Model selection for pion photoproduction

J. Landay, M. Döring, C. Fernández-Ramírez, B. Hu, and R. Molina

Phys. Rev. C **95**, 015203 — Published 12 January 2017

DOI: 10.1103/PhysRevC.95.015203

Model Selection for Pion Photoproduction

J. Landay,[1,*] M. Döring,[2,3,†] C. Fernández-Ramírez,[4,‡] B. Hu,[1] and R. Molina[1]

[1] *The George Washington University, Washington, DC 20052, USA*
[2] *Institute for Nuclear Studies (INS); Astronomy, Physics and Statistics Institute of Sciences (APSIS), The George Washington University, Washington, DC 20052, USA*
[3] *Theory Center, Thomas Jefferson National Accelerator Facility, Newport News, VA 23606, USA*
[4] *Instituto de Ciencias Nucleares, Universidad Nacional Autónoma de México, Ciudad de México 04510, Mexico*

Partial-wave analysis of meson and photon-induced reactions is needed to enable the comparison of many theoretical approaches to data. In both energy-dependent and independent parametrizations of partial waves, the selection of the model amplitude is crucial. Principles of the S-matrix are implemented to different degree in different approaches; but a many times overlooked aspect concerns the selection of undetermined coefficients and functional forms for fitting, leading to a minimal yet sufficient parametrization. We present an analysis of low-energy neutral pion photoproduction using the Least Absolute Shrinkage and Selection Operator (LASSO) in combination with criteria from information theory and K-fold cross validation. These methods are not yet widely known in the analysis of excited hadrons but will become relevant in the era of precision spectroscopy. The principle is first illustrated with synthetic data; then, its feasibility for real data is demonstrated by analyzing the latest available measurements of differential cross sections ($d\sigma/d\Omega$), photon-beam asymmetries (Σ), and target asymmetry differential cross sections ($d\sigma_T/d \equiv T\, d\sigma/d\Omega$) in the low-energy regime.

PACS numbers: 11.80.Et, 02.70.Rr, 25.20.-x, 11.30.Rd,

I. INTRODUCTION

The understanding of the strong interaction in the hadronic energy regime is an important unresolved issue that has regained a lot of attention in the last years due to the advances in detection techniques, accelerator technologies, first principle Quantum Chromodynamics (QCD) analyses, and S-matrix theory amplitude analysis techniques [1–6]. These developments have lead to a broad effort to build and perform experiments that are or will be collecting an unprecedented amount of data on hadron reactions, e.g., BELLEII [7, 8], BESIII [9], CLAS12 [10], CMS [11], COMPASS [12], ELSA [13, 14], GlueX [15], J-PARC [16], KLOE2 [17], LHCb [18], MAMI [19] and PANDA [20].

Partial-wave analysis of hadronic reactions is a prerequisite for many theoretical approaches to access information from the experimental data, especially if the comparison to hadron resonances is the goal. Narrowing the focus to photoproduction reactions, their decomposition into partial waves (multipoles) is usually performed through an energy-dependent (ED) parametrization of the amplitude. As long as data are not abundant and precise enough, it is not yet possible to perform a (truncated partial-wave) complete experiment [21, 22] in the resonance region. A parametrization in energy is needed for the determination of resonances, or as a stabilizing starting point for single-energy (SE) solutions in which energy-binned data are fitted independently. For both ED and SE analyses, the selection of fit parameters is a fundamental problem that we address in this study for the case of neutral pion photoproduction in the low energy region. We use this well-studied reaction as a benchmark for various techniques and as a template for future works because of the well-established theoretical framework and the availability of high-quality data which allow an (almost) model-independent SE extraction [23, 24].

In the analysis of photoproduction experiments, the parametrization of the amplitude is chosen according to the considered energy range. For low-energy neutral pion photoproduction, Heavy Baryon Chiral Perturbation Theory (HBChPT) [25–33] and Relativistic Baryon Chiral Perturbation Theory (RBChPT) [34] provide effective parametrizations. Both deliver equally good descriptions of the latest experimental data up to $E_\gamma \simeq 170$ MeV in the laboratory frame [23, 35]. Polynomial parametrizations which incorporate unitarity in the S wave have also proved to be an excellent description of the data up to $E_\gamma \simeq 185$ MeV where the $\Delta(1232)$ contribution begins to be relevant [23, 36, 37]. ChPT calculations including isospin breaking have been performed in Refs. [38–40] and the inclusion of the $\Delta(1232)$ resonance as an explicit degree of freedom in RBChPT allows to extend the agreement between theory and data up to $E_\gamma \simeq 200$ MeV [41, 42]. The RBChPT calculation in [34] has been extended to include also electroproduction of charged pions [43].

In an explicit parametrization of partial waves one cannot incorporate an infinite number (the series has to be truncated) and we need to determine how many terms one needs to incorporate in order to provide an accurate,

* jlanday@gwmail.gwu.edu
† doring@gwu.edu
‡ cesar.fernandez@nucleares.unam.mx

yet minimally parameterized, description of the physics involved. In Refs. [36, 44] it was determined that in the low-energy neutral pion photoproduction region we need to incorporate up to D waves and that higher partial waves can be safely dismissed. Although D waves are not necessary to describe the experimental data, not including them prevents the accurate extraction of the S wave, which is small since it vanishes in the chiral limit and provides insight into chiral symmetry breaking [45]. The structure of the observables in terms of the multipoles can be found in [36, 46, 47].

Beyond energies in which a systematic treatment is possible, effective field theory approaches [48–56], K-matrix parametrizations, or related approaches are used in [57–59]. At high energies, Regge parametrizations are very effective [60, 61]. Formulations to provide amplitudes that cover the entire energy region from threshold to the highest energies are under development [62, 63]. Yet, sometimes partial waves are parameterized purely phenomenologically in terms of functions that are in agreement with basic S-matrix principles such as coupled-channel two-body unitarity, the correct threshold behavior or Fermi-Watson's theorem [64, 65], but that are otherwise left free to ensure a high degree of model independence as in the SAID approach [57].

In general, new high-precision polarization data indeed lead to more consistent multipole solutions among different analysis groups although discrepancies remain [66]. A step towards the goal of matching solutions has been done recently. Providing the necessary information for other groups to carry out correlated χ^2 fits of πN partial waves, the statistical influence of elastic pion-nucleon scattering has been quantified [67]. This gives a more statistical foundation for multi-reaction analyses by many groups in the search for new excited baryons. Complementary information from hadron beams could also lead to more consistent solutions among the different partial-wave analysis approaches [68]. Considerations of which observables and which precision are necessary to discriminate models are also a necessary step forward to find definite answers in baryon spectroscopy [69].

The selection of fit parameters in most of these approaches is important. If the amplitude is under-parameterized, the quality of the data description is not satisfactory and the quality of the extracted amplitudes is difficult to assess. Over-parametrization can result in limited predictability of the amplitude outside the fitted data range and inflated uncertainties. Furthermore, problems in the data themselves (incompatibility of data, systematics, or even statistics) may be interpreted as significant physics in over-parameterized fits.

Most notably, in many approaches resonances are introduced in the parametrization as explicit terms, that will unavoidably improve the fit quality at the cost of potentially false positive resonance signals [70]. To control this problem, groups use mass scan techniques in which, ideally, a minimum of the χ^2 as a function of the resonance mass appears in more than one analyzed channel [71]. In the SAID approach, resonances appear dynamically generated, meaning that poles in the amplitude can appear without manual intervention, if required by data [57]. In Refs. [72, 73], the most probable resonance content (Bayesian evidence) is determined considering kaon photoproduction. Bayesian priors have also been used to restrict the low-energy constants in effective field theories to natural values and estimate the truncation errors [74, 75].

If a flexible background with resonance terms on top of it is provided, the task consists in minimizing the number of resonances and only accept them as physically significant if the background cannot provide a satisfactory description. The Least Absolute Shrinkage and Selection Operator (LASSO) [76–78] provides a tool to scan a plethora of different models, in particular multiple combinations of different resonances. Manually, such a scan would be impossible due to the large number of combinations, but the LASSO provides an automatized, "blindfolded" technique [79].

Having the above-mentioned extensions for future work in mind, we concentrate in this study on the question of how to select the simplest amplitude for a real-life example of photoproduction reactions. We choose low-energy neutral-pion photoproduction, $\gamma p \to \pi^0 p$, for which data of unprecedented precision exist from the same experimental setup at MAMI [23, 35], thus minimizing the potential impact from conflicting data or systematic uncertainties. The differential cross section $d\sigma/d\Omega$, photon beam asymmetry Σ, and target polarization differential cross section $d\sigma_T/d\Omega \equiv T\, d\sigma/d\Omega$ are analyzed. The entire presentation is kept as pedagogical as possible, even quoting textbook formulae for easier reference. The most relevant references on statistical analysis are [77, 78, 80].

II. FORMALISM

A. Parametrization

An energy-dependent parametrization is formulated for both real and imaginary parts of the three P waves, as well as for the real parts of E_{0+} and the four D-wave multipoles,

$$\mathrm{Re,Im}\,\mathcal{M}_{L\pm} = \frac{q_{\pi^0}^{\ell}}{m_{\pi^+}^{\ell+1}} \sum_{i=0}^{i_{\max}} \frac{a_i}{10^{-i}} \left(\frac{\omega_{\pi_0} - m_{\pi_0}}{m_{\pi^+}}\right)^i \quad (1)$$

where q_{π^0} is the center-of-mass momentum of the neutral pion, $\omega_{\pi_0}^2 = m_{\pi^0}^2 + q_{\pi^0}^2$, and a_i are the fit parameters. The quantity $\mathcal{M}$ stands for the electric ($E_{L\pm}$) and magnetic ($M_{L\pm}$) multipoles, or alternatively, the partial waves P_1,

P_2 and P_3 for the P waves, related to each other by

$$\begin{aligned} E_{1+} =& \frac{1}{6}(P_1 + P_2) \\ M_{1+} =& \frac{1}{6}(P_1 - P_2 + 2\,P_3) \\ M_{1-} =& \frac{1}{3}(P_3 + P_2 - P_1)\ . \end{aligned} \tag{2}$$

The parametrization in energy consists of a factor providing the correct threshold behavior and a Taylor expansion in momentum-squared around the neutral pion threshold. This expansion is theoretically justified by ChPT calculations [25–32]. In principle, it is preferable to use a set of polynomials in energy that is orthogonal in the fitted energy window, which reduces the correlations among parameters. However, in this particular example we could not observe any improvement when doing so.

Furthermore, in Eq. (1) for the real parts of the multipoles, one has $\ell = L$, while for the imaginary parts of the P-wave multipoles, $\ell = 3L + 1{=}4$ as can be obtained from Watson's theorem. We do not provide any imaginary parts for the D-waves because, on one hand, they are extremely small and, on the other end, restricting them to be real fixes the overall-phase ambiguity. Fit parameters are generically called a_i throughout, and we omit the indices specifying to which partial wave (real or imaginary part) they belong. The same applies to the cut-off $i_{\max}$. For the real parts of the P-waves and D-waves, $i_{\max} = 4$ while for the imaginary parts of the P-wave, $i_{\max} = 0$. Other factors in Eq. (1) serve to make all parameters dimensionless and of natural size. For the S-wave multipole E_{0+}, a real-valued term of the form of Eq. (1) with $i_{\max} = 4$ is supplemented by a term of the form

$$\Delta E_{0+} = i\frac{q_{\pi^+}}{m_{\pi^+}^2}\sum_{i=0}^{i=2}\frac{a_i}{10^{-i}}\left(\frac{q_{\pi^+}}{m_{\pi^+}}\right)^{2i} \tag{3}$$

to take into account the $\pi^+ n$ threshold cusp. The term provides an imaginary part above the $\pi^+ n$ threshold and contributes to the real part of E_{0+} below it. In total, there are $i_{\max} = 46$ free fit parameters. The number of available parameters is simply chosen such that every multipole can be grossly over-fitted.

B. Criteria from information theory

With this parametrization at hand, we turn to the LASSO method to select the *simplest* model [76–78]. The penalized χ^2_T is defined as follows:

$$\chi^2_T(\lambda) = \chi^2(\lambda) + P(\lambda), \tag{4}$$

with

$$P(\lambda) = \lambda^4 \sum_{i=1}^{i_{\max}} |a_i|\ . \tag{5}$$

In practice, one scans an entire range of λ, continuously minimizing χ^2_T. From there, we turn to various criterions from information theory in order to determine the optimal λ. Note that the power of four in Eq. (5) is simply chosen to provide a more convenient graphical representation of these criteria in the following plots. The three criteria that we use to find an optimal λ are the Akaike Information Criterion (AIC) [82, 83], a finite sample size corrected version of the AIC (AICc) [84], and the Bayesian Information Criterion (BIC) [85]. The three are defined as

$$\begin{aligned} \text{AIC} =& 2k - 2\log(L) = 2k + \chi^2\ , \\ \text{AICc} =& \text{AIC} + \frac{2k(k+1)}{n-k-1}\ , \\ \text{BIC} =& k\log(n) - 2\log(L) = k\log(n) + \chi^2\ , \end{aligned} \tag{6}$$

where k is the number of parameters which changes dynamically as a function of λ, n is the number of data points, and L is the likelihood. We define the number of degrees of freedom (d.o.f.) for each fit as d.o.f. $= n - k$. For all three of the criteria, the optimal value of λ is given by the respective minimum. This is because the AIC and the BIC take on small values for models with low test error. Assuming a Gaussian model, the BIC is proportional to the AIC, however the BIC tends to penalize models with more parameters due to the factor $\log(n)$ which allows for a more distinct minimum to be seen and a better indication of which model to use. For a further comparison between the AIC, AICc and the BIC, see Refs. [77, 78].

In Fig. 1 we show the different criteria as a function of the penalty λ in a simple simulation of fitting 22 synthetic data generated from a low-order polynomial with a model that includes that low-order polynomial as solution and also allows for over-fitting. One recognizes here a clear difference between the AIC and the AICc. The correction for the finite number of data points indeed leads to a larger AICc at small λ, providing a better identification of the minimum than the AIC. In the case of pion photoproduction considered later, the number of data is so large, though, that the difference between AIC and AICc becomes irrelevant.

C. Cross validation

Another method to find a model with minimal parametrization yet with accurate description of data is cross validation [76–78]. In short, data are randomly divided into a training set and a validation set. For a given λ, the penalized χ^2_T of the training set is minimized and the χ^2_V of the validation set is determined from that fit (without refitting). While χ^2_T is clearly a monotonously increasing function of λ, the situation is different for χ^2_V. For very large λ, χ^2_V is large as well, because the data are under-fitted. However, as λ decreases a point is reached below which χ^2_T is over-fitted, i.e., non-physical structures such as statistical fluctuations in the training set

FIG. 1. The AIC, AICc, and BIC criteria for a simple χ^2 fit to 22 data (see text). All three methods exhibit the minimum at the same value of λ indicating the optimal, i.e., simplest model. The correction term for a finite number of data in the AICc indeed produces a more pronounced minimum than for the AIC. The discontinuities occur every time a fit parameter is set to zero by LASSO.

are described. These fluctuations are different in the validation set, such that the validation χ^2_V becomes *larger* again as λ decreases further. The minimum in χ^2_V is then regarded as the *sweet spot* for λ, i.e., the point where the fit optimally describes the data without describing fluctuations.

The method has been recently applied in [86] for the determination of parton distribution functions. Another example is given in [87]. There, the task consisted in effectively smoothing a function that is subject to oscillations from unphysical finite-volume effects. In that example, the unphysical effect was not given by fluctuations which demonstrates that LASSO in combination with cross validation is a method with broader range of applicability than needed here.

The minimum of χ^2_V itself carries uncertainties. In practical terms it is numerically demanding to carry out the above-mentioned separation of training and validation sets for all possible combinations of data. An approximate method is given by K-fold cross validation in which the (uncorrelated) data are randomly divided in a few (here: 5) partitions. In five different cross validation runs, four of the five partitions serve as training set while the fifth serves as validation set. The different outcomes are used to estimate the uncertainty of χ^2_V at each λ. One can then further constrain the search for the simplest model by selecting the $\lambda_{\rm opt} > \lambda_{\rm min}$ that is compatible within errors with the minimum of χ^2_V at $\lambda = \lambda_{\rm min}$. In practice, this means to search for the $\lambda_{\rm opt}$ at which $\chi^2_V(\lambda_{\rm opt}) = \chi^2_V(\lambda_{\rm min}) + \Delta$ where Δ is the uncertainty of χ^2_V at $\lambda = \lambda_{\rm min}$. This is referred to as the 1-σ rule [88, 89].

D. Bootstrap and non-Gaussian uncertainties

Although it is common knowledge, we briefly mention the bootstrap technique to keep the presentation as pedagogical as possible, and we specify the way bootstrapping is implemented in this study. For a very similar procedure, see Refs. [6, 90]. Once a model is selected, the propagation of uncertainties from data to results can be carried out using bootstrap. Here, the results are given by the multipoles and the cusp parameter at threshold, β_0 (see Sec. III B 1 for discussion). This resampling technique allows to trace non-linear uncertainties and is, thus, in principle, superior to methods using the covariance matrix. Here, we repeatedly perform complete random resamplings of all data points according to their uncertainties and then refit each resampled data set. From each set of the resulting fit parameters, the multipoles and the cusp parameter β_0 are evaluated.

If the resulting distribution, e.g., of a multipole at fixed energy W, is Gaussian one can just estimate the variance and determine the final uncertainty by its square root. If the distribution is very skewed it is more meaningful to determine the 68% confidence level (CL) interval by cutting off the 16% largest and 16% smallest values.

This leads to a related comment concerning the fitting of the beam asymmetry Σ that is one of the considered observables in this study. Usually, the statistical uncertainty in polarization observables provided by experiment is treated as Gaussian in partial wave analysis, as if it originated from the measurement of a cross section in the limit of many counts. However, these observables O are ratios of the difference of positive Poisson distributions divided by their sum, i.e., they are restricted to $|O| < 1$. Thus, strictly speaking, those uncertainties cannot be regarded as Gaussian and maximizing the likelihood cannot be achieved by minimizing the χ^2. The bias is maximal for $|O| \approx 1$. The size of the beam asymmetry Σ is far from this limit at the low energies considered in this study and we neglect the bias here.

III. RESULTS

A. LASSO in a Benchmark Model

For a controlled test of the discussed methods, and before dealing with experimental data, we test our ideas with synthetic data. In doing so we proceed in the following way:

1. We generate synthetic data from a given set of multipoles and study to which precision and accuracy we can reconstruct that known set. To that end we first build a benchmark model ($\mathcal{B}$-model) which is a reduced version of the one described by Eqs. (1) and (3). We build it setting all the a_i to zero except for: a_0 and a_1 for the real parts of every S and P wave in Eq. (1) and a_0 in Eq. (3), totaling 9 parameters, i.e. 2 for each P-wave multipole and 3 for the S wave. All the imaginary

parts of the P waves are consequently set to zero. No D waves are included. If we include the D waves from the Born terms of photoproduction, this model would correspond to the one used in [23, 24, 35, 37] to analyze the experimental data. In this way, we keep the $\mathcal{B}$-model as realistic as possible. The synthetic data are generated around that solution at the same energies and scattering angles as the real data and with the same error bars.

2. These synthetic data are then analyzed with the full 46-parameter model as defined in the previous section, minimizing the penalized $\chi^2_T = \chi^2 + P$ for different λ according to Eq. (4).

In Fig. 2(a) the total χ^2_T and the contribution from data alone (χ^2) is indicated. The difference is given by the penalty P. Both curves rise as λ increases, and therefore, as discussed, one needs additional criteria to determine the optimal value, or range of values, for λ. In Fig. 2(b) the χ^2 per degree of freedom χ^2/d.o.f. is shown. It exhibits a minimum at around $\lambda = 3$ which already provides an initial impression about where to look for the simplest model. Yet, note that a simple Pearson's χ^2 test at 90% lower CL would rule out all fits up to $\lambda \approx 4$ as overfits, such that it is difficult to justify the use of the χ^2/d.o.f. itself as a tool to determine the optimal value for λ.

The λ-dependence of the fit parameters a_i is shown in Fig. 2(c). We have chosen here a logarithmic scale; on a linear scale, it becomes obvious that the parameters effectively approach zero once the penalization is large enough. Yet, the picture shows that, sometimes, parameters that had died out can in principle reappear at larger values of λ. Figure 2(c) suggests that for $\lambda > 3$ many parameters are effectively zero and the optimal value for λ is expected in that region. Anticipating the final result, the figure shows effectively non-zero parameters in red while effectively zero parameters are shown in gray. The horizontal line indicates the cut-off below which a parameter is counted as zero. It is remarkable to observe the parameters drop by three orders of magnitude because it means that LASSO is also capable of disentangling the extreme correlations between parameters present at $\lambda = 0$, as the unnaturally large parameter values at $\lambda = 0$ indicate.

As Fig. 2(d) shows, the AIC(c) and BIC criteria confirm this picture quantitatively, with all three considered criteria exhibiting a minimum at the same $\lambda = 3$. The BIC shows the cleanest signal as the region to the left of the minimum shows the steepest slope. The cross-validation χ^2_V shown in Fig. 2(e), obtained through 5-fold validation [77, 78], exhibits a broad, shallow minimum from $\lambda \approx 2.5$ through $\lambda \approx 3.5$ which does not very well determine the optimal λ. To have an impression of what the simplest allowed model according to the 1-σ rule [77, 78] could be, we can continue the upper end of the error bar, at the optimal $\lambda = 3$ found before, horizontally until it intersects with the central value of χ^2_V at $\lambda = 3.8$. This indicates the simplest model compatible with cross validation within errors. Combining the findings from the AIC(c), BIC and cross validation, vertical

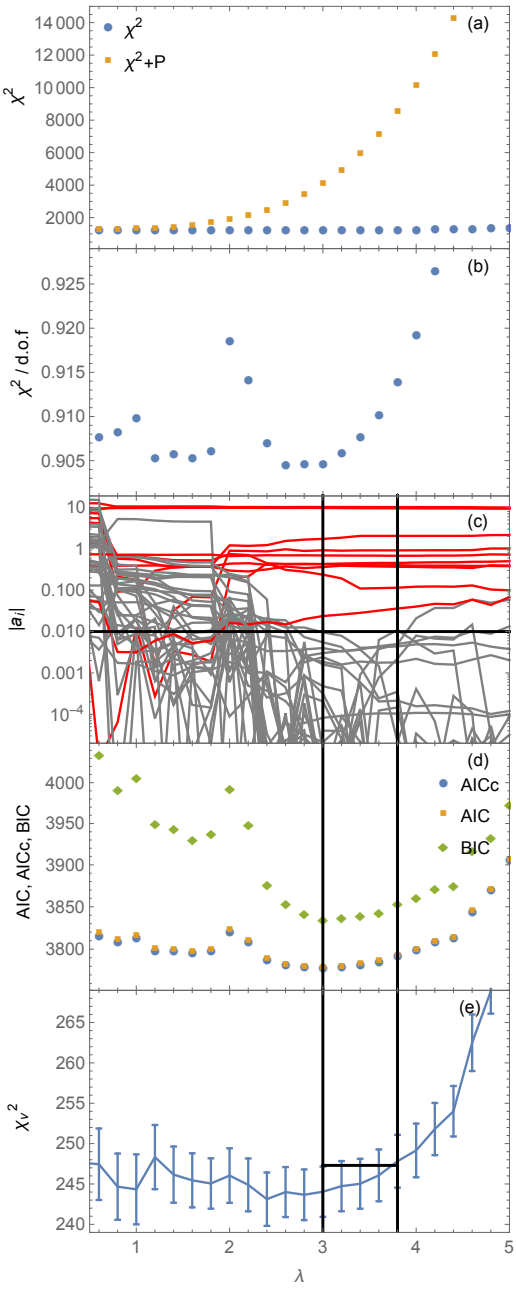

FIG. 2. LASSO, information criteria, and cross validation for the $\mathcal{B}$-model. (a) Total $\chi^2_T(\lambda) = \chi^2(\lambda) + P(\lambda)$ with penalty P (orange points) and $\chi^2(\lambda)$ from data alone (blue points). (b) The χ^2/d.o.f. with the number of parameters dynamically updated for each λ. (c) Absolute value of the parameters a_i as function of λ in a logarithmic scale. The red lines indicate the finally chosen parameters, the gray lines show the unnecessary parameters. (d) AIC, AICc and BIC. (e) The cross-validation χ^2_V. See text for further explanations.

lines at $\lambda = 3$ and $\lambda = 3.8$ enclose a region of optimal λ. If we return now to Fig. 2 (c), it becomes apparent that in that region the number of effective non-zero parameters indeed does not change, meaning that the precise value of λ to choose the simplest model does not matter as long as it is within that range.

3. Although trivial, the last step consists in setting all parameters that have been ruled out by LASSO to exactly zero and refit the model with the remaining parameters to the synthetic data. In summary, the LASSO reduces a model with 46 parameters to a simpler one with 10 parameters, remarkably close to the true number of parameters (9).

1. *Discussion*

In Fig. 3, the different steps in the model selection process are illustrated. The blue lines show the partial waves used to generate the synthetic data and provide, thus, the benchmark solution ($\mathcal{B}$-model) to be reproduced. The over-fit at $\lambda = 0$ is indicated with the orange lines and bands. All uncertainty bands have been obtained by bootstrap as explained in Sec. II D. Clearly, at the cost of fitting fluctuations, the benchmark solution is missed, even within the large uncertainties. Note that this is a 46-parameter fit to altogether 1373 data points, meaning that from these numbers it is per se not obvious that we have an overfit. However, the χ^2 of this fit ($\chi^2 = 1232$) is below the lower 95% confidence limit of Pearson's χ^2 test at $\chi^2 = 1243$, indicating overfitting.

We stress here the well-known defect of the rule-of-thumb, that a χ^2/d.o.f. ≈ 1 indicates a good fit. For a few points, a χ^2/d.o.f. significantly larger than one might be perfectly acceptable, while a χ^2/d.o.f.$= 1.1$ for a fit to 10,000 data points is very bad. Pearson's χ^2 test gives here a better answer with respect to under-fitting.

Finally, the 46-parameter fit has also greatly reduced predictability. Here, we have only included data up to a cut-off of $W \leq 1120$ MeV as indicated with the vertical lines in the figure. As expected, beyond that value, the fit shows large oscillations and an uncontrolled increase of the uncertainties (orange bands).

The fit at $\lambda = 3$ is shown with the brown lines in Fig. 3 (as discussed before, we could have taken any value up to $\lambda = 3.8$). Removing all parameters that are effectively zero and refitting the remaining 10 parameters, the final solution is indicated with the red lines and uncertainty bands. It remarkably well reproduces the benchmark solution (blue lines) and even produces reasonable and well-constrained partial waves beyond the range of fitted data (indicated with vertical lines).

The one additional parameter of the 10-parameter fit, compared to the benchmark solution (9 parameters), is given by a_1 in Eq. (3). Yet, as Fig. 3 shows, the benchmark solution for Im E_{0+} is rather well reproduced. At higher energies, the benchmark solution is just slightly outside the red narrow uncertainty band indicating the 68% confidence region.

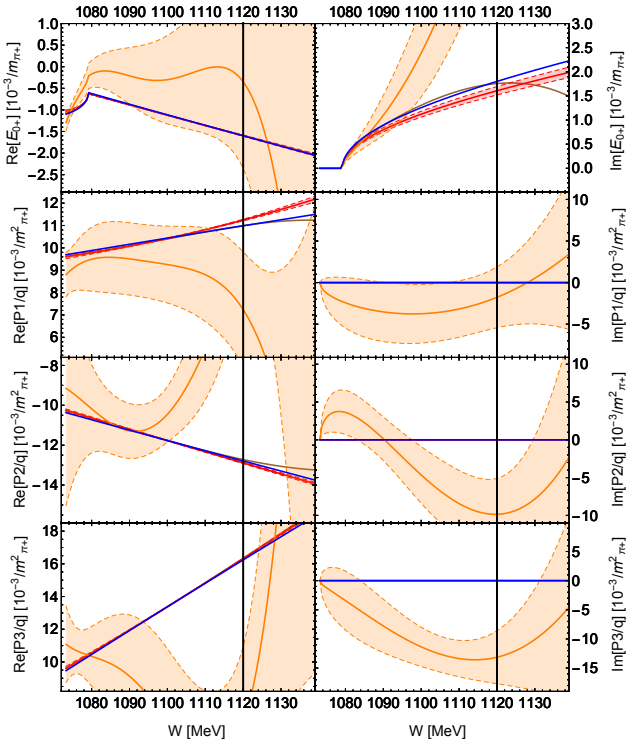

FIG. 3. S and P-wave partial waves from the fits to the $\mathcal{B}$-model (D waves are not shown because the result obtained was identically zero as expected). The blue lines show the $\mathcal{B}$-model used to generate synthetic data (benchmark solution). The orange lines and bands show the unconstrained 46-parameter fit with zero penalty $\lambda = 0$. The brown lines show the fit at optimal $\lambda = 3$. Removing the parameters that are effectively zero and refitting produces the final result indicated with red lines and uncertainty bands. The vertical lines at $W = 1120$ MeV indicate the upper limit of fitted data. Note that for Im P_i, all but the unconstrained (orange) results for $\lambda = 0$, are identically zero. Also in other cases the curves lie on top of each other such that only the blue line is visible. Uncertainties are computed using bootstrap.

It is also remarkable that in the simplest model LASSO is capable of setting all D waves to zero, which are finite in the over-parametrized model ($\lambda = 0$) but absent in the benchmark solution. Furthermore, the imaginary parts of all P waves are found to vanish (cf. Fig. 3), in agreement with the benchmark solution.

One should note that in the fit to photoproduction data there is an overall-phase problem that is usually solved by fixing the phase of one multipole. Here, the full 46-parameter model indeed contains only real D waves while the phase of all other multipoles are undetermined. Yet, those D waves are very small (at $\lambda = 0$) such that the overall phase problem is in fact not very well fixed. Part of the wide error bands for the $\lambda = 0$ case can,

thus, be attributed to a rather uncontrolled variation of the overall phase, leading to a large reshuffling between real and imaginary parts and causing the very poorly determined multipoles.

Finally, one can discuss other goodness-of-fit criteria beyond Pearson's χ^2 test. For example, it is common to find a best fit by combining the χ^2 test –that is rather restrictive to under-fitting but more tolerant to over-fitting– with the F test that determines, for nested models, if a more complex model leads to a significant improvement. For that, the χ^2s of two models with k (model 1) and $m+k$ (model 2) parameters fitted to n data points are compared. If the true values of the m extra parameters of the more complex model vanish, it can be shown that

$$y = \frac{(\chi_1^2 - \chi_2^2)/k}{\chi_2^2/(n-m-k)} \tag{7}$$

is $F(k, n-m-k)$ distributed [91]. For the explicit form of the F-distribution, see, e.g., Ref. [80]. A value of y beyond a chosen CL limit thus indicates that the more complex model 2 is significantly better than the simpler model 1 (which cannot be judged from the χ^2 values alone). Here, we find that y obtained from the 46-parameter fit and the optimal (simplest) 10-parameter fit is $y = 1.64$ which is below the 90% CL interval ending at $y = 2.63$, indicating that the overfit is indeed not significantly better than the simplest fit.

Other common goodness-of-fit-criteria are the Shapiro-Wilk [92], Anderson-Darling [93, 94], and the Kolmogorov-Smirnov [94–96] tests. They provide means to compare a sampled probability distribution against a theoretically expected one. In particular, they can be used to test fit residuals against their expected Normal $N[0,1]$ distribution. This applies also for weighted data as considered here as long as the fit residuals are divided by the respective experimental (statistical) uncertainty. In particular, these tests are by construction more sensitive than Pearson's χ^2 test because, on one hand, they are sensitive to the sign of the fit residual and not only its square, and, more importantly, they test the entire distribution instead of only the bulk property given by the sum of individual χ^2. The p-values of the three tests are shown in Fig. 4 as a function of λ.

For all considered values of λ the p-values are high; usually, fits with $p > 0.05$ are considered acceptable. For small values of λ the tests score high although one is in the region of over-fitting. In the region of optimal λ between $\lambda \approx 3$ and $\lambda \approx 4$, the p values are also consistently $p > 0.4$; however, there is no clear trend that would allow one to use these criteria themselves to find an optimal value for λ.

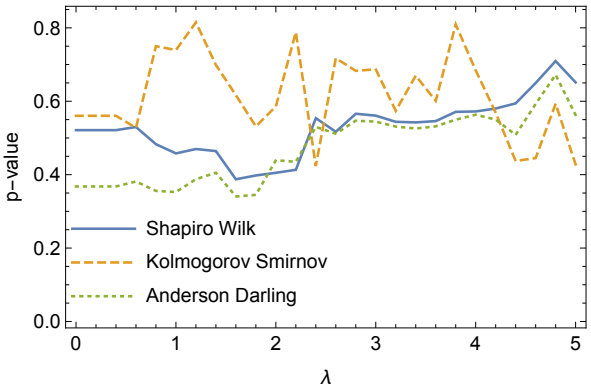

FIG. 4. Different goodness-of-fit criteria as a function of the penalty parameter λ.

B. LASSO for real data

For the reaction $\gamma p \to \pi^0 p$, precise low-energy data for the differential cross section $d\sigma/d\Omega$ and photon beam asymmetry Σ are available from MAMI [23]; for earlier measurements see Refs. [97–100]. The target polarization differential cross section $d\sigma_T/d\Omega$ has been measured recently at MAMI as well [35]. At energies beyond those considered here, these observables have also been remeasured with unprecedented precision [101, 102]. Electroproduction of π^0 mesons close to threshold has been recently measured by CLAS [103]. We analyze here the data from [23, 35] for $d\sigma/d\Omega$, Σ, and $d\sigma_T/d\Omega$ from the $\pi^0 p$ threshold up to $W = 1120$ MeV.

For the analysis of real data, we have slightly reduced the model space used for the $\mathcal{B}$-model. Before, the D waves were kept real to fix the overall-phase ambiguity. However, this is not effective at low energies where D waves are extremely small as discussed. Then, for complex S and P waves, the ambiguity reappears at low energies. We have therefore set all imaginary parts of the P waves to zero. For the D waves themselves, we have kept them fixed at the real values given by the Born terms of photoproduction. This is the same procedure as in [23]. In total, the number of available parameters is 23. Note that none of these changes have to do with the model selection methods discussed here. The former are additional constraints imposed from other considerations, because we aim at a result that is comparable with the analysis of [23].

Performing the LASSO scan as in the case of the $\mathcal{B}$-model, in combination with the information theory criteria and cross validation, we obtain the results shown in Fig. 5. The fit is compatible with Pearson's χ^2 test for any λ shown. The comparison of Fig. 5 with Fig. 2 shows quite similar features. Again, the BIC delivers the most pronounced minimum, or minimal plateau. We choose an optimal $\lambda = 2.8$ which coincides with the minimum of the BIC which is also the last point of the plateau. For the cross validation χ^2_V there is no minimum at all in this case. Only an upper bound for λ can be determined by proceeding as before for the $\mathcal{B}$-model, i.e., continuing the upper end of the error bar at $\lambda = 2.8$ to the right as indicated (in analogy to the 1-σ rule). This leads to

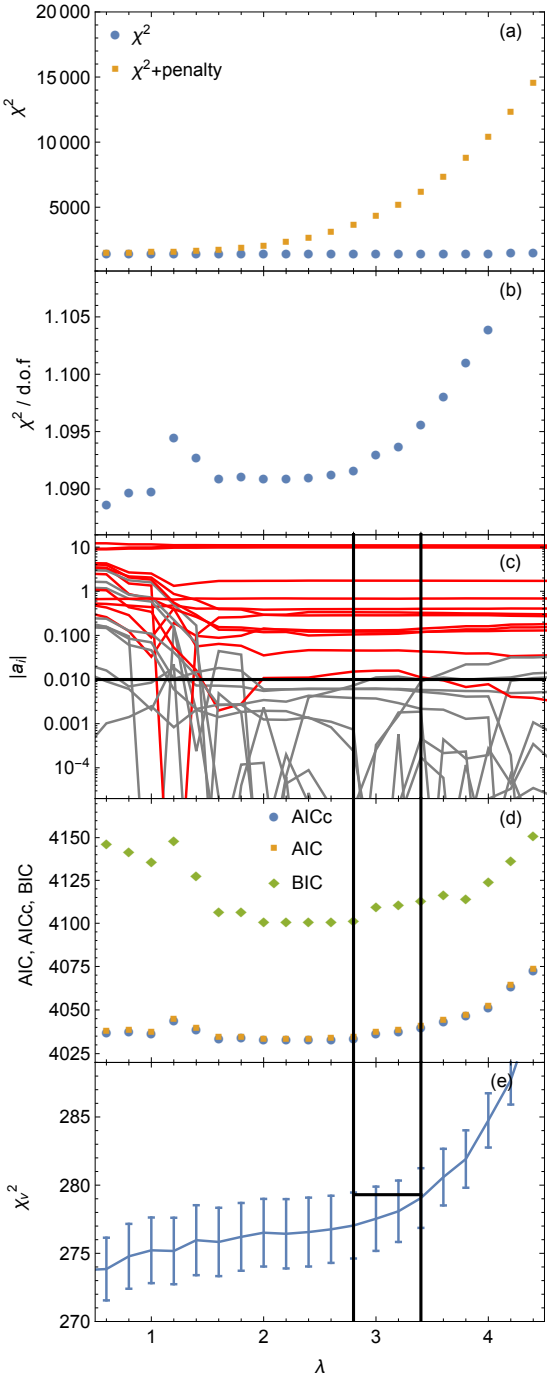

FIG. 5. The λ dependence of the same quantities as discussed in Fig. 2 but for the situation with actual data from experiment.

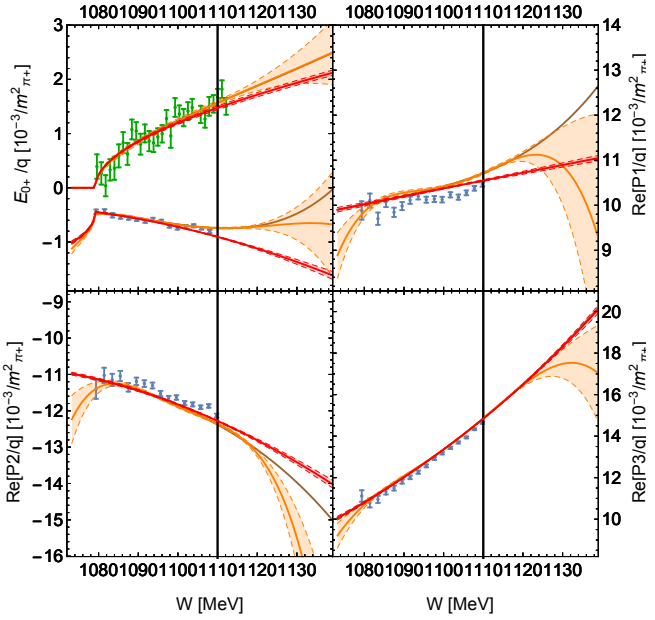

FIG. 6. S- and P-wave partial waves from the analysis of real data. Labeling of the curves and uncertainty bands as in Fig. 3. The simplest model is indicated with the red lines and bands. The blue circles show the single-energy solution of [23] for the real parts of S- and P-wave partial waves. The green triangles indicate the single-energy extraction of Im E_{0+} from [35]. Uncertainties are computed using bootstrap.

a maximal value of $\lambda = 3.4$ that is compatible within errors. As Fig. 5(c) shows, a parameter rises briefly above the cut-off threshold for values of $2.8 < \lambda < 3.4$. We do not see this as a problem but rather as a feature of the LASSO method demonstrating that the method indeed scans a large classes of models, and even rehabilitates parameters that were found to be zero for smaller λ. From the discussion it becomes clear that in this case it is not possible to fix the precise number of parameters without doubts; yet, we have seen before for the $\mathcal{B}$-model that it is possible to determine that number approximately. Overall, the number of model parameters is reduced from 23 to 13 (in the $\mathcal{B}$-model: from 46 to 10).

1. *Discussion*

The partial waves are shown in Fig. 6. They share similar properties as observed for the $\mathcal{B}$-model in Fig. 3.

For the unconstrained, $\lambda = 0$ case with 23 parameters, the uncertainty band in the region of fitted data is much narrower than in the $\lambda = 0$ case for the $\mathcal{B}$-model. This may be partly explained by the smaller maximal number of parameters available; another reason is the discussed overall-phase ambiguity which is avoided here by allowing only for real P-wave multipoles.

For the intermediate range of energies, the solutions

and uncertainty bands are sometimes difficult to distinguish in Fig. 6. A closer inspection reveals that the uncertainty bands of the 23-parameter fit are still wider than the ones of the 13-parameter simplest model. We also observe largely widened error bands for the unconstrained fit at very low energies and beyond the fitted region (vertical bands) indicating the reduced predictability of the $\lambda = 0$ fit. The simplest model (red lines and bands) shows good qualitative agreement with the SE solution from [23] for the real parts of the partial waves, although they do not coincide perfectly. There are differences such that here we allow for a variation of the cusp parameter at threshold, $\beta_0 = a_0 m_{\pi^+}^{-1}$ in Eq. (3), which was held fixed in [23] and, second, we include here the new $d\sigma_T/d\Omega$ data from [35]. The optimal (simplest) solution is also very close to the result of [48] (not shown here) that includes isospin breaking in a similar way as implemented here, fulfills Fermi-Watson's theorem and extends up to the resonance region.

The imaginary part of E_{0+} of the simplest model agrees well with the SE extraction performed in [35] (green data points) as Fig. 6 shows. For the cusp parameter at threshold we obtain $\beta_0 = (2.41 \pm 0.05) \cdot 10^{-3}\, m_{\pi^+}^{-1}$. This agrees with the value from [35]: $\beta_0 = (2.2 \pm 0.2[\text{stat.}] \pm 0.6[\text{syst.}]) \cdot 10^{-3}\, m_{\pi^+}^{-1}$. Note, however, the very different size of the statistical uncertainty which comes here from a global fit to all data while in [35] it was obtained from a two-parameter fit to the SE solutions shown in Fig. 6 with the green triangles. Note that the systematic uncertainties of β_0 are estimated to be 0.6 in the above units [35] which is larger than the statistical ones. This shows that systematic effects are not negligible, but we have refrained from analyzing these effects here because the focus of this paper is different. It should be mentioned that the values for β_0 found here and in [35] are smaller than the value of $\beta_0 = 3.35 \cdot 10^{-3}\, m_{\pi^+}^{-1}$ obtained in ChPT calculations [39, 40] using pionic atoms data.

In Figs. 7, 8, and 9, the experimental data and the best fit result are shown, corresponding to the 13 parameter fit indicated with the red lines in Fig. 6. The description of the data is consistently good. In particular, the inclusion of the new $d\sigma_T/d\Omega$ data does not have much influence in the multipoles other than reducing their uncertainties.

As a concluding remark it should be mentioned that the AIC and BIC compare the relative quality of models and are, thus, also applicable in situations in which no good fit in absolute terms can be achieved (e.g., when the achievable χ^2 is too large for all models). This robustness is a relevant aspect in the analysis of photoproduction data, especially for excited baryons at higher energies. There, data from different experiments are sometimes plagued by underestimated systematic uncertainties or even contradictory (for the selection of consistent data sets see, e.g., Refs. [104, 105]). Then, no good χ^2 might be achieved and one has to proceed using relative comparisons of models as proposed here.

IV. CONCLUSIONS

The LASSO provides a tool to scan large classes of models through the ability to set fit parameters to zero. Together with criteria from information theory and cross validation, it is possible to select a *simplest* model with a minimum of fit parameters. Here, we concentrate on single-meson photoproduction. The methods are, however, not restricted to this case and also applicable, e.g., in the context of excited meson production experiments as recently discussed by Guegan, Hardin, Stevens, and Williams [79].

In the example of pion photoproduction at low energies, we show that the LASSO in combination with additional criteria decreases the uncertainties of extracted multipoles and increases predictability, as expected. First, a benchmark model ($\mathcal{B}$-model) was considered to show these properties. The overall-phase problem was not entirely fixed in that model. Still, LASSO showed capable of simplifying this partially ill-defined problem, reducing an initial number of 46 fit parameters to 10 and recovering the known benchmark solution with remarkable accuracy and precision. In particular, properties of the benchmark solution such as vanishing imaginary parts of the P waves and absence of D waves could be detected. The considered criteria for the determination of the optimal penalty λ –AIC, AICc, and BIC– consistently exhibit minima at the same value of λ. The 1-σ rule in cross validation also defines a whole range of λ to be considered as optimal, and indeed in that range the number of relevant parameters stays constant.

When moving on to the analysis of real data, similar observations could be made. The simplest model shows remarkable agreement with single-energy extractions of multipoles from other studies. Additionally, these techniques provide more stable and reliable extrapolations of the models for energies above and below the fitted region. This feature can be exploited to determine physical observables at threshold, e.g. the discussed cusp parameter β_0. Besides their usefulness in the analysis of experimental data, the proposed methods could also be applied in the analysis of lattice QCD eigenvalues because the infinite-volume extrapolation of coupled-channel systems on the lattice necessarily requires an expansion of the amplitude in energy [106], similar to what has been discussed here. The discussed techniques also promise for a systematic and automatic work flow in the analysis of the excited baryon and meson spectra, providing a perspective towards finding more conclusive answers in hadron spectroscopy.

ACKNOWLEDGMENTS

The authors thank Maxim Mai for discussions. This work is supported by the National Science Foundation (PIF grant No. 1415459) and by The George Washington University through the Columbian College Facilitat-

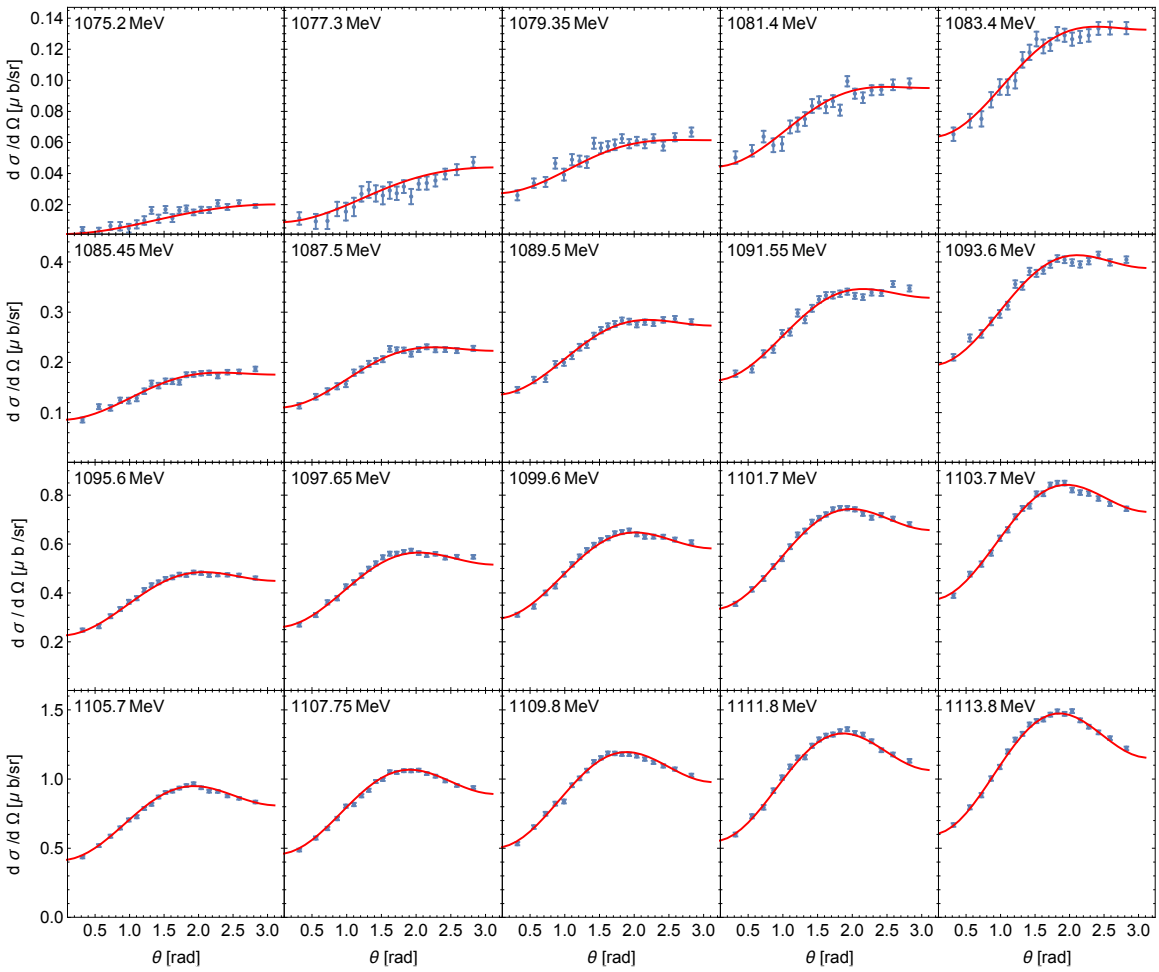

FIG. 7. Differential cross section. Data from [23]. The labels in the plot indicate the scattering energy W in MeV. The red lines show the central value of the simplest model as determined in this study (indicated with the red lines in Fig. 6).

ing Funds (CCFF). M.D. is also supported by the U.S. Department of Energy, Office of Science, Office of Nuclear Physics under contract DE-AC05-06OR23177 and grant No. DE-SC0016582. C.F.-R. is supported in part by CONACYT (Mexico) under grant No. 251817, by research grant IA101717 from PAPIIT-DGAPA (UNAM) and by Red Temática CONACYT de Física de Altas Energías (Red FAE, Mexico).

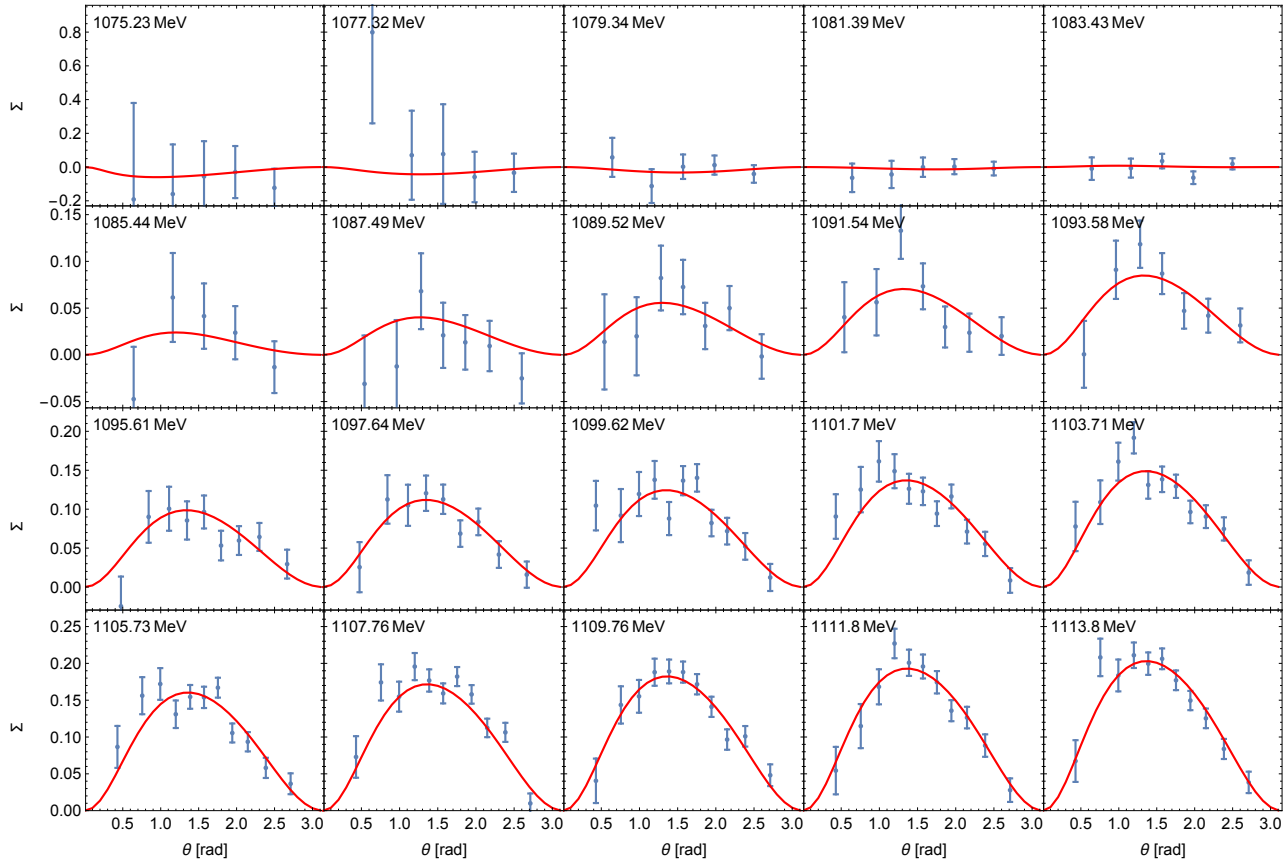

FIG. 8. Beam asymmetry. Data from [23]. See Fig. 7 for further description.

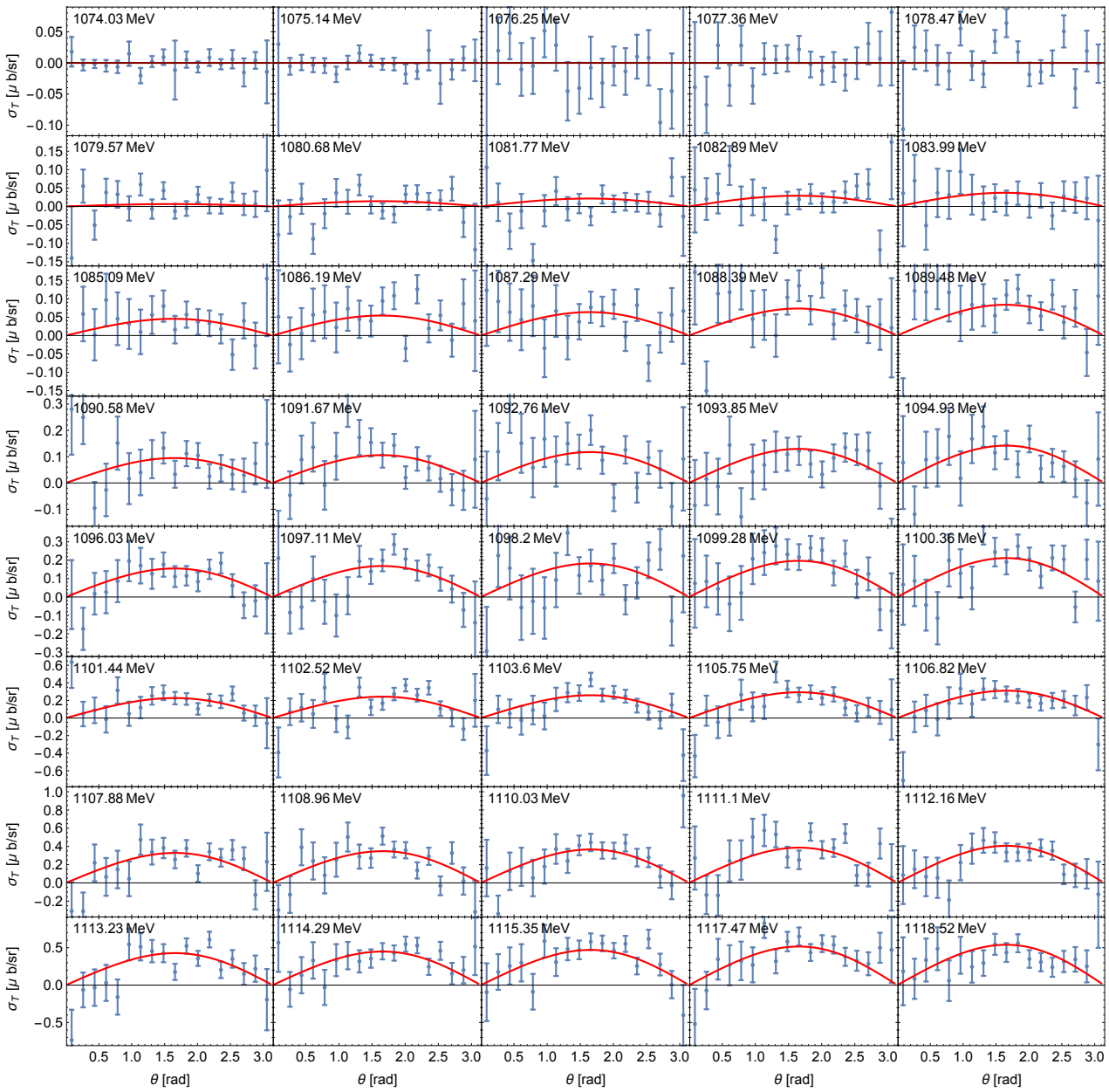

FIG. 9. Polarized differential cross section $\sigma_T = d\sigma/d\Omega\, T$. Data from [35]. See Fig. 7 for further description.

[1] M. Battaglieri *et al.*, Acta Phys. Polon. B **46**, 257 (2015).
[2] R. A. Briceño *et al.*, Chin. Phys. C **40**, 042001 (2016).
[3] J. J. Dudek, M. R. Shepherd, and R. E. Mitchell, Nature **534**, 487 (2016).
[4] C. B. Lang, L. Leskovec, M. Padmanath and S. Prelovsek, arXiv:1610.01422 [hep-lat].
[5] C. Fernández-Ramírez, J. Phys.: Conf. Ser. **761**, 012078 (2016).
[6] C. Fernández-Ramírez, I. V. Danilkin, D. M. Manley, V. Mathieu, and A. P. Szczepaniak, Phys. Rev. D **93**, 034029 (2016).
[7] T. Abe *et al.*, *Belle II Technical Design Report* [arXiv:1011.0352 [hep-ex]].
[8] T. Aushev *et al.*, *Physics at Super B Factory* [arXiv:1002.5012 [hep-ex]].
[9] D. M. Asner *et al.*, *Physics at BESIII* [arXiv:0809.1869 [hep-ex]].
[10] CLAS Collaboration, *CLAS12 Technical Design Report* (2008), https://www.jlab.org/Hall-B/clas12_tdr.pdf.
[11] H. Evans [ALICE, ATLAS, CMS and LHCb Collaborations], *Production and Spectroscopy of Heavy Hadrons at the LHC* [arXiv:1110.5294 [hep-ex]].
[12] P. Abbon *et al.* [COMPASS Collaboration], Nucl. Inst. Meth. A **577**, 455 (2007) [arXiv:hep-ex/0703049].
[13] K. H. Althoff *et al.*, Nucl. Inst. Meth. **61**, 1 (1968).
[14] K. H. Althoff *et al.*, Part. Accel. **27**, 101 (1990).
[15] M. R. Shepherd [GlueX Collaboration], AIP Conf. Proc. **1182**, 816 (2009).
[16] K. Tanida, JPS Conf. Proc. **10**, 010019 (2016).
[17] M. Silarski [KLOE-2 Collaboration], *The KLOE-2 experiment at DAFNE* [arXiv:1609.05467 [hep-ex]].
[18] A. A. Alves *et al.* [LHCb Collaboration], JINST **3** S08005 (2008).
[19] A. Denig, AIP Conf. Proc. **1735**, 020006 (2016).
[20] J. Messchendorp [PANDA Collaboration], *The PANDA Experiment at FAIR - Subatomic Physics with Antiprotons* [arXiv:1610.02804 [nucl-ex]].
[21] Y. Wunderlich, R. Beck, and L. Tiator, Phys. Rev. C **89**, 055203 (2014).
[22] R. L. Workman, L. Tiator, Y. Wunderlich, M. Döring, and H. Haberzettl, arXiv:1611.04434 [nucl-th].
[23] D. Hornidge *et al.* [A2 and CB-TAPS Collaborations], Phys. Rev. Lett. **111**, 062004 (2013)
[24] C. Fernández-Ramírez, EPJ Web Conf. **73**, 04007 (2014).
[25] V. Bernard, N. Kaiser, and U.-G. Meißner, Phys. Lett. B **382**, 19 (1996).
[26] V. Bernard, N. Kaiser, and U.-G. Meißner, Phys. Lett. B **378**, 337 (1996).
[27] V. Bernard, N. Kaiser, and U.-G. Meißner, Z. Phys. C **70** (1996) 483.
[28] V. Bernard, N. Kaiser, U.-G. Meißner, and A. Schmidt, Nucl. Phys. A **580**, 475 (1994).
[29] V. Bernard, N. Kaiser, T.-S. H. Lee, and U.-G. Meißner, Phys. Rept. **246**, 315 (1994).
[30] V. Bernard, N. Kaiser, and U.-G. Meißner, Nucl. Phys. B **383**, 442 (1992).
[31] V. Bernard, N. Kaiser, J. Gasser, and U.-G. Meißner, Phys. Lett. B **268**, 291 (1991).
[32] V. Bernard, B. Kubis, and U.-G. Meißner, Eur. Phys. J. A **25**, 419 (2005).
[33] C. Fernández-Ramírez and A. M. Bernstein, Phys. Lett. B **724**, 253 (2013).
[34] M. Hilt, S. Scherer, and L. Tiator, Phys. Rev. C **87**, 045204 (2013).
[35] S. Schumann *et al.* [A2 Collaboration], Phys. Lett. B **750**, 252 (2015).
[36] C. Fernández-Ramírez, A. M. Bernstein, and T. W. Donnelly, Phys. Rev. C **80**, 065201 (2009).
[37] C. Fernández-Ramírez, PoS CD **12**, 065 (2013).
[38] V. Baru, C. Hanhart, M. Hoferichter, B. Kubis, A. Nogga, and D. R. Phillips, Phys. Lett. B **694**, 473 (2011).
[39] M. Hoferichter, B. Kubis, and U.-G. Meißner, Phys. Lett. B **678**, 65 (2009).
[40] V. Baru, C. Hanhart, M. Hoferichter, B. Kubis, A. Nogga, and D. R. Phillips, Nucl. Phys. A **872**, 69 (2011).
[41] A. N. Hiller Blin, T. Ledwig, and M. J. Vicente Vacas, Phys. Lett. B **747**, 217 (2015).
[42] A. N. Hiller Blin, T. Ledwig, and M. J. Vicente Vacas, Phys. Rev. D **93**, 094018 (2016).
[43] M. Hilt, B. C. Lehnhart, S. Scherer, and L. Tiator, Phys. Rev. C **88**, 055207 (2013).
[44] C. Fernández-Ramírez, A. M. Bernstein, and T. W. Donnelly, Phys. Lett. B **679**, 41 (2009).
[45] A. M. Bernstein, Phys. Lett. B **442**, 20 (1998).
[46] G. Knöchlein, D. Drechsel, and L. Tiator, Z. Phys. A **352**, 327 (1995).
[47] A. M. Bernstein, M. W. Ahmed, S. Stave, Y. K. Wu, and H. R. Weller, Annu. Rev. Nucl. Part. Sci. **59**, 115 (2009).
[48] D. Rönchen, M. Döring, F. Huang, H. Haberzettl, J. Haidenbauer, C. Hanhart, S. Krewald, U.-G. Meißner, and K. Nakayama, Eur. Phys. J. A **50**, 101 (2014); Erratum: [Eur. Phys. J. A **51**, 63 (2015)].
[49] D. Rönchen, M. Döring, H. Haberzettl, J. Haidenbauer, U.-G. Meißner, and K. Nakayama, Eur. Phys. J. A **51**, 70 (2015).
[50] H. Kamano, S. X. Nakamura, T.-S. H. Lee, and T. Sato, Phys. Rev. C **88**, 035209 (2013).
[51] F. Huang, M. Döring, H. Haberzettl, J. Haidenbauer, C. Hanhart, S. Krewald, U.-G. Meißner, and K. Nakayama, Phys. Rev. C **85**, 054003 (2012).
[52] L. Tiator, S. S. Kamalov, S. Ceci, G. Y. Chen, D. Drechsel, A. Švarc, and S. N. Yang, Phys. Rev. C **82**, 055203 (2010).
[53] C. Fernández-Ramírez, E. Moya de Guerra, and J. M. Udías, Ann. Phys. (NY) **321**, 1408 (2006).
[54] C. Fernández-Ramírez, E. Moya de Guerra, and J. M. Udías, Phys. Lett. B **660**, 188 (2008).
[55] A. Gasparyan and M. F. M. Lutz, Nucl. Phys. A **848** 126 (2010).
[56] B. Juliá-Díaz, T.-S. H. Lee, A. Matsuyama, T. Sato, and L. C. Smith, Phys. Rev. C **77** 045205 (2008).
[57] R. L. Workman, M. W. Paris, W. J. Briscoe, and I. I. Strakovsky, Phys. Rev. C **86**, 015202 (2012).
[58] A. V. Anisovich, R. Beck, E. Klempt, V. A. Nikonov, A. V. Sarantsev, and U. Thoma, Eur. Phys. J. A **48**, 15 (2012).
[59] D. Drechsel, S. S. Kamalov, and L. Tiator, Eur. Phys.

J. A **34**, 69 (2007).
[60] A. Sibirtsev, J. Haidenbauer, F. Huang, S. Krewald, and U.-G. Meißner, Eur. Phys. J. A **40**, 65 (2009).
[61] V. Mathieu, G. Fox, and A. P. Szczepaniak, Phys. Rev. D **92**, 074013 (2015).
[62] V. Mathieu, I. V. Danilkin, C. Fernández-Ramírez, M. R. Pennington, D. Schott, A. P. Szczepaniak, and G. Fox, Phys. Rev. D **92**, 074004 (2015).
[63] J. Nys, V. Mathieu, C. Fernández-Ramírez, A. N. Hiller Blin, A. Jackura, M. Mikhasenko, A. Pilloni, A. P. Szczepaniak, G. Fox, and J. Ryckebusch [JPAC Collaboration], arXiv:1611.04658 [hep-ph].
[64] K. M. Watson, Phys. Rev. **95**, 228 (1954).
[65] E. Fermi, Suppl. Nuovo Cimento **2**, 17 (1955).
[66] A. V. Anisovich *et al.*, Eur. Phys. J. A **52**, 284 (2016).
[67] M. Döring, J. Revier, D. Rönchen, and R. L. Workman, Phys. Rev. C **93**, 065205 (2016).
[68] W. J. Briscoe, M. Döring, H. Haberzettl, D. M. Manley, M. Naruki, I. I. Strakovsky, and E. S. Swanson, Eur. Phys. J. A **51**, 129 (2015).
[69] J. Nys, J. Ryckebusch, D. G. Ireland, and D. I. Glazier, Phys. Lett. B **759**, 260 (2016).
[70] S. Ceci, M. Döring, C. Hanhart, S. Krewald, U.-G. Meißner and A. Švarc, Phys. Rev. C **84**, 015205 (2011).
[71] R. A. Arndt, Y. I. Azimov, M. V. Polyakov, I. I. Strakovsky, and R. L. Workman, Phys. Rev. C **69**, 035208 (2004).
[72] L. De Cruz, T. Vrancx, P. Vancraeyveld, and J. Ryckebusch, Phys. Rev. Lett. **108**, 182002 (2012).
[73] L. De Cruz, J. Ryckebusch, T. Vrancx, and P. Vancraeyveld, Phys. Rev. C **86**, 015212 (2012).
[74] S. Wesolowski, N. Klco, R. J. Furnstahl, D. R. Phillips, and A. Thapaliya, J. Phys. G **43**, 074001 (2016).
[75] R. J. Furnstahl, N. Klco, D. R. Phillips, and S. Wesolowski, Phys. Rev. C **92**, 024005 (2015).
[76] R. Tibshirani, J. R. Stat. Soc. B **58**, 267 (1996).
[77] T. Hasti, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Springer-Verlag, New York, USA, 2009).
[78] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning* (Springer-Verlag, New York, USA, 2013).
[79] B. Guegan, J. Hardin, J. Stevens, and M. Williams, JINST **10**, P09002 (2015).
[80] I. Miller and M. Miller, *John E. Freund's Mathematical Statistics with Applications* (Pearson, Boston, USA, 2014).
[81] V. Bernard, N. Kaiser, and U.-G. Meißner, Eur. Phys. J. A **11** (2001) 209.
[82] H. Akaike, IEEE Trans. Autom. Control **19**, 716 (1974).
[83] K. Burnham and D. Anderson, *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach* (Springer-Verlag, New York, 2003).
[84] J. E. Cavanaugh, Statist. Probab. Lett. **31**, 201 (1997).
[85] G. Schwarz, Ann. Stat. **6**, 461 (1978).
[86] N. Sato, W. Melnitchouk, S. E. Kuhn, J. J. Ethier, and A. Accardi, Phys. Rev. D **93**, 074005 (2016).
[87] D. Agadjanov, M. Döring, M. Mai, U.-G. Meißner, and A. Rusetsky, JHEP **1606**, 043 (2016).
[88] L. Breiman and P. Spector, Int. Stat. Rev. **60**, 291 (1992).
[89] R. Kohavi, *A study of cross-validation and bootstrap for accuracy estimation and model selection* in IJCAI'95 Proceedings of the 14th International Joint Conference on Artificial Intelligence, edited by C. S. Mellish (Morgan Kaufmann, San Francisco, CA, USA, 1995), p. 1137.
[90] A. N. Hiller Blin, C. Fernández-Ramírez, A. Jackura, V. Mathieu, V. I. Mokeev, A. Pilloni, and A. P. Szczepaniak, Phys. Rev. D **94**, 034002 (2016).
[91] L. Wilkinson and G. E. Dallal, Technometrics **23**, 377 (1981).
[92] S. S. Shapiro, M. B. Wilk, Biometrika **52**, 591 (1965).
[93] T. W. Anderson and D. A. Darling, Ann. Math. Stat. **23**, 193 (1952).
[94] M. A. Stephens, J. Amer. Statist. Assoc. **69**, 730 (1974).
[95] A. N. Kolmogorov, Giornale dell'Istituto Italiano degli Attuari **4**, 83 (1933).
[96] N. Smirnov, Ann. Math. Stat. **19**, 279 (1948).
[97] A. Schmidt *et al.*, Phys. Rev. Lett. **87**, 232501 (2001); Erratum: [Phys. Rev. Lett. **110**, 039903 (2013)].
[98] J. C. Bergstrom, R. Igarashi, and J. M. Vogt, Phys. Rev. C **55**, 2016 (1997).
[99] M. Fuchs *et al.*, Phys. Lett. B **368**, 20 (1996).
[100] R. Beck *et al.*, Phys. Rev. Lett. **65**, 1841 (1990).
[101] P. Adlarson *et al.* [A2 Collaboration], Phys. Rev. C **92**, 024617 (2015).
[102] J. R. M. Annand *et al.* [A2 Collaboration], Phys. Rev. C **93**, 055209 (2016).
[103] K. Chirapatpimol *et al.* [Hall A Collaboration], Phys. Rev. Lett. **114**, 192503 (2015).
[104] R. Navarro Pérez, J. E. Amaro, and E. Ruiz Arriola, Phys. Rev. C **88**, 064002 (2013) Erratum: [Phys. Rev. C **91**, 029901 (2015)].
[105] R. Navarro Pérez, J. E. Amaro, and E. Ruiz Arriola, Phys. Rev. C **89**, 064006 (2014).
[106] M. Döring and U.-G. Meißner, JHEP **1201**, 009 (2012).