Equation (5.1). Figure 5.2 shows the same pdfs as Figure 5.1, but I have added pdfs extracted by letting $\delta\mathcal{O}^{(k)}_{\rm theo} = 0$, i.e. without any truncation error. The uncertainty in these (yellow-colored) pdfs comes from the uncertainty of the LECs only. By comparing figures (a) and (b) it is evident that the truncation

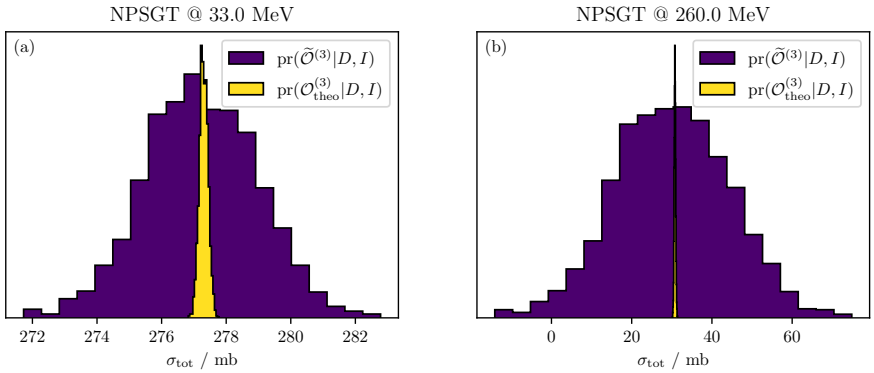

Figure 5.2: The purple pdfs $\mathrm{pr}(\widetilde{\mathcal{O}}^{(3)}|\mathcal{D}_{\mathrm{cal}}, I)$ are identical to Figure 5.1. The yellow pdfs $\mathrm{pr}(\mathcal{O}^{(3)}_{\mathrm{theo}}|D, I)$ are identical to the purple ones except that the truncation error has been neglected. Note that the maximum heights of the pdfs are arbitrary and chosen in such a way that the pdf shapes may be easily compared visually.

error becomes increasingly dominant at higher energies, and that the truncation error contributes the majority of the uncertainty in both cases. The parameter uncertainty provides a non-negligible contribution at 33 MeV, but the truncation error is totally dominating at 260 MeV. This indicates that extracting parameter pdfs is important for predicting low-energy cross sections, but that point estimates (e.g. MAP points found by optimizing the relevant posterior, eqivalent to approximating the pdf with a delta function) may suffice when predicting observables close to or above scattering energies equal to the pion production threshold. Neglecting the truncation error would however be inadvisable in either case.

A ppd like those shown in Figure 5.1 (a) or (b) is the final product of a study designed to predict an observable of intrinsic interest. Here, however, the main purpose of extracting ppds is to assess the validity of the statistical model of the EFT errors, and manual inspection of each ppd in its entirety is impractical as the validation data set $\mathcal{D}_{\mathrm{val}}$ encompasses over two thousand cross sections. Some quantity that summarizes the most important aspects of the ppd is therefore necessary. In frequentist statistics this is often achieved with a confidence interval. The Bayesian analog is called a credibility (or credible) interval (CI). Multiple choices of CIs exist, for example the so-called equal-tailed interval (ETI) in which there is equal probability mass below and

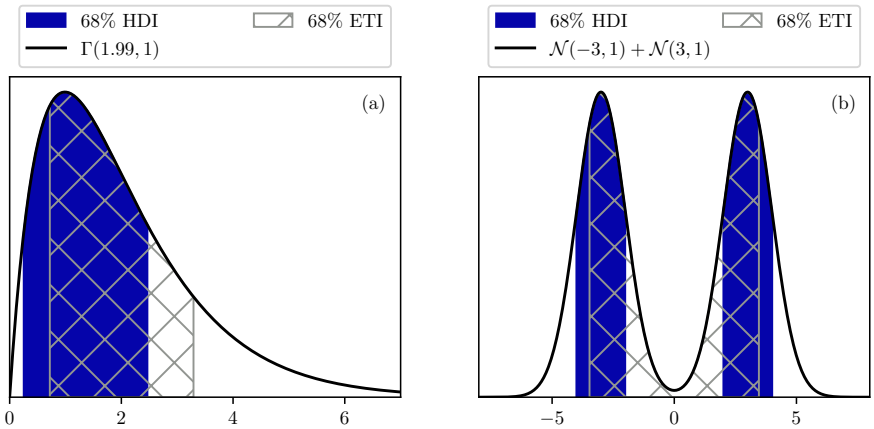

Figure 5.3: Differences between 68% ETIs and HDIs for two example pdfs. The pdfs are (a) $\Gamma(1.99, 1)$ and (b) $\mathcal{N}(-3, 1) + \mathcal{N}(3, 1)$.

above the interval, i.e. for a $(1 - a) \cdot 100\%$ ETI

$$\mathrm{pr}(X \leq q_a) = \mathrm{pr}(X \geq q_{1-a}) \tag{5.2}$$

where q_a and q_{1-a} are the a and $1-a$ quantiles of the pdf, respectively, and X is some stochastic variable. Another example is the highest posterior density interval (HPDI or HDI) [77], which in the one-dimensional case is the smallest possible interval encapsulating the specified probability mass. Visualizations of 68% ETIs and HDIs are shown in Figure 5.3 for two example pdfs. The ETI of the skewed gamma distribution in (a) excludes highly probable points to the left of the mode for the sake of keeping the tails equal, while the narrower HDI includes all the most probable points. In (b), the HDI becomes disjoint as it excludes the low probability region between the two modes. HDIs are in my opinion a natural choice for summarizing ppds of scattering observables as they do not exclude any high-probability regions of the ppd, and do exclude low-probability regions in case of multimodal ppds. I will hence employ HDIs throughout the remainder of this chapter. Figure 5.4 shows the same distribution as in Figure 5.1a with a corresponding 95% HDI shown in purple.

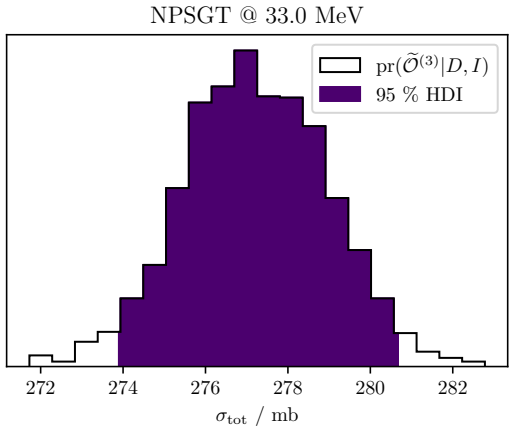

Figure 5.4: The same distribution as in Figure 5.1a with a 95% highest density interval indicated in purple.

5.2 Analyzing selected observables

At this point I am in a position to extract a predictive distribution for any low-energy *np* or *pp* elastic scattering cross section by applying the procedure outlined above for different observables and/or control variables T_{lab} and θ. Such distributions are of interest if a specific cross section is desired (e.g. if experimental data is unavailable), but more importantly for this work it allows me to check the validity of the statistical model by comparing the theoretical predictions with the $N = 2018$ data points in the validation data set $\mathcal{D}_{\mathrm{val}}$. If I, for example, compute a 50% HDI for each of the N data points, these HDIs should cover the empirical data roughly 50% of the time if the statistical model is working well. I will perform this exercise in a systematic way in the next section, but first I will present a few examples of ppds for specific observables across a limited range of control variables.

Figure 5.5 (a)-(b) shows predictions of the *np* total cross section across a wide range of energies with a separate ppd for each energy. In (a) I have plotted 100 predictive samples at each energy to give a visual representation of the ppds, and in (b) I have plotted 68% and 95% HDIs of the same ppds using 2000 samples to increase the precision of the intervals. All three studied chiral orders are included: LO in green, NLO in blue, and NNLO in purple. Also shown (in orange) are empirical data points with error bars, and

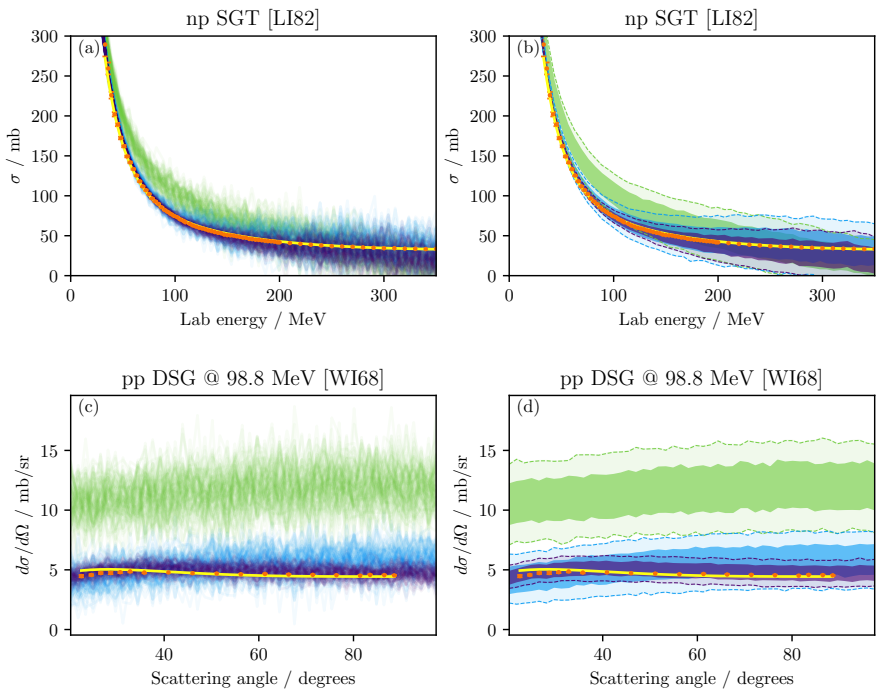

Figure 5.5: Posterior predictive distributions $\mathrm{pr}(\widetilde{\mathcal{O}}^{(k)}|\mathcal{D}_{\mathrm{cal}}, I)$ at LO (green), NLO (blue), and NNLO (purple) for two scattering observables. a) Uncorrelated samples of $\mathrm{pr}(\widetilde{\mathcal{O}}^{(k)}|\mathcal{D}_{\mathrm{cal}}, I)$ for the *np* total cross section. Empirical data with error bars is shown in orange and N3LO point estimates in yellow. b) 68% (dark shaded regions) and 95% (light shaded regions) HDIs of the ppds in a). c) Uncorrelated samples of $\mathrm{pr}(\widetilde{\mathcal{O}}^{(k)}|\mathcal{D}_{\mathrm{cal}}, I)$ for the *pp* differential cross section at $T_{\mathrm{lab}} = 98.8$ MeV. d) 68 and 95% HDIs of the ppds in c). The empirical data is gathered from the Granada database; see Tables II-III in Ref. [32] for the references indicated in brackets.

theoretical N3LO point estimates (yellow) calculated as $\mathcal{O}^{(4)}_{\text{theo}}(\vec{\alpha}_{\text{prel}})$ where $\vec{\alpha}_{\text{prel}}$ are preliminary LEC values found by optimization (see Section 2.3.1). The error bars for the empirical data are nearly invisible as the data is very precise. Note that the 95% HDI represented by the purple dashed region at $T_{\text{lab}} = 33$ MeV in Figure 5.5 (b) is also shown more clearly in Figure 5.4. As expected, the predictive distributions widen with increasing energy, and the predictions for these observables appear to converge to the empirical data with increased chiral order. Unlike NLO and NNLO, the LO MAP points differ significantly from the empirical results between (approximately) 50-200 MeV, indicating that this order does not include enough physics to accurately reproduce empirical results at higher energies. Other authors have drawn similar conclusions, see e.g. Ref. [18]. As we will see, the CIs here are overly confident and should be wider to reflect this lack of accuracy at LO. As expected, the NNLO HDIs are consistently narrower than the corresponding NLO intervals. The LO intervals show an interesting trend, however: they are significantly wider than their higher-order counterparts at lower energies, but are about the same width as at NNLO at high energies. This is not expected and indicates that the expansion coefficient $\bar{c}$ (see Table 2.2) has been underestimated, resulting in overly confident LO predictions at high energies in particular.

Figure 5.5 (c)-(d) show predictions analogous to (a)-(b), but for the *pp* differential cross section at 98.8 MeV across a range of scattering angles θ. The trends are similar to figures (a)-(b), with the predictions converging to the empirical results with increasing order. The LO predictions do not overlap the data in this case, while the NLO-NNLO predictions reproduce the data well. The NNLO results capture the subtle trends visible in both the empirical data and the N3LO point estimates. Figure 5.6 show predictions of spin observables at energies above the pion production threshold: *pp* spin correlations at 294.4 MeV in (a)-(b) and *np* beam polarization at 325 MeV in (c)-(d). The LO ppds are remarkably narrow in both cases, reinforcing the suspicion that the truncation error is underestimated. Figure (b) also shows that the NNLO predictions reproduce empirical data better than the N3LO point estimates in this particular case.

5.3 A frequentist probability makes a surprise appearance

The ppd plots presented in the previous section provide valuable hints and insights about how different observables are reproduced, but are the ppds

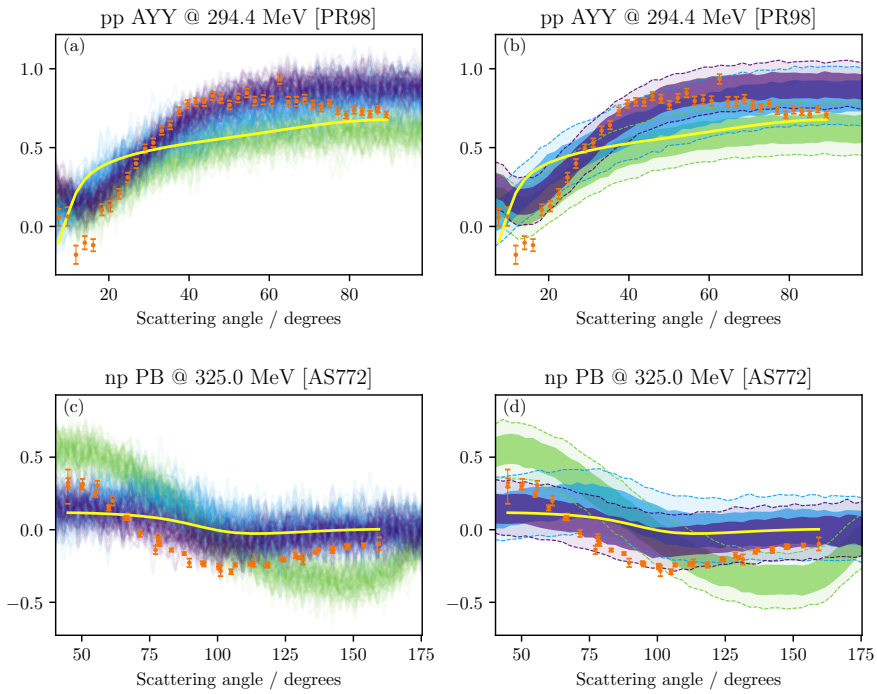

Figure 5.6: Posterior predictive distributions $\mathrm{pr}(\widetilde{\mathcal{O}}^{(k)}|\mathcal{D}_{\mathrm{cal}}, I)$ at LO (green), NLO (blue), and NNLO (purple) for two scattering observables. a) Uncorrelated samples of $\mathrm{pr}(\widetilde{\mathcal{O}}^{(k)}|\mathcal{D}_{\mathrm{cal}}, I)$ for the *pp* A_y spin correlation at 294.4 MeV. Empirical data with error bars is shown in orange and N3LO point estimates in yellow. b) 68% (dark shaded regions) and 95% (light shaded regions) HDIs of the ppds in a). c) Uncorrelated samples of $\mathrm{pr}(\widetilde{\mathcal{O}}^{(k)}|\mathcal{D}_{\mathrm{cal}}, I)$ for the *np* beam polarization at $T_{\mathrm{lab}} = 325$ MeV. d) 68 and 95% HDIs of the ppds in c). The empirical data is gathered from the Granada database; see Tables II-III in Ref. [32] for the references indicated in brackets.

actually reliable? To answer this question I investigate the so-called coverage probability of the credibility intervals (see e.g. Ref. [78] for an introduction with examples). As mentioned previously, a $100 \cdot p\%$ CI should cover about $100 \cdot p\%$ of the corresponding empirical data in order to be reliable. A coverage probability plot is generated by systematically computing CIs of probability p for predictions corresponding to each datum in a data set and checking whether each interval covers the datum. The CI probability p is then varied between 0-1. The resulting coverage probabilities $\tilde{p}$ are plotted against the specified probability p, and should ideally be equal to each other for all p.

The coverage probabilities are shown in Figure 5.7 (a) using the usual color coding: green for LO, blue for NLO, purple for NNLO. As suspected based on the prediction plots in the previous section, the LO truncation error is considerably underestimated, which can be seen by observing that the LO HDI coverage falls below the $p \cdot 100\%$ expectation. The NLO and NNLO predictions are also overconfident, but to a lesser degree. The root cause can be traced to the method of computing $\bar{c}(k)$; recall that $\bar{c}$ for the truncation error at order k is calculated as the RMS value of all expansion coefficients $\vec{c}_n$ from order $n = 0$ to order $n = k$. The issue here is that the coefficients $\vec{c}_0$ are uninformed by order-by-order differences, and essentially determined by the choice of $\mathcal{O}_{\text{ref}}$. With the particular choice of reference scale I employ, this results in $\bar{c}(0) = 1.17$, but by inspecting the values of $\vec{c}_n$ where $n > 0$ we expect that a more realistic value for $\bar{c}$ is around 3-4, i.e. the truncation error at LO should probably be about 3-4 times larger than currently specified. The inclusion of $\vec{c}_0$ in the RMS determination of $\bar{c}$ at the higher orders effectively reduces the truncation error. The effect is partly counteracted by the larger expansion coefficients at orders 2 and 3.

The observed success rates should follow a binomial distribution under the assumption that the observables are uncorrelated. We use the continuous version of the binomial distribution, the β distribution, to assign a likelihood for the proposition that p is the true success rate and compute 95% confidence intervals for this likelihood. These confidence intervals are shown in gray in Figure 5.7. The observed success rates should largely overlap with the computed confidence intervals if the observed success rates are consistent (at the 2σ level) with a true success rate of p [21]. In Figure 5.7 (a), this is not the case, and we can conclude that the true success rate is lower than p. However, it should be noted that the assumption that all observables are uncorrelated is not quite realistic.

It is prudent to investigate the effect of using different methods of computing $\bar{c}$. Ideally, the LEC posteriors should be resampled after each modification

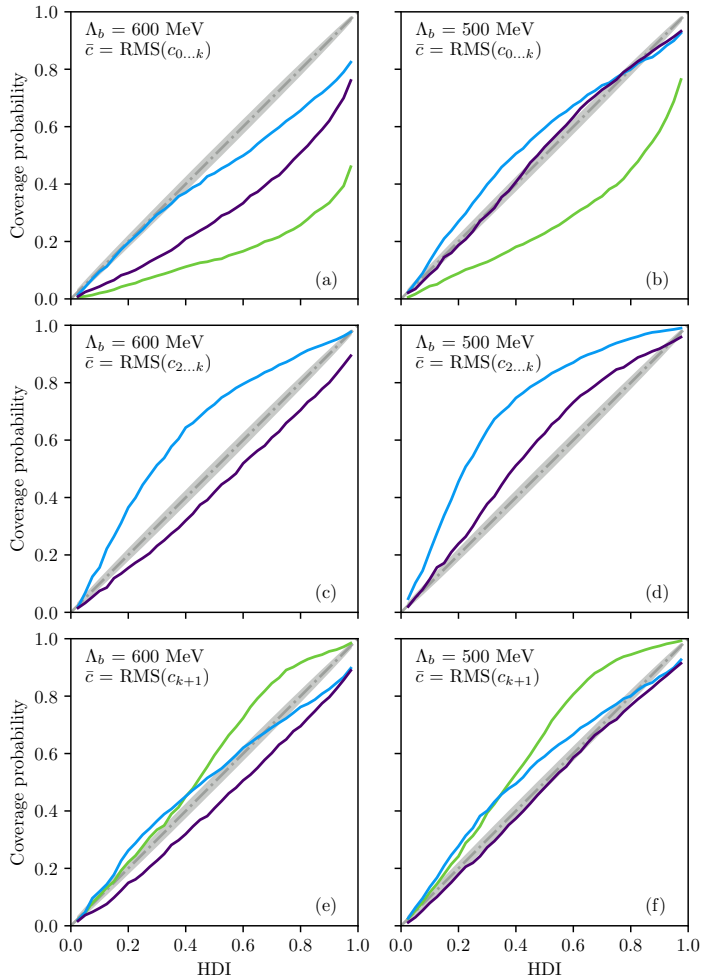

Figure 5.7: Coverage probabilities of predictions at LO (green), NLO (blue), and NNLO (purple) of the observables in $\mathcal{D}_{\text{val}}$. The first column was obtained with the assumption $\Lambda_b = 600$ MeV and the second with $\Lambda_b = 500$ MeV. Each row corresponds to different methods of computing $\bar{c}(k)$; $\bar{c}$ is computed from the RMS values of c at (1) orders $0, \ldots, k$ (2) orders $2, \ldots, k$ (3) order $k+1$. The computed $\bar{c}$ values are (a) $\bar{c}(0) = 1.17, \bar{c}(2) = 2.08, \bar{c}(3) = 2.72$ (b) $\bar{c}(0) = 1.17, \bar{c}(2) = 1.65, \bar{c}(3) = 2.22$ (c) $\bar{c}(2) = 4.95, \bar{c}(3) = 4.19$ (d) $\bar{c}(2) = 3.44, \bar{c}(3) = 2.77$ (e) $\bar{c}(0) = 4.95, \bar{c}(2) = 2.72, \bar{c}(3) = 4.12$ (f) $\bar{c}(0) = 3.44, \bar{c}(2) = 1.65, \bar{c}(3) = 1.99$.

of $\bar{c}$, but this is very time consuming. Fortunately, since the LEC uncertainties are relatively small compared to the truncation error for an overall majority of the data (see e.g. Figure 5.2), I merely resample the pdf for the truncation error. Figure 5.7 (b) shows the effect of excluding $\vec{c}_0$ from the calculations of $\bar{c}(k)$, i.e. letting $\bar{c}(k) = \mathrm{RMS}(\vec{c}_{2,\ldots,k})$. This method is obviously inapplicable at LO so only results for NLO and NNLO are shown. This immediately improves the predictive power of the statistical model. Indeed, the NLO predictions are a bit conservative, but the NNLO predictions are very close to the ideal, although still not quite consistent with a true success rate p according to the gray confidence intervals. The overly conservative predictions at NLO may be understood by inspecting Table 2.2. The RMS value of the expansion coefficients $\vec{c}_2$, here used to characterize the truncation error at NLO, are rather large ($\bar{c}_2 = 4.95$). But the expansion coefficients $\vec{c}_3$, which better represent this truncation error, are smaller, indicating that $\bar{c} = 4.95$ is an overestimation. I have also computed $\bar{c}_4 = \mathrm{RMS}(\vec{c}_4) = 4.12$, which together with the lower-order results indicates that the size of the truncation error tends to oscillate: odd-order contributions are small relative to even orders, where new contact LECs are introduced. This somewhat irregular convergence pattern of the chiral expansion, which complicates the estimation of $\bar{c}$, calls into question the assumption I make that all expansion coefficients are drawn from the same underlying distribution.

Inspired by the reasoning above I have tried one more method for computing $\bar{c}$: using the RMS value of the expansion coefficients $\vec{c}_{k+1}$, i.e. the first omitted order. Intuitively, this is the most logical choice as it directly estimates the contribution of the first omitted order, which provides the largest contribution to the overall truncation error. As can be seen in Figure 5.7 (c), this method further improves the performance of the statistical model, with coverages that are close to the ideal diagonal curve; the NNLO coverages, in particular, largely overlap with the 95% confidence intervals for a true success rate p. If possible, this method should be preferred over the others, but information from higher orders are nearly always unavailable. If so, the method showcased in (b) should be preferred.

The breakdown scale Λ_b is an important property of the physical system and it is of interest to see the sensitivity of predictions to different assumptions about this quantity. In Figures 5.7 (d)-(f) I repeat all the coverage probability calculations described thus far, with one important difference: I have scaled the EFT breakdown scale by introducing a dimensionless scaling factor λ_b according to [21]

$$\Lambda_b = \lambda_b \Lambda_b^{\mathrm{def}}. \tag{5.3}$$

where Λ_b^{def} is the (in this work) default value $\Lambda_b^{\text{def}} = 600$ MeV. By inspecting Equation (2.18) it is easy to see that decreasing (increasing) Λ_b yields smaller (larger) values for the expansion coefficients, since (2.18) can be rewritten as

$$c_n = \frac{\mathcal{O}_{\text{theo}}^{(n)}(\vec{\alpha}_{\text{prel}}) - \mathcal{O}_{\text{theo}}^{(n-1)}(\vec{\alpha}_{\text{prel}})}{\mathcal{O}_{\text{ref}} \cdot (\frac{\max(m_\pi, p)}{\lambda_b \Lambda_b^{\text{def}}})^n} = (\lambda_b \Lambda_b^{\text{def}})^n \frac{\mathcal{O}_{\text{theo}}^{(n)}(\vec{\alpha}_{\text{prel}}) - \mathcal{O}_{\text{theo}}^{(n-1)}(\vec{\alpha}_{\text{prel}})}{\mathcal{O}_{\text{ref}} \cdot \max(m_\pi, p)^n}, \tag{5.4}$$

i.e. the expansion coefficients are rescaled by a factor λ_b^n. The calculated expansion coefficients have so far been rather large yet resulting in overconfident ppds, suggesting that $\lambda_b = 1$ may be too high. This finding is somewhat at odds with Ref. [21], but direct comparison is difficult not least because the validation data set used here is orders of magnitude larger in size. Melendez et al. [23] expanded the analysis in Ref. [21] and found that $\lambda_b = 1$ generally performs fairly well, without ruling out other values. To see the effect of decreasing λ_b, I let $\lambda_b = 5/6$, yielding the rescaled breakdown scale $\Lambda_b = 500$ MeV and $\bar{c}$ values closer to the naturalness expectation. The size of the truncation error has a complex dependency on Λ_b according to Equations (2.20) and (2.21), since the latter can be rewritten as

$$(\sigma_{\text{theo}}^{(k)})^2 = \frac{\bar{c}^2 \mathcal{O}_{\text{ref}}^2 \cdot \max(m_\pi, p)^{2(k+1)}}{(\Lambda_b)^{2(k+1)} - (\Lambda_b)^{2k} \max(m_\pi, p)^2}. \tag{5.5}$$

Figures 5.7 (d)-(f) show that the overall performance of the statistical model improves with the rescaled breakdown scale, especially at NNLO. It also decreases the model's sensitivity to the exact value of $\bar{c}$. This indicates that the true EFT breakdown scale may be smaller than 600 MeV. Rather than taking Λ_b to be a given quantity and checking whether the predictive ppds are consistent with data, Melendez et al. [23] developed a method for inferring Λ_b by deriving Bayesian posteriors for Λ_b. Such posteriors, conditioned on extracted expansion coefficients at predetermined kinematic points, can then be sampled to yield a pdf for the EFT breakdown scale. In Paper A, we also extracted a pdf $\text{pr}(\bar{c}, Q|D, I)$ for $\bar{c}$ and Q.

It is instructive to study how the statistical model performs with regards to different subsets of the validation data. This is shown in Figure 5.8, where I have used $\Lambda_b = 600$ MeV and $\bar{c}(k) = \text{RMS}(\vec{c}_{2,\ldots,k})$. In (a), only total cross sections are considered. We see that the credibility intervals appear too conservative here, but keep in mind that the size of this data subset is relatively small with just 84 points, accounting for 4.2% of the full validation data set $\mathcal{D}_{\text{val}}$. The points are also highly correlated since (e.g.)

$\sigma_{\text{tot}}(198\text{ MeV}) \approx \sigma_{\text{tot}}(200\text{ MeV})$ and the precision of the data is high. Subfigures (b) and (c) take into account data below 100 MeV and above 290 MeV, respectively; overall, the coverages are similar, indicating that the statistical model works roughly equally well in both energy ranges. The biggest difference is that the NLO HDIs are overly conservative at high energies. Further increasing the granularity beyond what is shown here, e.g. by studying each observable type (see Table 2.1) separately, do not reveal much more in the way of systematic differences beyond what is shown in Figure 5.8. The NLO ppds are consistently too conservative, while the NNLO ppds are generally somewhat overconfident. An intriguing exception to this is that the NNLO coverage of DSG observables (figure omitted) follow the ideal diagonal curve, i.e. these predictions are neither too confident or conservative. Tracing the cause of this result is difficult; it may well be a coincidence. Another possibility may be that the prescription for the reference value of DSG observables is more reliable than that of spin observables (recall that I use $\mathcal{O}_{\text{ref}} = 0.15$ for spin observables and $\mathcal{O}_{\text{ref}} = \mathcal{O}_{\text{LO}}$ for non-spin observables). Of course, I also use $\mathcal{O}_{\text{ref}} = \mathcal{O}_{\text{LO}}$ for SGT observables, but as noted previously these data are few and strongly correlated. The DSG data are far more numerous and the corresponding coverages are thus supported by more statistics.

Finally, as a central feature of EFTs is the systematic improvability, it is of interest to see not only if the predictions are consistent with empirical data, but also with higher-order point estimates. To do this I replicate Figure 5.7 (b) but exchange the empirical data for N3LO point estimates ("N3LO data"). The coverage probabilities of the NLO ppds are indistinguishable between Figures 5.7 (b) and 5.9, which is in line with expectations. There is, however, a notable difference regarding the NNLO coverages; where the coverage probabilities against empirical data are consistently slightly too low, their equivalents versus N3LO data are too high at HDI probabilities below 50% and too low above 50%. Here, a closer look at individual observable types reveal strong systematic trends. The coverage probability is very high for differential cross sections (DSG) and spin polarizations (P, PB), but very low for spin correlations (e.g. AXX, AYY). This indicates that the N3LO results align closely with the NNLO predictions for the DSG, P, and PB observables, while they differ substantially from both the NNLO predictions and the empirical data for e.g. AXX, AYY, and AZX. Figure 5.6 shows an explicit example of this trend. The conclusion that can be drawn here is that the expected convergence pattern from NNLO to N3LO fails to materialize in this case. The probable cause is that the utilized LECs $\vec{\alpha}_{\text{prel}}$ at N3LO lead to relatively poor predictions of spin correlation observables.

Overall, the reliability of the statistical model is high considering its simplicity, as long as $\bar{c}$ is modelled with some care. While multiple improvements can (and should) be implemented (see Chapter 6), the model is indeed useful provided that $\bar{c}$ is estimated without contamination from the uninformative first order expansion coefficients $\vec{c}_0$.

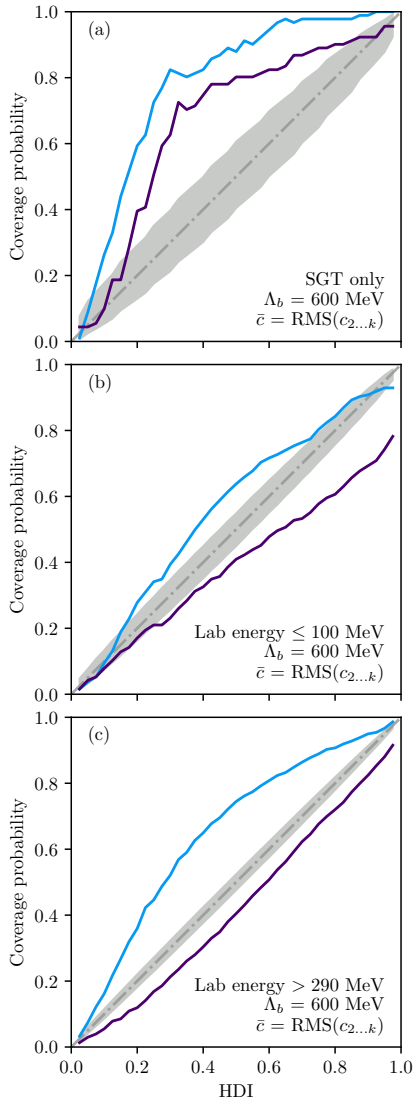

Figure 5.8: Coverage probabilities of predictions at NLO (blue) and NNLO (purple) of subsets of the observables in $\mathcal{D}_{\text{val}}$. I compute $\bar{c}$ as $\mathrm{RMS}(\vec{c}_{2,\ldots,k})$ as in Figures 5.7 (c) and (d) and use $\Lambda_b = 600$ MeV. (a) Only SGT observables, (b) all validation data with $T_{\text{lab}} \leq 100$ MeV, (c) all validation data with $T_{\text{lab}} > 290$ MeV.

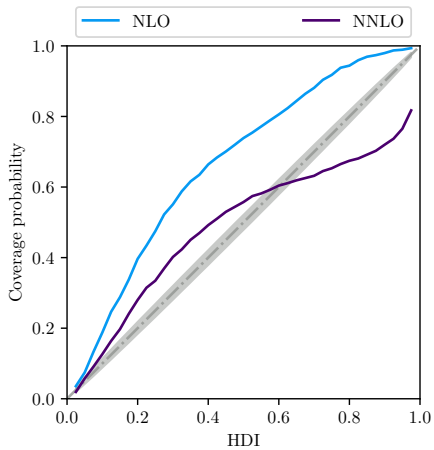

Figure 5.9: Coverage probabilities corresponding to those in Figure 5.7 (c), except that the coverages are computed against N3LO point estimates (or "N3LO data" for short) rather than empirical data. The N3LO data types match the empirical data $\mathcal{D}_{\text{val}}$ in every way (i.e. quantity, observable types, control variables, etc.), except that the value of each datum is a theoretical prediction at N3LO instead of an experimental result.

Chapter 6

Conclusion and outlook

The χEFT model of the nuclear interaction potential [4, 7–13] encompasses multiple sources of uncertainty, all of which influence predictions of low-energy nuclear observables. In this work I employ Bayesian analysis to define posterior distributions of the unknown LECs that govern interaction strengths at LO, NLO, and NNLO in Weinberg power counting. These posteriors, conditioned on experimental data and other information and assumptions, take into account the two dominant sources of uncertainty in χEFT: the uncertainty of the data and the theoretical uncertainty arising from the truncation of the EFT expansion at a finite order. I model the truncation errors for different observables as uncorrelated random variables drawn from the same underlying distribution [21, 24]. Once defined, the posteriors need to be numerically evaluated. Straightforward function evaluation of the posteriors, i.e. random sampling without any guidance, becomes futile beyond LO due to the rapid increase in dimensionality of the LEC domain, forcing us to turn to MCMC sampling [52, 79]. With the intent of decreasing sample autocorrelations and increasing efficiency I investigate the prowess of HMC [27, 28] for sampling the LEC posteriors. Equipped with the LEC posteriors I produce ppds for low-energy scattering cross sections, and perform model checking by comparing theoretical predictions with experimental data.

I find that using HMC brings significantly increased sampling efficiency at all three studied chiral orders, resulting in runtimes that—for a given sampling error—are up to six times shorter than if `emcee` [55, 56], a popular implementation of an affine invariant ensemble MCMC algorithm, is used. These gains enable us to either

1. sample more challenging pdfs—e.g. LEC posteriors with more parameters or more included diagrams—than previously possible, or

2. increase the statistical precision of previously extracted pdfs.

The increase in efficiency emerges from HMC's ability to move long distances through the parameter space without straying into low-probability areas while simultaneously maintaining a high acceptance probability for the proposed samples. These great strides diminish autocorrelations between subsequent samples, which leads to a large ESS. In order to gauge the sampler's efficiency, the increase in ESS must then be weighed against the large per-sample cost incurred by the need to solve Hamilton's equations for each proposed MCMC step, which in turn requires access to gradients of the posterior with regards to the LECs. My results show that, with proper tuning, the benefit of shorter autocorrelation times outweigh the increased per-sample cost. Furthermore, the benefit of using HMC is expected to increase along with the dimensionality of the posterior, as HMC theoretically scales better than other algorithms in this regard [27]. However, it can be a rather formidable task to tune the hyperparameters of HMC algorithm in practice, especially finding a suitable mass matrix $\mathcal{M}$. Successful tuning cannot be guaranteed a priori, which introduces an element of risk attached to the deployment of HMC. Employing an easier-to-use algorithm, e.g. `emcee`, may therefore be advisable as a safer option for less challenging sampling problems, e.g. the LO posterior in this work, even though the sampling efficiency is likely to suffer.

The successfully extracted LEC posteriors allow me to inspect to what extent the statistical model for the EFT errors yields predictive distributions consistent with empirical data, quantified using an empirical coverage measure. The employed model makes simplifying assumptions especially regarding the correlation structure of the expansion coefficients $\vec{c}_n$ that are introduced to characterize the EFT truncation error, and it is not obvious at the outset whether these simplifications are acceptable. My results indicate that the extracted observable ppds are approximately consistent with empirical data if the method used to estimate $\bar{c}$—the width of the Gaussian distribution from which all EFT expansion coefficients are assumed to be drawn—is modified so as to not be influenced by the zeroth-order expansion coefficients $\vec{c}_0$. In practice, this modification prevents the predictions from being overly confident. The predictive power of the statistical model is overall consistent across different types of low-energy scattering observables, energies, and scattering angles. The truncation error contributes greatly to the total error.

Several avenues of further research based on this work are possible. The most obvious outlook is to extend the study to extract LEC posteriors and

observable ppds at N3LO; all methods presented here are directly applicable to this problem, and even higher orders. Sampling the N3LO LEC posterior is a formidable challenge due to the large number of LECs at this order, which is ~ 30. I expect HMC to be able to efficiently sample the posterior assuming that an appropriate mass matrix can be extracted. Another improvement is to employ Gaussian processes to model the correlation structure of the expansion coefficients $\vec{c}_n$ [25] and infer, rather than prescribe a priori, the EFT breakdown scale Λ_b [23].

In my experience, the most important step for a successful application of HMC is the choice of a suitable mass matrix $\mathcal{M}$. I have primarily used HMC itself, with a preliminary mass matrix, to extract $\mathcal{M}$. However, other methods may be more efficient and/or reliable. For example, a short `emcee` or M-H sampling could potentially provide an excellent mass matrix. Another interesting idea is to use optimization and local second-order derivatives to extract an approximation of the parameter covariance matrix and use (the inverse of) this matrix as $\mathcal{M}$; this option bypasses the need for extensive preliminary samplings and could be very efficient. The extracted covariance matrix could potentially even be used to approximate the gradient of the posterior, required to solve Hamilton's equations; the efficiency gain would be enormous if this works. I have tried this approach briefly and the results are promising.

In Paper B, all empirical data on which the likelihoods are conditioned are *np* and *pp* scattering cross sections. First, this may be extended to include an *nn* scattering length and effective range, such that we can extract the $\widetilde{C}_{1S0}^{nn}$ LEC. Second, the inclusion of 3N scattering data would be very interesting, but evaluating such a likelihood is expensive; the efficiency gains $\mathcal{S}$ obtained with HMC (see Section 4.3.1) are instrumental for a successful sampling of such posteriors. The computational burden may be eased by employing efficient methods, such as emulators [80] or wave-packet continuum discretization [81], for evaluating the likelihood. Finally, the posteriors may be conditioned on bound state data, but it is currently unclear how to model the corresponding truncation error as bound states do not have well-defined external momenta.

An exciting prospect is the ability to extract reliable ppds for observables that cannot readily be experimentally measured, such as the *pp* fusion cross section [50]. Clearly, quantifying relevant LEC posteriors for models of the 2N and 3N interaction is crucial for computational statistics analyses of exotic nuclei and infinite nuclear matter.

Appendix A

montepython

Here I present a brief overview of the `montepython` package that I have developed. It contains the HMC sampler used throughout this thesis.

A.1 Installation

The package is hosted on GitHub. To acquire a copy, clone the repository:

```
git clone https://github.com/svisak/montepython.git
```

To install, go to the base directory and run

```
pip install .
```

or

```
pip install . --user
```

This will install `montepython` to your site-packages directory, which is in the Python path. If you do not want a system-wide (or user-wide) installation you can simply put the package in the same folder as your script.

A.2 Design

The code is designed to be easily extendable to incorporate different MCMC algorithms. At the time of writing, two algorithms are implemented: HMC

and Metropolis-Hastings (denoted `RWM`, short for "Random Walk Metropolis"). Despite the significant differences between these two algorithms, both share the same basic code, implemented in the abstract base class `MCMC`. The main user-facing method is `MCMC.run` which runs the sampler for a specified number of iterations, or alternatively the equivalent `MCMC.run_for` method which runs the sampler for a specified period of time. The `MCMC.run` method repeatedly calls the `MCMC.sample` method which proposes new samples, applies the Metropolis rejection criterion, and stores the resulting sample. The `MCMC` class also provides a variety of user-accessible methods to make the package user-friendly, including methods to access the MCMC chain, acceptance rates, and run time, and a method to save the MCMC chain to disk along with relevant metadata (e.g. the used mass matrix).

Specific MCMC algorithms are implemented by creating classes that extend the `MCMC` class and implement the abstract methods `MCMC.propose_state`, which as the name suggests proposes a new state, and `MCMC.joint_lnprob`, which returns the (log) probability of a state (a state is essentially a sample), see e.g. Equation (3.13). The log probability of a state is used to vet proposed states via the Metropolis rejection criterion.

The HMC algorithm is implemented in the `HMC` class. `HMC` creates a `Leapfrog` object, which implements the leapfrog integrator as defined in Section 3.4.4, in its constructor. The `HMC.propose_state` method draws a random momentum, calls the `Leapfrog.solve` method, and returns the proposed joint position-momentum state, which is then handed over to the `MCMC.sample` method to be either accepted or rejected. The link between energy and probability, see Section 3.4.1, is implemented in the `HMC.joint_lnprob` method.

The interface to define the posterior $\mathrm{pr}(\vec{\alpha}|D, I)$ is implemented in the abstract `Bayes` class. The user must extend this class and implement the `Bayes.evaluate` method. In this method, which takes a parameter vector $\vec{\alpha}$ as its only argument and is called repeatedly by the `Leapfrog.solve` method, the user is responsible for providing the log prior, log likelihood, and—in the case of HMC—negative log posterior values for different values of the parameter vector $\vec{\alpha}$. This approach gives the user a lot of flexibility regarding how the posterior is evaluated and in the case of HMC makes it easy to incorporate e.g. necessary derivatives from external applications in the gradient calculation.

A.3 Usage

This example demonstrates how to sample a multivariate Gaussian distribution. (See also the `README` file in the git repository, which contains a similar

example.) First import packages:

```
import montepython
import numpy as np
```

Then define the distribution to be sampled by extending the abstract base class Bayes:

```
# This class can be given any descriptive name as desired
class MultivariateGaussian(montepython.Bayes):

# Optional custom constructor
def __init__(self, mu, cov):
    super().__init__()
    self.mu = mu
    self.cov = cov
    self.cov_inv = np.linalg.inv(cov)

# This method is mandatory,
# Note how values are set using the
# self.set_* methods
def evaluate(self, alpha):
    # Prior, in this case an infinite uniform prior
    self.set_lnprior_value(0)

    # Likelihood
    diff = alpha - self.mu
    lnlikelihood_value = -0.5 * diff.T @ self.cov_inv @ diff
    self.set_lnlikelihood_value(lnlikelihood_value)

    # Gradient of the negative log posterior (HMC only)
    nlp_gradient_value = self.cov_inv @ diff
    self.set_nlp_gradient_value(nlp_gradient_value)
```

Note that the `nlp_gradient_value` is an array. It is only used in the HMC sampler, and not in the RWM sampler. Next instantiate the Bayes object:

```
ndim = 2 # Specify a bivariate N(0,1) distribution
mu = np.zeros(ndim)
cov = np.eye(ndim)
bayes = MultivariateGaussian(mu, cov)
```

Next, set run parameters. The `mass_matrix` specified here is the default one and only included for illustration purposes.

```
startpos = np.zeros(ndim)
ell = 20 # number of leapfrog steps
epsilon = 0.3 # leapfrog step size
mass_matrix = np.eye(ndim)
```

Set `stepsize` (scalar or matrix) instead of `ell`, `epsilon`, and `mass_matrix` if you are running RWM. Now initialize the sampler:

```
hmc = montepython.HMC(bayes,
                      startpos,
                      leapfrog_ell=ell,
                      leapfrog_epsilon=epsilon,
                      mass_matrix=mass_matrix)
# A vanilla Metropolis-Hastings sampler,
# montepython.RWM, is also available
```

And run it for e.g. 100 iterations:

```
hmc.run(100)
```

If it turns out that 100 samples was not enough, you can continue where you left off:

```
hmc.run(2000)
```

The length of the chain is now 2100. Alternatively, you can specify that the sampler runs for a specified amount of time, e.g.

```
hmc.run_for(1) # Default unit is 'hours'
hmc.run_for(2.5, unit='minutes')
hmc.run_for(30, unit='seconds')
```

The sampler will then have run for one hour and three minutes. The length of the chain will depend on the computational expense of each sample. You can then access some information about the chain:

```
print(hmc.acceptance_rate())
print(hmc.ndim())
chain = hmc.chain()
```

The MCMC samples are contained in `chain`, which is a NumPy `ndarray` of shape `(nsamples,ndim)` where `nsamples` is the number of collected samples. See the methods in the MCMC class for other information that you can access. You can optionally save the chain to disk in HDF5 format:

```
metadata = {}
metadata['some_info'] = 'HMC sampling of a Gaussian'
hmc.to_disk(**metadata)
# or simply hmc.to_disk(some_info='My info')
```

This saves the MCMC chain in a file called `hdf5/hmc.hdf5` assuming that the HMC algorithm is used. The (optional) dictionary `metadata` can hold many types of variables, such as strings, scalars, lists and arrays. In addition to the MCMC samples themselves, the `to_disk` method also saves a variety of metadata by default, such as acceptance rate, run time, and so on. You can set the HDF5 path, filename and dataset name:

```
path = 'path/to'
filename = 'toy_example.h5'
dataset_name = 'gauss2d'
hmc.to_disk(path=path,
            filename=filename,
            dataset_name=dataset_name)
```

If the dataset name conflicts with an existing dataset the default dataset name (based on the MCMC type and a timestamp) will be used instead. The chain can subsequently be retrieved from disk:

```
import h5py
f = h5py.File('path/to/toy_example.h5', 'r')
dset = f.get(dataset_name)
chain = dset[...]
acceptance_rate = dset.attrs['acceptance_rate']
print(acceptance_rate)
```

You can use e.g. `prettyplease`[1] to visualize the result:

```
import matplotlib.pyplot as plt
import prettyplease
fig = prettyplease.corner(chain)
plt.show()
```

[1] `https://github.com/svisak/prettyplease.git`

Appendix B

Covariance matrix for the uncorrelated EFT error

Let $X_1, \ldots, X_N$ be N independent, normally distributed random variables with mean 0 and variances $\sigma_1^2, \ldots, \sigma_N^2$. Then the sum $Y = \sum_{i=1}^{N} X_i$ is distributed according to

$$\mathrm{pr}(Y) = \mathcal{N}\left(0, \sum_{i=1}^{N} \sigma_i^2\right). \tag{B.1}$$

More generally, if we multiply each random variable X_i by a constant λ_i, the sum is distributed as

$$\mathrm{pr}\left(\sum_{i=1}^{N} \lambda_i X_i\right) = \mathcal{N}\left(0, \sum_{i=1}^{N} \lambda_i^2 \sigma_i^2\right) \tag{B.2}$$

and if $\sigma_i = \sigma, i = 1, \ldots, N$, then

$$\mathrm{pr}\left(\sum_{i=1}^{N} \lambda_i X_i\right) = \mathcal{N}\left(0, \sigma^2 \sum_{i=1}^{N} \lambda_i^2\right). \tag{B.3}$$

Consider now the EFT truncation error introduced in Chapter 2

$$\delta\mathcal{O}_{\mathrm{theo}}^{(k)} = \mathcal{O}_{\mathrm{ref}} \sum_{n=k+1}^{\infty} c_n Q^n \tag{B.4}$$

where each c_n is drawn from a normal distribution according to

$$\mathrm{pr}(c_n|\bar{c}) = \mathcal{N}(0, \bar{c}^2). \tag{B.5}$$

Following Equation (B.3), $\delta\mathcal{O}^{(k)}_{\mathrm{theo}}$ is distributed as

$$\mathrm{pr}(\delta\mathcal{O}^{(k)}_{\mathrm{theo}}|\bar{c}, Q) = \mathcal{N}\left(0, \bar{c}^2\mathcal{O}^2_{\mathrm{ref}} \sum_{n=k+1}^{\infty} Q^{2n}\right) \tag{B.6}$$

With

$$\Lambda = \mathcal{O}^2_{\mathrm{ref}} \sum_{n=k+1}^{N} Q^{2n} \tag{B.7}$$

we have

$$\Lambda - Q^2\Lambda = \mathcal{O}^2_{\mathrm{ref}}Q^{2(k+1)} - \mathcal{O}^2_{\mathrm{ref}}Q^{2(N+1)} \tag{B.8}$$

$$\Leftrightarrow \Lambda(1 - Q^2) = \mathcal{O}^2_{\mathrm{ref}}Q^{2(k+1)} - \mathcal{O}^2_{\mathrm{ref}}Q^{2(N+1)} \tag{B.9}$$

$$\Leftrightarrow \Lambda = \frac{\mathcal{O}^2_{\mathrm{ref}}Q^{2(k+1)} - \mathcal{O}^2_{\mathrm{ref}}Q^{2(N+1)}}{1 - Q^2}. \tag{B.10}$$

Letting $N \to \infty$ yields (since $0 < Q < 1$)

$$\Lambda \underset{N\to\infty}{=} \mathcal{O}^2_{\mathrm{ref}} \frac{Q^{2(k+1)}}{1 - Q^2} \tag{B.11}$$

and

$$\bar{c}^2\Lambda \underset{N\to\infty}{=} \bar{c}^2\mathcal{O}^2_{\mathrm{ref}} \sum_{n=k+1}^{\infty} Q^{2n} = \bar{c}^2\mathcal{O}^2_{\mathrm{ref}} \frac{Q^{2(k+1)}}{1 - Q^2} \tag{B.12}$$

Hence $\delta\mathcal{O}^{(k)}_{\mathrm{theo}}$ is distributed according to

$$\mathrm{pr}(\delta\mathcal{O}^{(k)}_{\mathrm{theo}}|\bar{c}, Q) = \mathcal{N}(0, \sigma^2_{\mathrm{theo}}) \tag{B.13}$$

with

$$\sigma^2_{\mathrm{theo}} = \bar{c}^2(\mathcal{O}_{\mathrm{ref}})^2 \frac{Q^{2(k+1)}}{1 - Q^2}. \tag{B.14}$$

In the multivariate case it can be shown that [24]

$$\mathrm{pr}(\delta\vec{\mathcal{O}}^{(k)}_{\mathrm{theo}}|\bar{c}, Q) = \mathcal{N}(0, \Sigma_{\mathrm{theo}}) \tag{B.15}$$

where, under the assumption that the truncation errors are uncorrelated,

$$\Sigma_{\mathrm{theo}} = \sigma^2_{\mathrm{theo},i}\delta_{ii}. \tag{B.16}$$

Here the index i refers to the ith datum.

Bibliography

[1] H. A. Bethe. "Nuclear many-body problem". *Phys. Rev.* **103** (1956), pp. 1353–1390.

[2] H. Hergert. "A guided tour of *ab initio* nuclear many-body theory". *Front. in Phys.* **8** (2020), p. 379.

[3] H. Yukawa. "On the interaction of elementary particles I". *Proc. Phys. Math. Soc. Jap.* **17** (1935), pp. 48–57.

[4] R. Machleidt and D. R. Entem. "Chiral effective field theory and nuclear forces". *Phys. Rept.* **503** (2011), pp. 1–75.

[5] K. G. Wilson. "Confinement of quarks". *Phys. Rev. D* **10** (1974). Ed. by J. C. Taylor, pp. 2445–2459.

[6] Z. Davoudi et al. "Nuclear matrix elements from lattice QCD for electroweak and beyond-Standard-Model processes". *Phys. Rept.* **900** (2021), pp. 1–74.

[7] S. Weinberg. "Phenomenological Lagrangians". *Physica A* **96**.1-2 (1979). Ed. by S. Deser, pp. 327–340.

[8] E. Epelbaum, H.-W. Hammer, and U.-G. Meissner. "Modern theory of nuclear forces". *Rev. Mod. Phys.* **81** (2009), pp. 1773–1825.

[9] H. Hammer, S. König, and U. van Kolck. "Nuclear effective field theory: status and perspectives". *Rev. Mod. Phys.* **92**.2 (2020), p. 025004.

[10] S. Weinberg. "Nuclear forces from chiral Lagrangians". *Phys. Lett. B* **251** (1990), pp. 288–292.

[11] S. Weinberg. "Effective chiral Lagrangians for nucleon-pion interactions and nuclear forces". *Nucl. Phys. B* **363** (1991), pp. 3–18.

[12] S. Weinberg. "Three body interactions among nucleons and pions". *Phys. Lett. B* **295** (1992), pp. 114–121.

[13] C. Ordóñez and U. van Kolck. "Chiral Lagrangians and nuclear forces". *Phys. Lett. B* **291**.4 (1992), pp. 459–464.

[14] D. R. Entem, R. Machleidt, and Y. Nosyk. "High-quality two-nucleon potentials up to fifth order of the chiral expansion". *Phys. Rev. C* **96**.2 (2017), p. 024004.

[15] H. W. Grießhammer. "Assessing theory uncertainties in EFT power countings from residual cutoff dependence". *PoS* **CD15** (2016), p. 104.

[16] C. Yang et al. "Power counting in chiral effective field theory and nuclear binding". *Phys. Rev. C* **103**.5 (2021), p. 054304.

[17] D. R. Entem and R. Machleidt. "Accurate charge dependent nucleon nucleon potential at fourth order of chiral perturbation theory". *Phys. Rev. C* **68** (2003), p. 041001.

[18] B. D. Carlsson et al. "Uncertainty analysis and order-by-order optimization of chiral nuclear interactions". *Phys. Rev. X* **6**.1 (2016), p. 011019.

[19] P. Reinert, H. Krebs, and E. Epelbaum. "Semilocal momentum-space regularized chiral two-nucleon potentials up to fifth order". *Eur. Phys. J. A* **54**.5 (2018).

[20] M. R. Schindler and D. R. Phillips. "Bayesian methods for parameter estimation in effective field theories". *Annals Phys.* **324** (2009), pp. 682–708.

[21] R. J. Furnstahl et al. "Quantifying truncation errors in effective field theory". *Phys. Rev. C* **92**.2 (2015), p. 024005.

[22] S. Wesolowski et al. "Bayesian parameter estimation for effective field theories". *J. Phys. G* **43**.7 (2016), p. 074001.

[23] J. A. Melendez, S. Wesolowski, and R. J. Furnstahl. "Bayesian truncation errors in chiral effective field theory: nucleon-nucleon observables". *Phys. Rev. C* **96**.2 (2017), p. 024003.

[24] S. Wesolowski et al. "Exploring Bayesian parameter estimation for chiral effective field theory using nucleon–nucleon phase shifts". *J. Phys. G* **46**.4 (2019), p. 045102.

[25] J. A. Melendez et al. "Quantifying correlated truncation errors in effective field theory". *Phys. Rev. C* **100**.4 (2019), p. 044001.

[26] J. Brynjarsdóttir and A. O'Hagan. "Learning about physical parameters: the importance of model discrepancy". *Inverse Probl.* **30**.11 (2014), p. 114007.

[27] R. M. Neal. "MCMC using Hamiltonian dynamics". *Handbook of Markov chain Monte Carlo*. Ed. by S. Brooks et al. Boca Raton: CRC Press, 2011, pp. 113–162.

[28] S. Duane et al. "Hybrid Monte Carlo". *Phys. Lett. B* **195**.2 (1987), pp. 216–222.

[29] B. A. Lippmann and J. Schwinger. "Variational principles for scattering processes. I". *Phys. Rev.* **79** (1950), pp. 469–480.

[30] M. I. Haftel and F. Tabakin. "Nuclear saturation and the smoothness of nucleon-nucleon potentials". *Nucl. Phys. A* **158** (1970), pp. 1–42.

[31] R. Navarro Pérez, J. E. Amaro, and E. Ruiz Arriola. "Partial-wave analysis of nucleon-nucleon scattering below the pion-production threshold". *Phys. Rev. C* **88** (2 2013), p. 024002.

[32] R. N. Pérez, J. E. Amaro, and E. R. Arriola. "Coarse-grained potential analysis of neutron-proton and proton-proton scattering below the pion production threshold". *Phys. Rev. C* **88** (6 2013), p. 064002.

[33] P. W. Lisowski et al. "Search for resonance structure in the np total cross section below 800 MeV". *Phys. Rev. Lett.* **49** (4 1982), pp. 255–259.

[34] *INS Data Analysis Center: SAID*. http://gwdac.phys.gwu.edu. Accessed: 2021-03-29.

[35] W. Jiang et al. "Accurate bulk properties of nuclei from $A = 2$ to ∞ from potentials with Δ isobars". *Phys. Rev. C* **102**.5 (2020), p. 054301.

[36] A. Ekström et al. "Accurate nuclear radii and binding energies from a chiral interaction". *Phys. Rev. C* **91**.5 (2015), p. 051301.

[37] D. Stump et al. "Uncertainties of predictions from parton distribution functions. 1. The Lagrange multiplier method". *Phys. Rev. D* **65** (2001), p. 014012.

[38] E. T. Jaynes and G. L. Bretthorst. *Probability theory: the logic of science*. Cambridge University Press, 2003.

[39] J. Dobaczewski, W. Nazarewicz, and P. Reinhard. "Error estimates of theoretical models: a guide". *J. Phys. G* **41** (2014), p. 074001.

[40] R. B. Wiringa, V. G. J. Stoks, and R. Schiavilla. "An accurate nucleon-nucleon potential with charge independence breaking". *Phys. Rev. C* **51** (1995), pp. 38–51.

[41] R. Machleidt. "The high precision, charge dependent Bonn nucleon-nucleon potential (CD-Bonn)". *Phys. Rev. C* **63** (2001), p. 024001.

[42] P. Reinert, H. Krebs, and E. Epelbaum. "Semilocal momentum-space regularized chiral two-nucleon potentials up to fifth order". *Eur. Phys. J. A* **54**.5 (2018), p. 86.

[43] U. van Kolck. "Few nucleon forces from chiral Lagrangians". *Phys. Rev. C* **49** (1994), pp. 2932–2941.

[44] M. Hoferichter et al. "Roy–Steiner-equation analysis of pion–nucleon scattering". *Phys. Rept.* **625** (2016), pp. 1–88.

[45] D. Siemens et al. "Reconciling threshold and subthreshold expansions for pion–nucleon scattering". *Phys. Lett. B* **770** (2017), pp. 27–34.

[46] D. E. Gonzalez Trotter et al. "Neutron-deuteron breakup experiment at E(n) = 13 MeV: Determination of the $^1S_0$ neutron-neutron scattering length a_{nn}". *Phys. Rev. C* **73** (2006), p. 034001.

[47] Q. Chen et al. "Measurement of the neutron-neutron scattering length using the pi-d capture reaction". *Phys. Rev. C* **77** (2008), p. 054002.

[48] G. A. Miller, B. M. K. Nefkens, and I. Slaus. "Charge symmetry, quarks and mesons". *Phys. Rept.* **194** (1990), pp. 1–116.

[49] I. Charpentier and J. Utke. "Fast higher-order derivative tensors with Rapsodia". *Optim. Method. Softw.* **24**.1 (2009), pp. 1–14.

[50] B. Acharya et al. "Uncertainty quantification for proton–proton fusion in chiral effective field theory". *Phys. Lett. B* **760** (2016), pp. 584–589.

[51] A. Gelman et al. *Bayesian data analysis*. Texts in statistical science series. Chapman & Hall/CRC, 2014.

[52] N. Metropolis et al. "Equation of state calculations by fast computing machines". *J. Chem. Phys.* **21**.6 (1953), pp. 1087–1092.

[53] W. K. Hastings. "Monte Carlo sampling methods using Markov chains and their applications". *Biometrika* **57** (1970), pp. 97–109.

[54] G. O. Roberts, A. Gelman, and W. R. Gilks. "Weak convergence and optimal scaling of random walk Metropolis algorithms". *Ann. Appl. Probab.* **7**.1 (1997), pp. 110–120.

[55] J. Goodman and J. Weare. "Ensemble samplers with affine invariance". *Comm. App. Math. Com. Sc.* **5**.1 (2010), pp. 65–80.

[56] D. Foreman-Mackey et al. "emcee: the MCMC hammer". *PASP* **125** (2013), pp. 306–312.

[57] A. Sokal. “Monte Carlo methods in statistical mechanics: foundations and new algorithms”. *Functional Integration: Basics and Applications*. Ed. by C. DeWitt-Morette, P. Cartier, and A. Folacci. Boston, MA: Springer US, 1997, pp. 131–192.

[58] J. M. Hammersley and K. W. Morton. “A new Monte Carlo technique: antithetic variates”. *Math. Proc. Cambridge* **52**.3 (1956), pp. 449–475.

[59] R. H. Swendsen and J.-S. Wang. “Replica Monte Carlo simulation of spin-glasses”. *Phys. Rev. Lett.* **57** (21 1986), pp. 2607–2609.

[60] C. J. Geyer. “Markov chain Monte Carlo maximum likelihood”. *Comp. Sci. Stat.: Proceedings of the 23rd Symposium Interface, 1991*. 1991.

[61] J. Skilling. “Nested sampling”. *AIP Conf. Proc.* **735**.1 (2004), pp. 395–405.

[62] D. S. Sivia and J. Skilling. *Data analysis: a Bayesian tutorial*. Oxford science publications. Oxford University Press, 2006.

[63] A. Griewank. “A mathematical view of automatic differentiation”. *Acta Numer.* **12** (2003), pp. 321–398.

[64] G. Van Rossum and F. L. Drake. *Python 3 reference manual*. Scotts Valley, CA: CreateSpace, 2009.

[65] C. R. Harris et al. “Array programming with NumPy”. *Nature* **585** (2020), pp. 357–362.

[66] Stan Development Team. *Stan modeling language users guide and reference manual, version 2.26*. 2021. URL: `https://mc-stan.org`.

[67] M. Betancourt. *A conceptual introduction to Hamiltonian Monte Carlo*. 2018. arXiv: `1701.02434 [stat.ME]`.

[68] L. Verlet. “Computer “experiments” on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules”. *Phys. Rev.* **159** (1 1967), pp. 98–103.

[69] M. D. Homan and A. Gelman. “The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo”. *J. Mach. Learn. Res.* **15**.1 (2014), pp. 1593–1623.

[70] A. Ekström et al. “Optimized chiral nucleon-nucleon interaction at next-to-next-to-leading order”. *Phys. Rev. Lett.* **110**.19 (2013), p. 192502.

[71] I. Vernon, M. Goldstein, and R. G. Bower. “Galaxy formation: a Bayesian uncertainty analysis”. *Bayesian Anal.* **5**.4 (2010), pp. 619–669.

[72] V. Roy. “Convergence diagnostics for Markov chain Monte Carlo”. *Annu. Rev. Stat. Appl.* **7**.1 (2020), pp. 387–412.

[73] A. Gelman and D. B. Rubin. “Inference from iterative simulation using multiple sequences”. *Stat. Sci.* **7**.4 (1992), pp. 457–472.

[74] S. Brooks and A. Gelman. “General methods for monitoring convergence of iterative simulations”. *J. Comput. Graphi. Stat.* **7** (1998), pp. 434–455.

[75] D. Foreman-Mackey. *Autocorrelation analysis & convergence — emcee 3.0.2 documentation*. 2021. URL: `https://emcee.readthedocs.io/en/stable/tutorials/autocorr/`.

[76] G. E. P. Box. “Science and statistics”. *J. Am. Stat. Assoc.* **71**.356 (1976), pp. 791–799.

[77] K. P. Murphy. *Machine learning: a probabilistic perspective*. Adaptive computation and machine learning series. MIT Press, 2012.

[78] A. E. Gelfand. “Model determination using sampling-based methods”. *Markov chain Monte Carlo in practice*. Ed. by W. R. Gilks, S. Richardson, and D. J. Spiegelhalter. London, UK: Chapman & Hall, 1996. Chap. 9, pp. 145–161.

[79] S. Brooks et al. *Handbook of Markov chain Monte Carlo*. Chapman and Hall/CRC Handbooks of Modern Statistical Methods Ser. CRC Press LLC, 2011.

[80] D. Frame et al. “Eigenvector continuation with subspace learning”. *Phys. Rev. Lett.* **121**.3 (2018), p. 032501.

[81] S. B. S. Miller, A. Ekström, and C. Forssén. “Accelerating nucleon-nucleon scattering calculations” (2021). arXiv: `2106.00454 [nucl-th]`.