

## A Bayesian approach to chiral extrapolations

---

**Matthias R. Schindler\***

*Department of Physics and Astronomy Ohio University*  
*Athens, OH 45701*  
USA  
E-mail: [schindle@ohio.edu](mailto:schindle@ohio.edu)

**Daniel R. Phillips**

*Department of Physics and Astronomy Ohio University*  
*Athens, OH 45701*  
USA  
E-mail: [phillips@phy.ohiou.edu](mailto:phillips@phy.ohiou.edu)

The determination of low-energy constants from data is an important component of most effective field theory programs, including that of chiral perturbation theory. We propose a novel method based on Bayesian probability theory which allows us to address several shortcomings of the standard approach to parameter extraction. Using a toy-model we argue that the Bayesian approach is ideally suited for the application in effective field theories. We also discuss the application to lattice QCD data.

*6th International Workshop on Chiral Dynamics*  
*July 6-10 2009*  
*Bern, Switzerland*

---

\*Speaker.

## 1. Introduction

An effective field theory (EFT) is a low-energy approximation to an underlying theory. It allows for a model-independent description of phenomena at an energy scale  $m$  that is much lower than an underlying scale  $\Lambda$ . The Lagrangian of the EFT is constructed by including all terms that are consistent with the symmetries of the underlying theory. Each of the terms in the Lagrangian is accompanied by a so-called low-energy constant (LEC) that incorporates the effects of high-energy degrees of freedom on the low-energy dynamics. The EFT leads to a perturbative expansion for observables at the low-energy scale if the LECs are of order  $\mathcal{O}(1)$  in units of the high-energy scale, i.e. if they are “natural” with respect to  $\Lambda$ . In principle, these LECs can be determined from the underlying theory. In practice, however, there are only a few cases in which the LECs can be rigorously derived from the underlying theory, and in all other instances the only model-independent way to determine the LECs is by comparison with experimental data.

The standard approach to the extraction of LECs from data is to calculate an observable at some given order and then perform a fit of this EFT expression using methods like least squares or maximum likelihood. There are several issues with this approach that we are going to address:

1. Which order in the EFT expansion should be used to perform the fit?
2. How can the naturalness requirement on the LECs be incorporated?
3. What is the appropriate energy regime to perform the fit? In most cases more data is available for higher energies, but the reliability of the EFT calculation decreases as the energy is increased.

With data sets that include a large number of very precise measurements, these issues are not of any significance. If, however, only limited and imprecise data is available, these issues manifest themselves as sensitivity of the extracted LECs on the way the fit is performed.

In order to avoid the above-mentioned issues we have developed an approach that is based on Bayesian probability theory [1]. We argue that Bayesian methods (for an introduction see e.g. [2]) are ideally suited for the extraction of LECs. In the Bayesian approach prior knowledge on the parameters can be easily included in the process of estimating these parameters. When combined with the concept of marginalization, applied to the order of the fit function, the derived method resolves the first two issues in the above list. We also show that this method is not sensitive to higher-energy data within certain bounds.

## 2. Bayesian probability theory

Consider a general EFT for which the LECs are denoted by  $\mathbf{a} = \{a_i | i = 1, \dots, M\}$ . In the following we will restrict the discussion to extracting a subset  $\mathbf{a}_{res}$  of these unknown parameters from some given data  $D = \{(d_k, \sigma_k) | k = 1, \dots, N\}$ , where  $d_k$  is an individual measurement at  $x_k$  with associated uncertainty  $\sigma_k$ . We are therefore interested in the probability density

$$\text{pr}(\mathbf{a}_{res} | D), \tag{2.1}$$

where  $\text{pr}(A|B)$  denotes the conditional probability density of  $A$  given  $B$ . Bayes' theorem relates this probability density to the more familiar likelihood  $\text{pr}(D|\mathbf{a}_{res})$ ,

$$\text{pr}(\mathbf{a}_{res}|D) = \frac{\text{pr}(D|\mathbf{a}_{res})\text{pr}(\mathbf{a}_{res})}{\text{pr}(D)}. \quad (2.2)$$

Here,  $\text{pr}(\mathbf{a}_{res})$  is the so-called prior which incorporates any information available on the parameters prior to analysis of the data. The denominator can be obtained from the requirement that  $\text{pr}(\mathbf{a}_{res}|D)$  be normalized. The prior information we wish to include is the assumption of naturalness of the parameters. However, the notion of “naturalness” is not strictly defined. Here we employ the principle of maximum entropy to motivate a prior of the form [1]:

$$\text{pr}(\mathbf{a}|M, R) = \left( \frac{1}{\sqrt{2\pi}R} \right)^{M+1} \exp\left(-\frac{\mathbf{a}^2}{2R^2}\right). \quad (2.3)$$

Note that we have introduced several additional parameters:  $\mathbf{a} = (\mathbf{a}_{res}, \mathbf{a}_{marg})$  denotes the complete set of LECs at a given order, including the higher-order LECs  $\mathbf{a}_{marg}$  that we do *not* wish to extract,  $M$  is related to the order of the EFT calculation,<sup>1</sup> and  $R$  is a parameter that encodes the definition of naturalness as chosen here. Thus, while we have succeeded in defining the prior, this has come at the price of the introduction of these additional parameters. Since we are not interested in the exact values of these parameters and, in fact, one of our aims was to avoid having to fix the value of  $M$ , we apply marginalization to eliminate these “nuisance” parameters. The general marginalization description is given by

$$\text{pr}(A|C) = \int dB \text{pr}(A, B|C), \quad (2.4)$$

that means unwanted parameters are integrated out. We apply marginalization to the higher-order LECs  $\mathbf{a}_{marg}$ , the order of the EFT calculation and the “naturalness parameter”  $R$ . This last marginalization thus takes into account the uncertainty in the definition of naturalness. The final probability density is given by (for a derivation see Ref. [1])

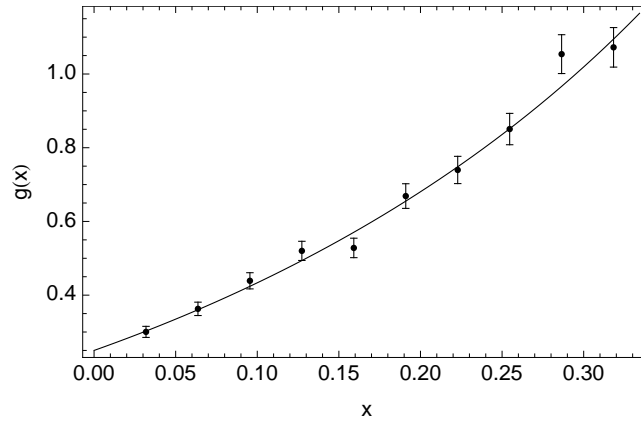
$$\text{pr}(\mathbf{a}_{res}|D) = \sum_M \int dR \int d\mathbf{a}_{marg} \frac{\text{pr}(D|\mathbf{a}, M)\text{pr}(\mathbf{a}|M, R)\text{pr}(M)\text{pr}(R)}{\text{pr}(D)}. \quad (2.5)$$

Since Bayes' theorem was employed several times in the derivation of Eq. (2.5) we are forced to introduce priors for  $M$  and  $R$ . We do not assume any particular knowledge of these parameters. However, since  $M$  is a “location parameter” and  $R$  is a “scale parameter” we use different priors. The prior for  $M$  is a constant, while  $\text{pr}(R) = \frac{1}{R}$  (see Ref. [1] for more details). The parameters and the associated uncertainties are determined from the first and second moments of the pdf,

$$\langle a_i \rangle = \int d\mathbf{a}_{res} a_i \text{pr}(\mathbf{a}_{res}|D), \quad (2.6)$$

$$\sigma_{a_i}^2 = \langle a_i^2 \rangle - \langle a_i \rangle^2. \quad (2.7)$$

<sup>1</sup>In general,  $M$ , the number of LECs, is not identical to the order of the calculation.



**Figure 1:** Generated artificial data. The solid line is the function  $g(x)$ .

### 3. Application to a toy problem

In order to demonstrate the advantages of our proposed method we consider an application to a toy problem. We generate pseudo-data using the function

$$g(x) = \left( \frac{1}{2} + \tan\left(\frac{\pi}{2}x\right) \right)^2 \quad (3.1)$$

for  $x \geq 0$ . Our aim is to extract the first two coefficients  $a_0, a_1$  of a polynomial

$$f(x) = \sum_{j=0}^P a_j x^j \quad (3.2)$$

from the pseudo-data, where  $P$  denotes the order of the polynomial. The function  $g(x)$  might not have any direct physical application, but it exhibits a number of features that are common in EFT applications.  $g(x)$  is non-analytic in  $x \in \mathbb{R}$ , but for  $x < 1$  it can be approximated to arbitrary precision by a power series. The first few terms in this power series are given by

$$g(x) \approx 0.25 + 1.57x + 2.47x^2 + 1.29x^3 + 4.06x^4 + \dots \quad (3.3)$$

The coefficients of at least the first ten terms are “natural”, however, their magnitude is not decreasing for increasing order.

The pseudo-data we wish to analyze, covering the range  $0 < x \leq 1/\pi$  are shown in Fig. 1. The prior information available is that the data are normally distributed (which reduces the problem to a minimum  $\chi^2$  one in the standard approach) and that the coefficients of the polynomial are  $\mathcal{O}(1)$ . The results of a standard least-squares fit at various orders of the polynomial, which does not take into account the information on the naturalness of the parameters, are shown in Tab. 1. While the quadratic fit reproduces the underlying values of  $a_0$  and  $a_1$  reasonably well and with a relatively low  $\chi^2$ , without knowledge of the underlying values it might be difficult to decide why the quadratic fit should be preferred. One should also note the lack of convergence, especially for  $a_1$ , as one goes to higher orders and the fast growth of the uncertainties. An experienced practitioner might be able

$P$	$\chi^2/d.o.f.$	$a_0$	$a_1$
1	2.23	$0.203 \pm 0.014$	$2.51 \pm 0.10$
2	1.06	$0.260 \pm 0.022$	$1.31 \pm 0.39$
3	1.13	$0.235 \pm 0.038$	$2.14 \pm 1.08$
4	1.13	$0.177 \pm 0.067$	$4.76 \pm 2.70$
5	0.99	$0.327 \pm 0.133$	$-3.56 \pm 6.94$
6	1.32	$0.314 \pm 0.297$	$-2.73 \pm 18.5$
7	1.47	$1.05 \pm 0.792$	$-56.3 \pm 56.5$

**Table 1:** Fit result for standard  $\chi^2$  approach.

to discern which of the various fits to trust most, however our aim is to eliminate the need for this post-analysis judgement.

We now apply our method based on the use of Bayes' theorem and marginalization to the data. The naturalness of the parameters is included in the analysis with the use of the prior of Eq. (2.3). We marginalize over the polynomial order from  $P = 2$  to  $P = 8$ . For the “naturalness parameter”  $R$  we choose  $R = 0.1 - 10$ . We find

$$a_0 = 0.246 \pm 0.021, \quad (3.4)$$

$$a_1 = 1.63 \pm 0.37, \quad (3.5)$$

in good agreement with the underlying values. We have avoided the need to choose a specific order for the fit; instead the uncertainties in the results for  $a_0$  and  $a_1$  include contributions from the marginalization over  $P$ . And while the results are influenced by our inclusion of the “naturalness prior”, the lack of exact knowledge of  $R$  again contributes to the final uncertainties of the parameters via marginalization. We therefore believe that our method not only leads to improved extraction of the parameters of interest, but also includes some of the uncertainties related to such an extraction in a more systematic way than the standard approach.

We have performed an analogous analysis with a different data set that contains the same number of data points, but for which  $0 < x \leq 2/\pi$ . The result for the standard  $\chi^2$  fit are shown in Tab. 2. With more data points closer to the radius of convergence the problems of the standard approach are exacerbated. While the fourth-order fit gives results not too far from the underlying values, without knowledge of these “true” values it is not clear which result to trust. In our Bayesian approach, again choosing  $P = 2 - 8$  and  $R = 0.1 - 10$ , we find

$$a_0 = 0.241 \pm 0.048, \quad (3.6)$$

$$a_1 = 2.23 \pm 0.74. \quad (3.7)$$

These values are again in agreement with the underlying values, and reproduce them much better than the standard  $\chi^2$  results. We consider it a strength of our method that the results are not as sensitive to high- $x$  data, allowing for the use of larger data sets.

$M$	$\chi^2/d.o.f.$	$a_0$	$a_1$
2	5.35	$0.392 \pm 0.033$	$-0.387 \pm 0.351$
3	1.47	$0.141 \pm 0.058$	$4.32 \pm 0.946$
4	1.48	$0.246 \pm 0.106$	$1.79 \pm 2.35$
5	1.46	$0.00697 \pm 0.217$	$8.67 \pm 5.94$
6	0.46	$0.995 \pm 0.516$	$-24.0 \pm 16.6$
7	0.50	$0.180 \pm 1.41$	$5.98 \pm 51.0$

**Table 2:** Fit results for standard  $\chi^2$  approach with  $x_{max} = 2/\pi$ .

#### 4. Application to lattice data

One possible application of the outlined method is the extraction of LECs in chiral perturbation theory (ChPT) from lattice data. In particular, we have studied the determination of the chiral limit value of the nucleon mass and the nucleon sigma term. There are several additional issues that need to be addressed. In these exploratory studies we again used pseudo-data generated at a set of pion mass values from the ChPT form of the nucleon mass. Our results suggest that for the naturalness prior of Eq. (2.3) larger values of  $R$  are suppressed, and the main contribution to the integral over  $R$  comes from the region  $R \sim 1 - 2$ . It should be noted that the numerical values of the dimensionless low-energy coefficients to which the naturalness assumption applies depend on the value of the underlying scale  $\Lambda$ . This manifests itself in a certain sensitivity of the extracted LECs on the choice of  $\Lambda$ . In addition we also want to make use of detailed information on some of the parameters that appear in the ChPT expression of the nucleon mass, such as the pion decay constant and the axial coupling of the nucleon. This information allows for use of more sophisticated priors. We are continuing our investigation of these issues [3].

#### 5. Conclusions

Extraction of the values of parameters relevant to low-energy dynamics from pertinent data is an important part of effective-field-theory calculations. We have presented a novel approach to this problem that is based on Bayesian probability theory. In this approach, prior information regarding the parameters of interest can be taken into account during the data analysis. This also allows for a more systematic inclusion of uncertainties related to truncations in the EFT. Application to a toy problem shows that our method results in an improved extraction of the low-energy constants of interest. We are continuing to study the application of these ideas to lattice QCD data and chiral perturbation theory.

#### Acknowledgments

We thank the organizers for a very interesting workshop and gratefully acknowledge discussions with J. A. McGovern. This work was supported by the US DOE under grant no. DE-FG02-93ER40756.

## References

- [1] M. R. Schindler and D. R. Phillips, *Annals Phys.* **324** (2009) 682 [Erratum-ibid. **324** (2009) 2051] [arXiv:0808.3643 [hep-ph]].
- [2] D. S. Sivia, *Data analysis: a Bayesian tutorial*, Oxford University Press, 1996.
- [3] J. A. McGovern, D. R. Phillips, M. R. Schindler, in preparation.