

## Model-independent extrapolation of MUonE data with Padé and D-Log approximants

Diogo Boito<sup>1</sup>,<sup>2</sup> Cristiane Y. London<sup>1,2</sup>, Pere Masjuan<sup>2</sup>, and Camilo Rojas<sup>2</sup>

<sup>1</sup>*Instituto de Física de São Carlos, Universidade de São Paulo, CP 369, 13560-970, São Carlos, São Paulo, Brazil*

<sup>2</sup>*Grup de Física Teòrica, Departament de Física, Universitat Autònoma de Barcelona, and Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology (BIST), Campus UAB, E-08193 Bellaterra, Barcelona, Spain*



(Received 17 June 2024; accepted 27 July 2024; published 10 October 2024)

The MUonE experiment is designed to extract the hadronic contribution to the electromagnetic coupling in the spacelike region  $\Delta\alpha_{\text{had}}(t)$  from elastic  $e\mu$  scattering. The leading-order hadronic vacuum polarization contribution to the muon  $g-2$ ,  $a_{\mu}^{\text{HVP,LO}}$ , can then be obtained from a weighted integral over  $\Delta\alpha_{\text{had}}(t)$ . This, however, requires knowledge of  $\Delta\alpha_{\text{had}}(t)$  in the whole domain of integration, which cannot be achieved by experiment. In this work, we propose to use Padé and D-Log Padé approximants as a systematic and model-independent method to fit and reliably extrapolate the future MUonE experimental data, extracting  $a_{\mu}^{\text{HVP,LO}}$  with a conservative but competitive uncertainty, using no or very limited external information. The method relies on fundamental analytic properties of the two-point correlator underlying  $a_{\mu}^{\text{HVP,LO}}$  and provides lower and upper bounds for the result for  $a_{\mu}^{\text{HVP,LO}}$ . We demonstrate the reliability of the method using toy datasets generated from a model for  $\Delta\alpha_{\text{had}}(t)$  reflecting the expected statistics of the MUonE experiment.

DOI: [10.1103/PhysRevD.110.074012](https://doi.org/10.1103/PhysRevD.110.074012)

### I. INTRODUCTION

The recent measurements of the anomalous magnetic moment of the muon  $a_{\mu} = (g-2)/2$ , by the FNAL E989 experiment at Fermilab, in 2021 and 2023 [1,2], are in good agreement with the previous experimental result from the Brookhaven National Laboratory E821 experiment of 2006 [3]. The combination of the results leads to an experimental determination of  $a_{\mu}$  with an impressive uncertainty of only 0.19 ppm. As is well known, the 2020  $g-2$  theory initiative white paper [4] recommended result for  $a_{\mu}$  in the Standard Model (based on the results of Refs. [5–28]) is  $5.1\sigma$  lower than the new, combined experimental number—a tension that has attracted enormous attention in the past few years. This result relies on the dispersive description of the hadronic vacuum polarization (HVP) contribution to  $a_{\mu}$ ,  $a_{\mu}^{\text{HVP}}$ . If one employs instead the recent lattice QCD results for  $a_{\mu}^{\text{HVP}}$  obtained by the BMW Collaboration [29], the discrepancy between theory and experiment would be reduced to  $2.0\sigma$ .

Understanding the origin of the tension between the dispersive-based result and the lattice-based determination of  $a_{\mu}^{\text{HVP}}$  is of crucial importance. The detailed comparison is not completely straightforward [30–32] since in lattice QCD one has access to the Euclidean HVP, while the dispersive approach relies on data for  $e^+e^- \rightarrow (\text{hadrons})$  in the whole Minkowski domain. Independent information from other related processes may be crucial to fully resolve persistent discrepancies. A prominent example is the use of  $\tau$  decay data, which require a nontrivial treatment of isospin corrections [33,34]. In this context, the recently proposed MUonE experiment [35–37] would also be very welcome. The proposal is to extract the HVP in the Euclidean domain, directly from data, from the measurement of the elastic  $e\mu$  cross section using the 150-GeV muon beam from CERN's M2 beamline scattered off atomic electrons of a low- $Z$  target. The experiment could yield competitive results after three years of data taking and would be able to cover approximately 86% [38,39] of the integration interval required for the computation of  $a_{\mu}^{\text{HVP}}$  at leading order ( $a_{\mu}^{\text{HVP,LO}}$ ).

An important question is how to treat the remaining 14% of the  $a_{\mu}^{\text{HVP,LO}}$  integral not directly accessible to the MUonE experiment. In principle, one could simply resort to external information and use the dispersive approach, perturbative QCD, and/or lattice QCD results. Another option, arguably more interesting, is to extract  $a_{\mu}^{\text{HVP,LO}}$

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/). Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by SCOAP<sup>3</sup>.

exclusively from the MUonE data, which requires some form of extrapolation of the experimental results beyond the kinematically accessible region. This problem is, however, nontrivial since the experiment would have access to a narrow window in the Euclidean  $t$  variable, in the range  $-0.153 \leq t \leq -0.001 \text{ GeV}^2$ . In Ref. [38], a model inspired by one-loop QED is put forward as a fitting function to fit and extrapolate the MUonE data. A disadvantage of this approach is a potential model dependency that could bias the final results. An alternate strategy suggested in Ref. [39] relies on the extraction of derivatives of the hadronic contribution to the running of the electromagnetic coupling, supplemented with information from perturbative QCD and  $R(s)$  data. Another recent proposal, closer in spirit to our work, consists of using transfer theorems to build so-called “reconstruction approximants” which allow for a partial reconstruction of the HVP function to compute  $a_\mu^{\text{HVP}}$  [40]. Here, we propose the use of Padé approximants (PAs) and a variant of this method, as a systematic, simple, and model-independent way of fitting and extrapolating the MUonE results to compute  $a_\mu^{\text{HVP,LO}}$  from MUonE data.

The extrapolation of the HVP results in the Euclidean domain using PAs has been explored previously in the context of heavy-quark physics [41] and lattice QCD results [42]. The use of PAs in this problem is predicated on the fact that the HVP is a Stieltjes function [41]. In this case, convergence theorems for sequences of PAs apply and inequalities guarantee that certain PA sequences approach the function “from above” while others do so “from below” [41,43,44], providing a systematic way to bound the value of the function of interest. The same theorems play a crucial role in our work.

By construction, the usual PAs are not able to explore the Minkowski region coming from the Euclidean domain as they contain *only* poles and zeros, and the branch cuts can be, at best, emulated by the accumulation of singularities [44–46]. Therefore, we propose to accompany the PA study with the use of a variant of the method called D-Log Padé approximants, D-Logs [32,47,48], an extension of PAs which contain, by construction, not only poles and zeros but also branch cuts. In some cases, a systematic way to bound the value of the function is also provided pointing toward a similar convergence theorem. This can pave the way for future explorations where one could have a glimpse of the Minkowski region from fits in the Euclidean region in a model-independent and systematic way.

To assess the reliability and viability of our proposal, we adopt a simple but sufficiently realistic model for the HVP function introduced in Ref. [40]. We then build the approximants first to the exact Taylor expansion of the model function and later to pseudodata generated from the model. In this first step, we do not include uncertainties, as a proof of concept. We then generate realistic pseudodata following the expected uncertainties and kinematic range accessible to the MUonE experiment [35–38,49–51]. With

these datasets, we perform a systematic study of the use of PAs and D-Logs, as a way to fit and extrapolate MUonE data.

As we already mentioned, obtaining  $a_\mu^{\text{HVP}}$  solely from MUonE data is a nontrivial problem. Therefore, we perform a systematic study where we enlarge in each step the window in which we rely on extrapolated results. In our systematic investigation, we show that there is a trade-off in precision. It is possible to perform a reliable, robust, and model-independent extraction of  $a_\mu^{\text{HVP,LO}}$  using approximants solely from MUonE data, but with a somewhat larger error. An advantage of the method is that both PAs and D-Logs allow for a reliable estimate of the systematic error. If the window in which one uses the extrapolated results is reduced, the error diminishes, as could be expected.

This paper is organized as follows. In Sec. II, we introduce the basic elements to describe the HVP contribution to the  $(g-2)_\mu$  related to the running of the electromagnetic coupling constant. In Sec. III, we review the aspects of Padé theory that are the foundations of our work, in particular, Stieltjes functions and the convergence theorems of PAs and D-Logs applied to them. In Sec. IV, we present the model of Ref. [40] that we use to generate our toy datasets for the Euclidean HVP, while an example of the power of the convergence theorems is presented in Sec. V, where PAs and D-Logs are built from the exactly known Taylor series given by the model of Sec. IV. In Sec. VI, we employ our method in the idealized scenario where the data points have zero error. Section VII illustrates the application of our method to realistic datasets, following the expectations of the MUonE experiment. Our conclusions are given in Sec. VIII. Technical details about the fitting functions are relegated to the Appendix.

## II. THEORETICAL FRAMEWORK

The Standard Model computation of  $a_\mu$  can be divided into four different contributions, namely, from quantum electrodynamics (QED), electroweak effects, HVP, and hadronic light-by-light scattering. The dominant uncertainty arises from the HVP contribution [4], more specifically, from its leading-order contribution,  $a_\mu^{\text{HVP,LO}}$ . In the computation of  $a_\mu^{\text{HVP}}$  the main object is the polarization function associated with the electromagnetic current two-point correlator  $\Pi(q^2)$ , defined as

$$(q_\mu q_\nu - q^2 g_{\mu\nu})\Pi(q^2) = i \int d^4x \langle 0 | T \{ j_\mu^{\text{EM}}(x) j_\nu^{\text{EM}}(0) \} | 0 \rangle, \quad (1)$$

where the electromagnetic current is

$$j_\mu^{\text{EM}} = \frac{2}{3} \bar{u} \gamma_\mu u - \frac{1}{3} \bar{d} \gamma_\mu d - \frac{1}{3} \bar{s} \gamma_\mu s + \frac{2}{3} \bar{c} \gamma_\mu c + \dots \quad (2)$$

We define  $\bar{\Pi}(q^2) = \Pi(q^2) - \Pi(0)$ , and the function  $\bar{\Pi}(q^2)$  obeys the usual once-subtracted dispersion relation

$$\bar{\Pi}(q^2) = \frac{q^2}{\pi} \int_{m_\pi^2}^{\infty} ds \frac{\text{Im}\Pi(s)}{s(s - q^2 + i\epsilon)}. \quad (3)$$

In the dispersive approach,  $a_\mu^{\text{HVP,LO}}$  is obtained from the inclusive hadronic electroproduction cross section defined, with  $s = q^2$ , as

$$R(s) = \left( \frac{3s}{4\pi\alpha^2} \right) \sigma_{e^+e^- \rightarrow \text{hadrons}}(s) = 12\pi \text{Im}\Pi(s), \quad (4)$$

through the following weighted integral:

$$a_\mu^{\text{HVP,LO}} = \frac{\alpha^2}{3\pi^2} \int_{m_\pi^2}^{\infty} \frac{ds}{s} K(s) R(s), \quad (5)$$

where  $\alpha$  is the electromagnetic fine-structure constant and  $K(s)$  is the QED kernel function [52–54]

$$K(s) = \int_0^1 \frac{x^2(1-x)}{x^2 + (1-x)\frac{s}{m_\mu^2}} dx. \quad (6)$$

The analytical result for  $K(s)$  is given explicitly in Ref. [4].

An alternate representation for  $a_\mu^{\text{HVP,LO}}$  in terms of the correlator in the Euclidean,  $\Pi(Q^2)$  with  $Q^2 = -q^2 > 0$ , can be obtained interchanging the order of the integrals in  $s$  and  $x$  in Eq. (5) [55]. Using the analytical properties of  $\Pi(q^2)$  one can then write

$$a_\mu^{\text{HVP,LO}} = \frac{\alpha^2}{\pi} \int_0^1 dx (1-x) \Delta\alpha_{\text{had}}[t(x)], \quad (7)$$

where, following the notation employed by Bernecker and Meyer in Ref. [56], we defined

$$\Delta\alpha_{\text{had}}(t) = -4\pi \text{Re}[\bar{\Pi}_{\text{had}}(t)] \quad (8)$$

as the hadronic contribution to the running of the electromagnetic coupling  $\alpha$  and  $t$  is the spacelike variable given by

$$t = -\frac{x^2 m_\mu^2}{1-x}. \quad (9)$$

The MUonE experiment is designed to extract  $\Delta\alpha_{\text{had}}(t)$  from  $e\mu$  scattering data using 150-GeV muons scattered off atomic electrons. This allows, in principle, for a completely independent determination of  $a_\mu^{\text{HVP,LO}}$ . However, the experiment would be restricted approximately to the window  $x \in [0.2, 0.93]$  [35–38,49–51], which corresponds to  $-0.15 \lesssim t \lesssim -0.001 \text{ GeV}^2$ . To obtain  $a_\mu^{\text{HVP,LO}}$  from MUonE data without requiring external information, it is imperative to have a reliable method to extrapolate the MUonE data way outside the experimentally accessible window. With this purpose in mind, we will use a set of approximants as fitting functions to toy datasets for

$\Delta\alpha_{\text{had}}(t)$  simulating the expected results of the MUonE experiment.

Finally, uncertainty assessment is crucial and we want to keep track of the goodness of our extrapolation beyond the fit region. The extrapolation for  $0 \leq x < 0.2$  is safe, since it involves a small interval in  $t$ , not too far from the origin, namely,  $0 \leq t \lesssim 0.001 \text{ GeV}^2$ . The extrapolation for  $0.93 < x < 1$ , on the other hand, corresponding to  $0.15 \text{ GeV}^2 \lesssim t < \infty$ , is nontrivial and, in order to assess it carefully, we define  $a_\mu^{\text{HVP,LO}}(x_{\text{max}})$  as the partial contribution from  $x = 0$  up to  $x = x_{\text{max}}$  to  $a_\mu^{\text{HVP,LO}}$ , expressed as

$$a_\mu^{\text{HVP,LO}}(x_{\text{max}}) = \frac{\alpha^2}{\pi} \int_0^{x_{\text{max}}} dx (1-x) \Delta\alpha_{\text{had}}[t(x)]. \quad (10)$$

### III. PADÉ AND D-LOG PADÉ APPROXIMANTS

In this section, we give an overview of Padé theory including Padé approximants and their variant, D-Log Padé approximants, and a discussion of the advantages and disadvantages of each of them in the determination of  $a_\mu^{\text{HVP,LO}}$ .

A PA  $P_M^N(z)$  is a rational function given by

$$P_M^N(z) = \frac{Q_N(z)}{R_M(z)} = \frac{q_0 + q_1 z + \dots + q_N z^N}{1 + r_1 z + \dots + r_M z^M}, \quad (11)$$

where we adopted  $r_0 = 1$ . The standard technique to construct PAs to a function is by matching the first  $N + M + 1$  terms of the Taylor series expansion of Eq. (11) to that of the original function order by order, thereby fixing the coefficients of the PA unambiguously [43–45].

A D-Log Padé approximant, to which we refer simply as D-Log, in turn, is a variant that is very useful for functions with branch points or poles with higher multiplicity [44,47,48]. Let us consider, for example, the following function:

$$f(z) = A(z) \frac{1}{(\mu - z)^\gamma} + B(z), \quad (12)$$

where  $A(z)$  and  $B(z)$  are functions with little structure and analytic at  $z = \mu$ . We are primarily interested in the case in which  $f(z)$  has a branch point at  $z = \mu$  and, accordingly,  $\gamma$  is not necessarily an integer number. We can define now a new function  $F(z)$  which, near  $z = \mu$ , behaves as [44]

$$F(z) = \frac{d}{dz} \ln f(z) \approx \frac{\gamma}{(\mu - z)}. \quad (13)$$

Even though  $\gamma$  does not have to be an integer,  $F(z)$  has a simple pole and its residue is the exponent of the cut of  $f(z)$ . Thus, with  $\bar{P}_M^N(z)$  being the PA constructed to  $F(z)$  defined above, the DLog $_M^N(z)$ , or simply  $D_M^N(z)$ , of  $f(z)$  is given by the expression below [44,47,48],

$$\text{Dlog}_M^N(z) \equiv D_M^N(z) = f(0) \exp \left[ \int dz \bar{P}_M^N(z) \right]. \quad (14)$$

Because of the derivative in Eq. (13), the  $f(0)$  term is lost and must be reintroduced to correctly normalize the D-Log. The  $D_M^N$  then reproduces exactly the first  $M + N + 2$  coefficients of  $f(z)$  (one order more than the usual  $P_M^N$ ) and can be used to predict the  $(M + N + 3)$ th coefficient and higher.

In principle, this type of approximant offers a way to determine the branch point and the exponent of the cut of the original function  $f(z)$  from the study of the PA to  $F(z)$  around its pole. Since no assumption about  $\mu$  or  $\gamma$  is made, their estimates are exclusively obtained from the series coefficients.

With these methods, the convergence of sequences of both PAs and D-Logs to the original function is guaranteed for specific types of functions, for example, Stieltjes functions. A Stieltjes function is a function that can be represented by a Stieltjes integral

$$f(z) = \int_0^\infty \frac{d\phi(u)}{1+zu}, \quad (15)$$

where  $\phi(u)$  is a bounded nondecreasing function on the interval  $0 \leq u < \infty$  with finite positive moments given by

$$f_n = \int_0^\infty u^n d\phi(u), \quad (16)$$

for  $n = 0, 1, 2, \dots$ . A necessary and sufficient condition [44] for a function to be Stieltjes is that all the determinants  $\mathcal{D}_{m,n}^{(f)}$  given by

$$\mathcal{D}_{m,n}^{(f)} = \begin{vmatrix} f_m & f_{m+1} & \cdots & f_{m+n} \\ f_{m+1} & f_{m+2} & \cdots & f_{m+n+1} \\ \vdots & \vdots & & \vdots \\ f_{m+n} & f_{m+n+1} & \cdots & f_{m+2n} \end{vmatrix} \quad (17)$$

are positive for any  $m \geq 0$  and  $n \geq 0$ . These determinants produce constraints between the Taylor series coefficients  $f_n$  of the Stieltjes functions. They will also prevent the appearance of “defects,” or technically Froissart doublets, exact cancellations between the zeros of the numerator and the denominator [44], effectively reducing the order of the approximants. These will be essential in our analysis since a PA or D-Log constructed from a Stieltjes function has the same Stieltjes properties as the original function has [43,44]. The D-Log in Eq. (14) is, in general, not a rational approximant, but the function  $F(z)$  is meromorphic and, because of that, it is easily approximated by the PA  $\bar{P}_M^N$ , and the convergence of the approximation can be proven [43]. What is more, for some Stieltjes functions  $f(z)$ , it can be shown that  $F(z)$  will also be Stieltjes. In these cases,

convergence theorems for Stieltjes functions apply to the approximation of  $F(z)$  by  $\bar{P}_M^N$  as well. The exception to the convergence rule is the  $P_1^N$  sequence, as one can prove that approximants belonging to this sequence are not necessarily Stieltjes even when built to approximate a Stieltjes function [44].

It is possible to prove [41,57] that  $\Delta\alpha_{\text{had}}$ , which is related to the inclusive hadronic electroproduction cross section, is a Stieltjes function defined in the region  $-\infty < t \leq 0$ . Thus, Padé theory [43,44] assures us that the poles of approximants of the type  $P_M^{M+k}$  with  $k \geq -1$ , as well as the branch point of D-Logs of the type  $D_N^N$ ,  $D_N^{N+1}$ , and  $D_{N+1}^N$  to  $\Delta\alpha_{\text{had}}$ , are always real and positive with positive residues. Furthermore, since  $\Delta\alpha_{\text{had}}$  scales as  $\mathcal{O}(t^1)$  for small  $t$ , the PA sequences  $P_N^N$  and both  $P_{N+1}^{N+1}$  and  $P_{N+1}^N$  bound the original function. In our case, the fastest convergence is obtained with the  $P_{N+1}^{N+1}$ , then the convergence theorem for Stieltjes functions reads

$$P_1^1(t) \leq P_2^2(t) \leq \dots \leq \Delta\alpha_{\text{had}} \leq \dots \leq P_2^3(t) \leq P_1^2(t). \quad (18)$$

For the D-Log case, a similar pattern is found, in our case. The sequence  $D_N^N$  together with  $D_{N+1}^N$  and  $D_N^{N+1}$  bound the original function. We found the fastest convergence with the  $D_{N+1}^N$  sequence, following the pattern

$$D_1^1(t) \leq D_2^2(t) \leq \dots \leq \Delta\alpha_{\text{had}} \leq \dots \leq D_3^2(t) \leq D_2^1(t). \quad (19)$$

In addition to being model-independent, this method takes advantage of the convergence theorems of Padé theory. To construct the functions that will be used to fit the MUonE toy data in a form that makes contact with the convergence theorems, we will first compute the PAs and D-Logs to the (unknown) Taylor series of  $\Delta\alpha_{\text{had}}(t)$ , i.e., we will build canonical approximants to

$$\Delta\alpha_{\text{had}}(t) = a_1 t + a_2 t^2 + a_3 t^3 + \dots, \quad (20)$$

where the coefficients  $a_n$  are unknown. Then, after a change of variable from  $t$  to  $x$  using Eq. (9), we finally get our fitting functions as a function of  $x$ . With this technique, we can use our knowledge of Stieltjes functions in full to analyze the fit quality and provide constraints to the fit parameters.

In particular, from Carleman’s condition [44,58], the Taylor coefficients of a Stieltjes series cannot change sign and, in our case, they have to be negative. Moreover, the determinant condition imposes the following hierarchy for the coefficients:

$$0 > a_i > a_{i+1}, \quad i \in \mathbb{N}. \quad (21)$$

All these constraints will be used in the fit procedure and it is important to mention that they are model independent, relying only on  $\Delta\alpha_{\text{had}}(t)$  being a Stieltjes function.

Finally, we mention that, in our case, since  $\Delta\alpha_{\text{had}}(t)$  has  $a_0 = 0$ , the  $P_M^N(z)$  will match  $N + M$  coefficients instead of  $N + M + 1$ , and the  $D_M^N(z)$  matches the first  $N + M$  coefficients, instead of the first  $N + M + 2$ .

#### IV. A MODEL FOR THE EUCLIDEAN CORRELATOR

To test our method, we need to generate toy datasets based on a sufficiently realistic model. We will use the phenomenological model for the function  $\text{Im}\Pi_{\text{had}}(s)$  inspired by chiral perturbation theory and perturbative QCD introduced by Greynat and de Rafael in Ref. [40] to obtain  $\Delta\alpha_{\text{had}}$  through the dispersion relation in Eqs. (3) and (8). The model is given by

$$\begin{aligned} \text{Im}\Pi_{\text{had}}(s) &= \frac{1}{4\pi} \left(1 - \frac{4m_\pi^2}{s}\right)^{3/2} \\ &\times \left( \frac{|F(s)|^2}{12} + \sum_f Q_f^2 \Theta(s, s_c, \Delta) \right) \theta(s - 4m_\pi^2), \end{aligned} \quad (22)$$

with  $Q_f$  being the electric charge of the quark of flavor  $f$  and  $\theta(x)$  is the Heaviside  $\theta$  function. The function  $|F(s)|^2$  is the pion vector form factor, which is modeled simply by the  $\rho(770)$  contribution as

$$|F(s)|^2 = \frac{m_\rho^4}{(m_\rho^2 - s)^2 + m_\rho^2 \Gamma(s)^2}, \quad (23)$$

where the running width is

$$\begin{aligned} \Gamma(s) &= \frac{m_\rho s}{96\pi f_\pi^2} \left[ \left(1 - \frac{4m_\pi^2}{s}\right)^{3/2} \theta(s - 4m_\pi^2) \right. \\ &\quad \left. + \frac{1}{2} \left(1 - \frac{4m_K^2}{s}\right)^{3/2} \theta(s - 4m_K^2) \right], \end{aligned} \quad (24)$$

with  $f_\pi = 93.3$  MeV being the pion decay constant and  $m_{\pi/K/\rho}$  the corresponding meson masses. The function  $\Theta(s, s_c, \Delta)$  is defined as

$$\Theta(s, s_c, \Delta) = \left[ \frac{\arctan\left(\frac{s-s_c}{\Delta}\right) - \arctan\left(\frac{4m_\pi^2-s_c}{\Delta}\right)}{\frac{\pi}{2} - \arctan\left(\frac{4m_\pi^2-s_c}{\Delta}\right)} \right]. \quad (25)$$

The parameters  $s_c$  and  $\Delta$  will assume the same values employed in Ref. [40]:  $s_c = 1$  GeV<sup>2</sup> and  $\Delta = 0.5$  GeV<sup>2</sup>. We show in Fig. 1 the line shape of the integrand of Eq. (7) obtained from the use of this model where the gray band represents the experimentally accessible region with the designed MUonE experiment [35–38,49–51].

We consider this simple model to be sufficiently realistic because it leads to a representation of  $\Delta\alpha_{\text{had}}$  which is a

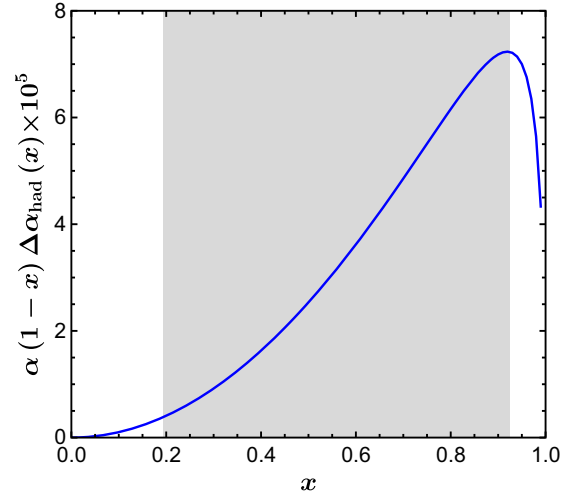


FIG. 1. The integrand to obtain  $a_\mu^{\text{HVP,LO}}$ , given in Eq. (7), calculated from the model of Greynat and de Rafael of Eq. (22). The gray area is the expected region that the MUonE experiment will cover.

Stieltjes function, as expected in QCD. Hence, the theorems given in Sec. III and the coefficient constraints of Eq. (21) are all valid. We also note that numerical checks of several hundred determinants of Eq. (17) built from the series of  $\frac{d}{dt} \ln \Delta\alpha_{\text{had}}(t)$  indicate that this function, which is required to construct the D-Logs, is also Stieltjes.

Computing the integral of Eq. (7) using the model in Eq. (22), we determine the value of  $a_\mu^{\text{HVP,LO}}$  from this model as

$$a_{\mu,\text{model}}^{\text{HVP,LO}} = 6992.4 \times 10^{-11}. \quad (26)$$

This result will serve as a guide for us to compare the values of  $a_\mu^{\text{HVP,LO}}$  obtained from our PAs and D-Log approximants.

#### V. APPROXIMANTS TO THE TAYLOR SERIES

In this section, we will canonically build PAs and D-Logs by matching each approximant's Taylor series to that of  $\Delta\alpha_{\text{had}}(t)$ , Eq. (20), which will be computed from the model of Eq. (22). In this case, the convergence theorems to Stieltjes functions apply. Thus, the convergence to the  $a_\mu^{\text{HVP,LO}}$  value of the model (26) is guaranteed and it should respect the pattern determined in Eqs. (18) and (19).

The Taylor series of  $\Delta\alpha_{\text{had}}$  as a function of  $t$  according to the model of Sec. IV is

$$\begin{aligned} \alpha\Delta\alpha_{\text{had}}(t) \times 10^5 &= -918t - 1752t^2 - 6066t^3 \\ &\quad - 31589t^4 - 214058t^5 + \mathcal{O}(t^6). \end{aligned} \quad (27)$$

##### A. Convergence of PA results

Results from the sequences  $P_N^N(t)$  and  $P_N^{N+1}(t)$  are shown in Fig. 2(a), in the range  $0 \leq x \leq 0.997$ , which corresponds to  $-4 \leq t \leq 0$  GeV<sup>2</sup>. The value from the

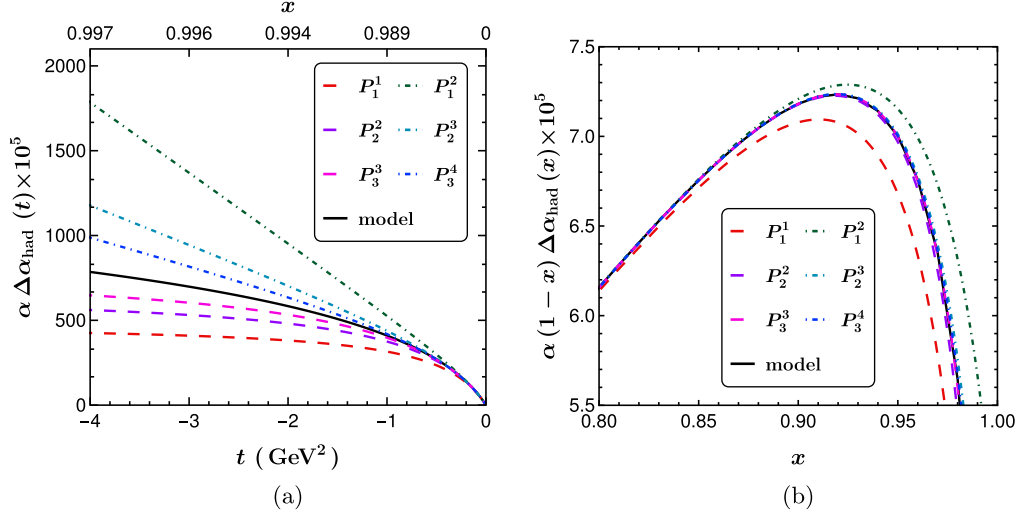


FIG. 2. PAs  $P_N^N(t)$  (dashed line) and  $P_N^{N+1}(t)$  (dot-dashed line) built from the Taylor series of  $\Delta\alpha_{\text{had}}(t)$ , Eq. (27), together with the model results (black line). (a)  $\alpha\Delta\alpha_{\text{had}}(t)$  and (b)  $\alpha(1-x)\Delta\alpha_{\text{had}}(x)$ , used to compute  $a_\mu^{\text{HVP,LO}}$  with Eq. (7).

model for  $\Delta\alpha_{\text{had}}$  is represented by the black line in Fig. 2(a). One can notice that both sequences approach consistently the true value of the model when the order of the PAs is increased and they also bound the model value as expected by the convergence theorem of Eq. (18): the PAs of the sequence  $P_N^N$  approach the function from below, while the ones belonging to the sequence  $P_N^{N+1}$  do so from above. Additionally, it is possible to observe that, for large values of  $|t|$ , the discrepancy between the approximants and the model is much more noticeable. Since the PAs are constructed from the Taylor series coefficients of  $\Delta\alpha_{\text{had}}(t)$ , i.e., in the region of  $t$  close to zero, it is to be expected that far from this region the PAs start to deviate from the true model value. Notice, however, that the PA convergence is dramatically faster compared with that of the Taylor expansion, which breaks off outside a radius of convergence around  $|t| = 4m_\pi^2 \sim 0.1 \text{ GeV}^2$ . The deviation of our approximants at large  $t$  is mitigated in the case of the integrand that appears in the calculation of  $a_\mu^{\text{HVP,LO}}$ , which can be seen in Fig. 2(b) as a function of  $x$ . Since the integrand goes to zero as  $x$  goes to 1, differences in the deep Euclidean region are suppressed, although still noticeable for lower-order PAs. One should also note that the change of variables from  $t$  to  $x$  maps the entire infinite interval  $t \in (-\infty, -4 \text{ GeV}^2]$  into the small interval  $x \in [0.997, 1]$ .

We know from Padé theory [44–46] that the approximants can mimic branch cuts by accumulating interleaved poles and zeros along the cut. For higher-order PAs the mimicking of the cut originated by the term  $(1 - 4m_\pi^2/t)^{3/2}$  in the model of Eq. (22), and whose branch point is located at  $t = 0.078 \text{ GeV}^2$ , was observed. Let us take  $P_5^6(t)$  as an example: this PA has poles at 0.089, 0.116, 0.187, 0.426, and 1.192  $\text{GeV}^2$ , which are interleaved with zeros at 0.089, 0.118, 0.198, 0.620, and 7.951  $\text{GeV}^2$ .

After building the PAs, we can turn to the estimate of the value for  $a_\mu^{\text{HVP,LO}}$ . The results can be seen in Fig. 3, where the black solid line represents the model’s value given in Eq. (26). Again, the pattern stated by the theorem in Eq. (18) is obeyed, with the PA sequence  $P_N^N(t)$  (illustrated by the dashed red line) reaching the model result from below and sequence  $P_N^{N+1}(t)$  (dashed green line) approaching from above. The convergence to the true value, which is guaranteed by theorems, can also be observed.

## B. Convergence of D-Logs and comparison with PA results

We can also build D-Logs to the Taylor series of  $\Delta\alpha_{\text{had}}(t)$ . As stated by the convergence theorems for

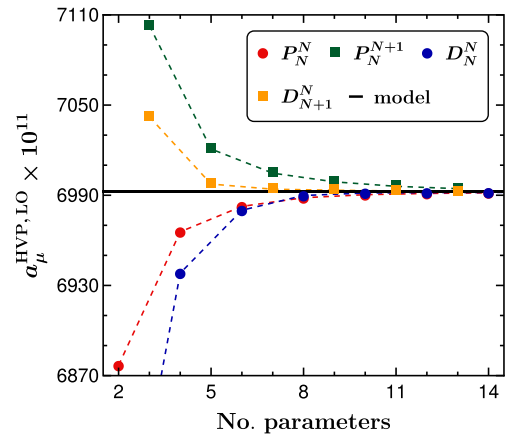


FIG. 3. Comparison between PA and D-Log estimates of  $a_\mu^{\text{HVP,LO}}$  and the value predicted by the model given in Eq. (26) (black solid line). “No. parameters” refers to the value of  $N + M$  used for each approximant.

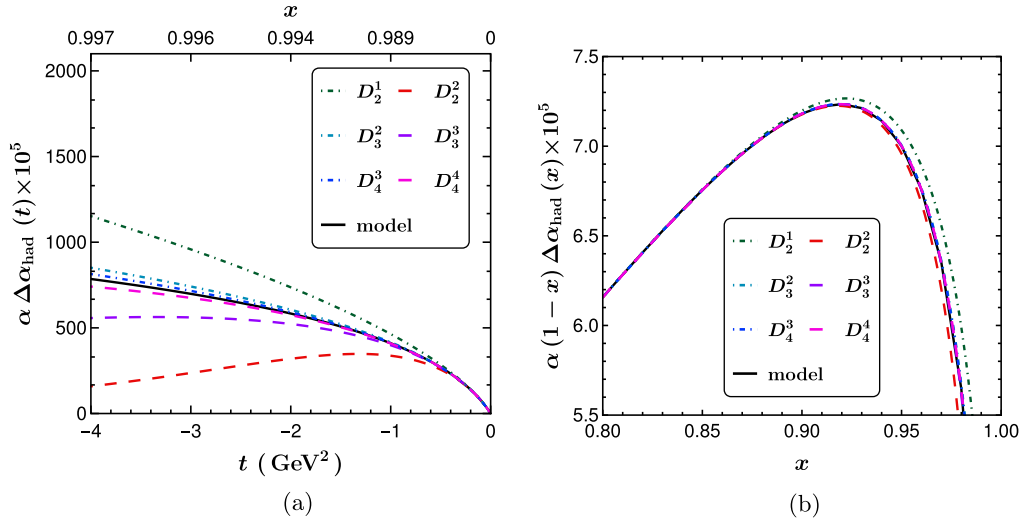


FIG. 4. D-Logs  $D_N^N(t)$  (dashed line) and  $D_{N+1}^N(t)$  (dot-dashed line) built from the Taylor series of  $\Delta\alpha_{\text{had}}(t)$ , Eq. (27), together with the model results (black line). (a)  $\alpha\Delta\alpha_{\text{had}}(t)$  and (b)  $\alpha(1-x)\Delta\alpha_{\text{had}}(x)$ . Plots in the same scale as Figs. 2(a) and 2(b).

Stieltjes functions, specific sequences of D-Logs to  $\Delta\alpha_{\text{had}}(t)$  bound the function from above and from below, as can be seen in Fig. 4, in complete analogy to the convergence pattern of PAs shown in Figs. 2(a) and 2(b).

Compared to PAs, D-Logs contain built-in branch cuts. For example,  $D_5^5(t)$  has four branch cuts and one exponential factor. The onset of the first branch cut for each D-Log approaches the two-pion production threshold of the model, and this is done hierarchically: the D-Log  $D_3^3(t)$  predicts the production threshold to be at  $0.095 \text{ GeV}^2$  and the D-Log  $D_9^9(t)$  at  $0.0798 \text{ GeV}^2$ . With this value, we can predict the charged pion mass to be  $0.141 \text{ GeV}$ , to compare with  $0.140 \text{ GeV}$  used in the model of Eq. (22). We also noticed all D-Logs have a branch cut that tends to be a square root with the singularity around  $0.5 \text{ GeV}^2$ . This branch cut can be interpreted as a signal of the  $\rho$  meson mass and an attempt to replicate the Breit-Wigner distribution used by the model in Eq. (23).

Finally, in Fig. 3 we compare the performance of both PAs and D-Logs provided the same amount of coefficients of the Taylor expansion are used. Subdiagonal  $D_{N+1}^N$  performs better than  $P_N^{N+1}$  when approaching from above for all orders, while diagonal PAs do better than diagonal D-Logs when approaching from below. Notice, however, that the latter comparison holds up only to the number of parameters up to 6, above which the diagonal D-Logs surpass PAs as they converge faster. Not only Cauchy convergence is observed in both sequences but absolute convergence as well. Both methods are useful when bounding the function. For a large number of parameters, D-Logs would converge faster, but for lower orders, none of the methods are systematically superior. We shall keep both methods for our fitting function study.

## VI. APPROXIMANTS TO DATA WITH NO ERROR

Let us consider now an idealized case, where the data points represent exactly what is expected from the model of Eq. (22), with zero error. It is convenient to work with data for  $\alpha\Delta\alpha_{\text{had}}(x) \times 10^5$  such that the Taylor coefficients  $a_n$  are of a natural size, which makes the fits simpler. The values of  $x$  were calculated as follows: the interval  $0.2 \leq x \leq 0.93$  was divided into 30 equally spaced bins and we take the center of each bin  $x_i$  without any error, as the representative values of  $x$  where we will generate the data points to be fitted with our approximants.

To obtain our fitting functions, both PAs and D-Logs will be written in terms of the Taylor series coefficients of  $\Delta\alpha_{\text{had}}(t)$  and later converted to the  $x$  variable by Eq. (9). These functions will then be used to perform the fits to the toy datasets where the fit parameters can be written in terms of the Taylor coefficients of  $\Delta\alpha_{\text{had}}(t)$ . PAs used as fitting functions were already applied in similar contexts and the convergence theorems are apparently satisfied in these cases [18,59–63]. It is important to emphasize, however, that the theorems presented in Sec. III are demonstrated in the case of canonical PAs, i.e., those built to the Taylor series coefficients. Strictly speaking, in the case of a fit to data points in a given interval, we conscientiously slightly depart from the conditions of the theorem as a fit to data can be interpreted as imposing one matching condition for each datum instead of  $N+M$  conditions at the same point. Therefore, we expect convergence theorems to be, as we will see, satisfied in this case as well, but the convergence velocity may be smaller.

To show in detail a concrete case of our procedure, we start by examining the PA  $P_1^1(t)$ . To construct the fitting function, we first compute  $P_1^1(t)$  as a function of the

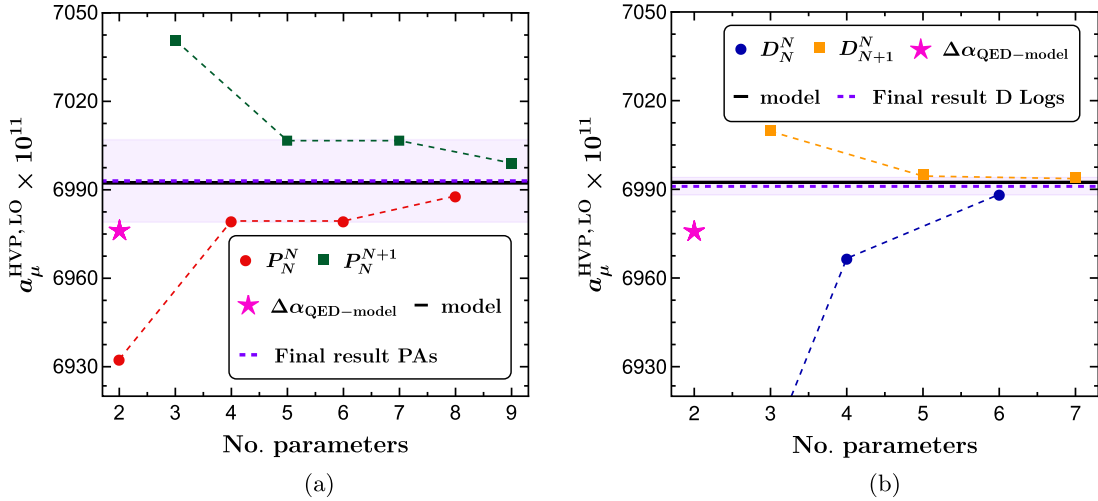


FIG. 5.  $a_\mu^{\text{HVP,LO}}$  from (a) PAs and (b) D-Logs fitted to a dataset with zero error. The pink star is the value obtained from the model of Eq. (34) and the purple dashed line is the final result coming from the (a) PAs and (b) D-Logs with the light purple band representing the systematic error of the result. The black line indicates the true value of the model given in Eq. (26).

unknown Taylor series coefficients  $a_n$  of  $\Delta\alpha_{\text{had}}(t)$  given by Eq. (20). We Taylor expand the PA and  $\Delta\alpha_{\text{had}}(t)$  and match the coefficients of both order by order. We then perform the change of variables of Eq. (9) which leads to the fitting function in terms of  $x$ ,<sup>1</sup>

$$\begin{aligned} P_1^1(t(x)) \equiv P_1^1(x) &= -\frac{a_1^2 m_\mu^2 x^2}{a_1 - a_1 x + a_2 m_\mu^2 x^2} \\ &= -\frac{b_1 m_\mu^2 x^2}{1 - x + b_2 m_\mu^2 x^2}, \end{aligned} \quad (28)$$

where the fit parameters are  $b_1 = a_1$  and  $b_2 = a_2/a_1$ . Because of the hierarchy of the Taylor coefficients of  $\Delta\alpha_{\text{had}}(t)$ , given in Eq. (21), we can conclude that  $b_1$  and  $b_2$  have to obey the following relations:  $b_1 < 0$  and  $b_2 > 1$ , which will be imposed in the fit.

To obtain the parameters of the PAs and D-Logs from these fits to these zero-error data, we will perform the minimization of the fit quality  $\mathcal{Q}^2$  given simply by

$$\mathcal{Q}^2 = \sum_{i=1}^{30} [\alpha \Delta\alpha_{\text{had}}(x_i) \times 10^5 - \mathcal{P}_M^N(x_i)]^2, \quad (29)$$

where  $\mathcal{P}_M^N(x)$  refers, generically, to a PA  $P_M^N(x)$  or to a D-Log  $D_M^N(x)$ . In reasonable fits, we expect very small values for  $\mathcal{Q}^2$ . For  $P_1^1(x)$ , after the minimization we get  $\mathcal{Q}^2 = 1.18 \times 10^{-3}$  and the values of the fit parameters  $b_1$  and  $b_2$  lead to the following Taylor series coefficients:

<sup>1</sup>We will use the same notation  $P_M^N(x)$  to refer to the fitting functions derived from the PA  $P_M^N(t)$  even though they are not a PA of order  $N + M$  in the variable  $x$ .

$$a_1 = -912 \text{ GeV}^{-2} \quad \text{and} \quad a_2 = -1489 \text{ GeV}^{-4}, \quad (30)$$

which differ from the true values by only 0.7% and 15%, respectively. Furthermore, the pole of this PA in  $t$  is at  $t_{\text{pole}} = 0.61 \text{ GeV}^2$ . As stated by the convergence theorems, the pole is located on the positive real axis of  $t$ .

The PA can then be used to estimate the value of  $a_\mu^{\text{HVP,LO}}$ . By employing the resulting  $P_1^1(x)$  in Eq. (7) we get

$$a_{\mu, P_1^1}^{\text{HVP,LO}} = 6933 \times 10^{-11}, \quad (31)$$

with an error of 0.9% compared to Eq. (26), the expected value within our model.

The same analysis can be made for the other PAs of the sequences  $P_N^N(t)$  and  $P_{N+1}^N(t)$ .<sup>2</sup> We will employ only these two sequences since they are expected to bound the real function in the theorem presented in Sec. III. We will build several approximants from each sequence, in order to study the convergence pattern, but the final estimate of  $a_\mu^{\text{HVP,LO}}$  will not include all results, as we detail below. Figure 5(a) shows the values predicted by the PAs of the sequences  $P_N^N$  and  $P_{N+1}^N$  for  $a_\mu^{\text{HVP,LO}}$ . It is possible to notice that the pattern of convergence of Eq. (18), where the  $P_N^N$  bound the true value from below and the ones of  $P_{N+1}^N$  from above, is satisfied in this case. One can note in Fig. 5(a) that in some cases when the order of the PA is increased there is no real improvement in the estimate of  $a_\mu^{\text{HVP,LO}}$ . This can be explained by the appearance of a defect, which consists of a pole partially canceled by a nearby zero, effectively reducing the order of the PA. It is known from PA

<sup>2</sup>The fitting functions obtained from these approximants are given explicitly in the Appendix.

theory [43,44] that approximants applied canonically to Stieltjes functions, i.e., to their Taylor series, cannot have Froissart doublets. Nonetheless, the theorem does not prevent the PA from having numerical defects, which are almost exact cancellations between poles and zeros, arising from the fitted parameters. Adding more coefficients into a fitting function also introduces larger correlations among them. As such, fitted coefficients will have errors and within their allowed fitted regions cancellations may emerge. We note, however, that the important fact that the two sequences of PAs bound the true value remains true.

A final estimate for  $a_{\mu}^{\text{HVP,LO}}$  from the PAs can then be obtained. For that, we will limit the analysis to the first two approximants in each sequence, i.e.,  $P_1^1, P_2^2, P_1^2$ , and  $P_2^3$ . This choice is motivated by the fact that we do not expect to be able to fit much more than five parameters in more realistic scenarios, with a limited number of data points and with the errors expected by the MUonE experiment. Computing the central value of the prediction as the mean between the four approximants we get  $a_{\mu, \text{PAS}}^{\text{HVP,LO}} = 6990 \times 10^{-11}$ , which differs from the expected value given in Eq. (26) by 0.04%. This can be improved if we take advantage of the expected convergence of the PA sequences dictated by Eq. (18). It is expected that the higher-order PAs will be closer to the true value. We can then take the average of the highest-order approximants of each sequence only, i.e.,  $P_2^2$  and  $P_3^3$ , since from the convergence behavior of Eq. (18), we expect the true value to lie between these two PAs estimates. Thus, computing this mean we get  $a_{\mu, \text{PAS}}^{\text{HVP,LO}} = 6993 \times 10^{-11}$ , which presents an error of less than 0.01% with respect to the model value. This result can be seen in Fig. 5(a) as the purple dashed line.

One can also estimate a theoretical uncertainty due to the truncation of the PA sequences. Taking the convergence pattern of Eq. (18) into consideration, the error will be defined as half of the distance between the two highest-order PAs ( $P_2^2$  and  $P_3^3$  in this case), which gives us a relative error of 0.19% represented in Fig. 5(a) by the light purple band. This systematic error has to be added to our final numbers in the spirit of a conservative estimate.

A similar analysis can be performed with D-Logs. As argued in the previous section, they introduce branch cuts, thus speeding the convergence, a key factor when dealing with a small number of parameters. In this case, we will use the approximants of the sequences  $D_N^N$  and  $D_{N+1}^N$ , since these are the sequences expected to bound the original function, as indicated in Eq. (19). To construct the fitting

functions, we start with an arbitrary Taylor series in the  $t$  variable up to order  $N + M$ , compute its logarithm, and then the derivative with respect to  $t$ , as indicated in Eq. (13). After that, we construct a PA,  $\bar{P}_M^N$ , to the thus obtained series and, employing Eq. (14), we obtain the D-Log expression in terms of  $t$ . We then perform the change of variables of Eq. (9) and write the generic fitting function in terms of  $x$ . The simplest D-Log, for example, is  $D_2^1(t)$  that before and after the change of variables reads

$$D_2^1(t) = \frac{-f_0 t}{(r_1 - t)^{\gamma_1}} \rightarrow D_2^1(x) = \frac{f_0 m_{\mu}^2 x^2 (1-x)^{-1+\gamma_1}}{(r_1 - r_1 x + m_{\mu}^2 x^2)^{\gamma_1}}, \quad (32)$$

with fit parameters  $f_0$ ,  $r_1$ , and  $\gamma_1$ . We recall that  $D_M^N$  requires  $N + M$  fit parameters. The next approximant is  $D_2^2(t)$ , which is given by

$$D_2^2(t) = \frac{-f_0 t e^{\beta t}}{(r_1 - t)^{\gamma_1}} \rightarrow D_2^2(x) = \frac{f_0 m_{\mu}^2 x^2 (1-x)^{-1+\gamma_1}}{(r_1 - r_1 x + m_{\mu}^2 x^2)^{\gamma_1}} e^{\beta \frac{m_{\mu}^2 x^2}{(1-x)}}, \quad (33)$$

with fit parameters  $f_0$ ,  $r_1$ ,  $\gamma_1$ , and  $\beta$ . All D-Logs used are provided in Table III in the Appendix.

A comparison of the results for  $a_{\mu}^{\text{HVP,LO}}$  obtained from the D-Log approximants is found in Fig. 5(b) where we show both sequences  $D_N^N$  and  $D_{N+1}^N$  as a function of the number  $N + M$  of fitted coefficients. Convergence for D-Logs is faster than PAs, especially for the subdiagonal sequence. In line with the criteria used for PAs, we employ  $D_3^2$  and  $D_3^3$  to obtain a final prediction for  $a_{\mu}^{\text{HVP,LO}}$  from the D-Logs. Computing the mean of these values, we obtain  $a_{\mu, \text{DLogs}}^{\text{HVP,LO}} = 6991.4 \times 10^{-11}$ , which exhibits an error of less than 0.02% compared to the model value and a theoretical uncertainty band of less than 0.05% [light purple band in Fig. 5(b)].

An important outcome of analyzing data with no errors is to provide an estimate of the theoretical uncertainty for other models proposed in the literature to fit and extrapolate the MUonE data. Let us take a closer look at the model proposed by Abbiendi in the letter of intent of the MUonE experiment [37], motivated by the one-loop QED calculation of the vacuum polarization. This fitting function is

$$\Delta\alpha_{\text{QED-model}}(t) = KM \left[ -\frac{5}{9} - \frac{4M}{3t} + \frac{2 \left( \frac{4M^2}{3t^2} + \frac{M}{3t} - \frac{1}{6} \right)}{\sqrt{1 - \frac{4M}{t}}} \log \left| \frac{1 - \sqrt{1 - \frac{4M}{t}}}{1 + \sqrt{1 - \frac{4M}{t}}} \right| \right], \quad (34)$$

where the parameters  $K$  and  $M$  are determined by the fit.

After the  $Q^2$  minimization, we obtain  $Q^2 = 1.84 \times 10^{-4}$  and the parameters determined by the fit are  $K = 6865.36 \text{ GeV}^{-2}$  and  $M = 0.06 \text{ GeV}^2$ . The  $a_\mu^{\text{HVP,LO}}$  value can be predicted by this model employing the expression above in Eq. (7), which in this scenario gives us  $a_{\mu, \text{QED-model}}^{\text{HVP,LO}} = 6976 \times 10^{-11}$ . Even though this function has only two parameters, as is the case of  $P_1^1(x)$ , this model implements the logarithmic dependency expected at large  $|t|$ , which facilitates a better approximation to the exact result. Thus, it is expected that  $\Delta\alpha_{\text{QED-model}}(t)$  has a more accurate result than  $P_1^1(x)$ , which can be verified in Fig. 5(a) where the pink star represents the prediction of  $\Delta\alpha_{\text{QED-model}}$ . Regarding the D-Logs,  $D_1^1(t)$ , which contains only two parameters, is unable to reproduce any type of singularity, whether poles or cuts. Therefore, one would not expect that its estimate of  $a_\mu^{\text{HVP,LO}}$  be more precise than the one provided by the QED-inspired model, which is corroborated by Fig. 5(b). The next D-Log,  $D_2^1(t)$ , with three parameters, contains a branch cut and achieves a relative error of 0.24% when compared to the true value of our model (26), an error very similar to one obtained from  $\Delta\alpha_{\text{QED-model}}$ , but using our model-independent method. When comparing it, however, to our final estimate from the PAs or D-Logs given in Figs. 5(a) and 5(b), respectively, the result from the use of  $\Delta\alpha_{\text{QED-model}}$  differs significantly more from the true model value of Eq. (26). *A priori* knowledge of the function has been used when defining the QED model in Eq. (34). Even though the logarithmic dependence is captured, the value of the fitted parameter  $M$  departs from the true one,  $m_\pi^2 \sim 0.02 \text{ GeV}^2$ . The accuracy of the result is thus model dependent and the systematic error is difficult to quantify. Whether the QED model in Eq. (34) would perform similarly with real data cannot be answered with guarantees.

As mentioned above, we can estimate the theoretical uncertainty in using  $\Delta\alpha_{\text{QED-model}}$  as the relative difference between the fitted value and the true model value given in Eq. (26). This gives us 0.24%, which is slightly larger than, but of the same order of, the one obtained with our PA

method. The systematic error from the use of  $\Delta\alpha_{\text{QED-model}}$  is, however, significantly larger than that of the D-Log approach. As already stated, these systematic uncertainties have to be considered in the final estimates of  $a_\mu^{\text{HVP,LO}}$  regardless of the fitting function used.

## VII. APPROXIMANTS TO DATA WITH REALISTIC ERRORS

We can now move to more realistic datasets, with fluctuations and uncertainties of the same order as those expected to be obtained in the MUonE experiment. We built 1000 toy datasets for the function  $\alpha\Delta\alpha_{\text{had}}(x) \times 10^5$  employing the model of Eq. (22) with 30 data points each, corresponding to the expected MUonE bin sizes. The 30 values of  $x$  are computed in the same way as in the previous section. The datasets were generated assuming a Gaussian distribution around the value of  $\alpha\Delta\alpha_{\text{had}}(x) \times 10^5$  given by the model of Eq. (22) with an error ranging from 0.7%, for larger values of  $x$ , up to 6.7% for  $x \approx 0.2$ . Additionally, the  $a_\mu^{\text{HVP,LO}}$  of each dataset was calculated: the data were used in the region  $x \in [0.2, 0.93]$ , and the model of Sec. IV, used in the data generation, was employed outside this interval. Calculating the median and the 68% confidence level (CL) of this distribution we get

$$a_\mu^{\text{HVP,LO}} = (6991_{-20}^{+22}) \times 10^{-11} \quad (\text{model value}). \quad (35)$$

Since in this case the extrapolation outside the data region is exact, this value gives an idea of the best possible result that we can expect from our PA or D-Log predictions given the available information in the datasets.

For each dataset, the parameters  $b_n$  of the PAs and  $f_0, \beta, r_n, \gamma_n$  of the D-Logs are determined by a  $\chi^2$  minimization. We apply penalizations if the result for the Taylor series coefficients  $a_n$  of  $\Delta\alpha_{\text{had}}$  do not follow the expected hierarchy given in Eq. (21) or the determinant condition discussed in Sec. III. This is done by employing a modified  $\chi^2$  function given by

$$\begin{aligned} \chi^2(\mathcal{P}_M^N) = & \sum_{i,j=1}^{30} [\alpha\Delta\alpha_{\text{had}}(x_i) \times 10^5 - \mathcal{P}_M^N(x_i)](C^{-1})_{ij}[\alpha\Delta\alpha_{\text{had}}(x_j) \times 10^5 - \mathcal{P}_M^N(x_j)] \\ & + n_{\text{d.o.f.}} \left[ \sum_{i=2}^{N+M} \theta(a_i - a_{i-1}) + \sum_{i=3}^{N+M} \theta(\mathcal{D}_{i-1,1}) + \sum_{i=5}^{N+M} \theta(-\mathcal{D}_{i-2,2}) \right], \end{aligned} \quad (36)$$

where  $C$  is the data covariance matrix,  $n_{\text{d.o.f.}}$  is the number of degrees of freedom,  $\theta(x)$  is the Heaviside  $\theta$  function, and  $\mathcal{D}_{m,n}$  is the determinant given in Eq. (17). The  $\chi^2$  penalties are scaled by  $n_{\text{d.o.f.}}$  to be of a natural size. Hence, if the hierarchy of  $a_i$  coefficients or the determinant

conditions are not satisfied, the  $\chi^2$  has a steep increase that forces the minimization algorithm to search for minima respecting the conditions expected for Stieltjes functions. The arguments inside  $\theta(x)$  are always written in terms of the fit parameters of each approximant.

Alternatively, one could neglect the second line from Eq. (36) and turn the aforementioned penalization into limits for the fit parameters. Both strategies lead to equivalent results.

For each fit, the approximant written as a function of the variable  $t$  is examined for the appearance of defects. As already explained in Sec. VI, there is no guarantee that numerical defects will not occur when approximants are used as fitting functions to real data. In PAs, defects are manifest in a nearly exact cancellation between a pole and a zero of the approximant. Their presence effectively decreases the order of the PA, which spoils the systematic study of the convergence of a given sequence. For the D-Logs, a similar effective reduction in the order of the approximants can happen in several circumstances. If the exponent of a cut  $\gamma_n$  is compatible with zero, this reduces the number of fitted coefficients by 2, for example. Equivalently, if two branch points are equal, i.e.,  $r_i = r_j (i \neq j)$ , the two cuts are merged, again lowering the number of coefficients by 2. In the case of the diagonal D-Logs, the exponential coefficient  $\beta$  may be compatible with zero within errors, reducing, in practice, the approximant by one order (see the Appendix for the explicit D-Log expressions). Finally, if  $\sum_n \gamma_n = 1$ , the pole at  $x = 1$ , resulting from the change of variable described in Eq. (9), is lost.

In summary, those approximants with numerical defects are discarded because they are redundant with lower-order ones.<sup>3</sup> For instance, for the PA  $P_2^2$  approximately 30% of the fits have to be disregarded due to the presence of defects, and for  $P_2^3$  this number is 56%. In the case of D-Logs, 0.7% of fits for  $D_3^2$  and 4% for  $D_3^3$  are not considered.

A related issue appears in fits where one of the parameters turns out to be almost zero without numerical cancellation. In our fits, this happens when using the PA  $P_2^2$  as a fitting function for several of the datasets.<sup>4</sup> This may be problematic since, with one of the parameters in the denominator equal to zero, this approximant is effectively reduced to  $P_1^2$ . To circumvent this issue, in our results for  $P_2^2$ , we impose that none of the parameters are zero, even if this leads to a larger  $\chi^2$ , which guarantees that we do not mix results from different approximants in the final statistical distributions.

After fitting, we use the resulting approximant to calculate  $a_\mu^{\text{HVP,LO}}$  for each of the accepted fits using

<sup>3</sup>When we inspect the approximants for the appearance of these defects, we employ a numerical tolerance of  $10^{-4}$ . This number is somewhat arbitrary, and we have checked that varying it by 1 order of magnitude does not alter the results significantly.

<sup>4</sup>This only happens for  $P_2^2$  and this is not observed when fitting the data without fluctuations, which indicates that this is an artifact related to the statistical errors.

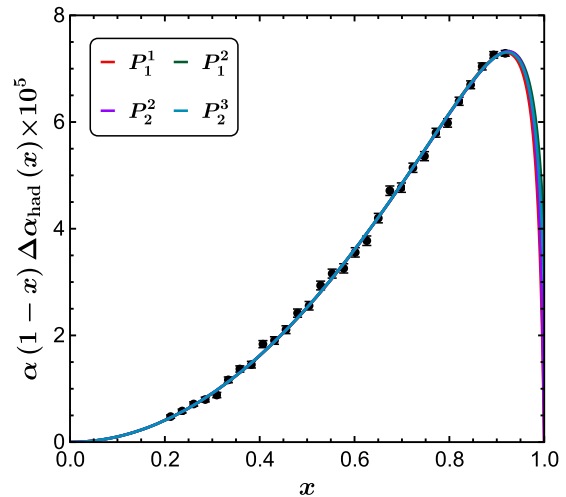


FIG. 6. Fitted PAs to the integrand  $\alpha(1-x)\Delta\alpha_{\text{had}}(x)$  from Eq. (7). Black points show one toy dataset used in the current exercise.

Eq. (7) [or with Eq. (10), for the partial contributions].<sup>5</sup> The results from each  $\mathcal{P}_M^N$  are then obtained as follows: the central value for  $a_\mu^{\text{HVP,LO}}$  is the median of the distribution of the results and the uncertainty is obtained within a 68% CL.

Given the size of the errors present in the datasets, now we are only able to obtain meaningful results from approximants with at most six fit parameters, which corresponds to the PAs  $P_1^1(x)$ ,  $P_1^2(x)$ ,  $P_2^2(x)$ , and  $P_2^3(x)$ , and the D-Logs  $D_2^1(x)$ ,  $D_2^2(x)$ ,  $D_3^2(x)$ , and  $D_3^3(x)$ . Higher-order approximants have an excessive number of parameters to fit which leads to gigantic uncertainties and unstable results. As an example of typical fit outcomes, for one of the 1000 datasets, we compare in Fig. 6 the data and the four PAs obtained after fitting to these data. We can notice that all approximants fit the data very well and no significant deviation between the approximants is present in the region where data are available. Similar results are found for D-Logs. Discrepancies in the region where extrapolation is necessary are, however, non-negligible. As discussed in Secs. II and V, the change of variable from  $t$  to  $x$  masks the fact that there is a significant extrapolation being performed and the integration from  $x \approx 0.93$ , where the data end, to  $x = 1$  still gives an important contribution to  $a_\mu^{\text{HVP,LO}}$ .

Before turning to our final numbers, it is instructive to examine the results from each approximant separately. We give in Table I the values for  $a_\mu^{\text{HVP,LO}}$  obtained after

<sup>5</sup>It is also possible to use the approximants only to extrapolate the data, i.e., use the data points to obtain  $a_\mu^{\text{HVP,LO}}$  in the region  $x \in [0.2, 0.93]$  and then apply the approximants only outside this interval. The results from both procedures are in very good agreement, since the fit quality is almost always excellent, which is confirmed by Fig. 6 and the  $\chi^2$  values of Table I.

TABLE I. Results for  $a_\mu^{\text{HVP,LO}}$  from the PAs and D-Logs used as fitting functions to toy datasets together with the final values for  $\chi^2/n_{\text{d.o.f.}}$  of the respective approximants. Final results for both methods are also presented.

	$a_\mu^{\text{HVP,LO}} \times 10^{11}$	$\chi^2/n_{\text{d.o.f.}}$		$a_\mu^{\text{HVP,LO}} \times 10^{11}$	$\chi^2/n_{\text{d.o.f.}}$
$P_1^1$	$6938 \pm 21$	$1.01^{+0.27}_{-0.25}$	$D_2^1$	$7052^{+66}_{-71}$	$1.01^{+0.26}_{-0.26}$
$P_1^2$	$7042^{+114}_{-104}$	$1.01^{+0.28}_{-0.26}$	$D_2^2$	$6956^{+96}_{-65}$	$1.05^{+0.28}_{-0.27}$
$P_2^2$	$6980^{+46}_{-34}$	$1.05^{+0.29}_{-0.27}$	$D_3^2$	$6999^{+48}_{-39}$	$1.10^{+0.29}_{-0.28}$
$P_2^3$	$6994^{+85}_{-49}$	$1.11^{+0.29}_{-0.31}$	$D_3^3$	$6977^{+72}_{-53}$	$1.14^{+0.30}_{-0.29}$
Final result	$6987^{+46}_{-34}$	...	Final result	$6988^{+48}_{-39}$	...

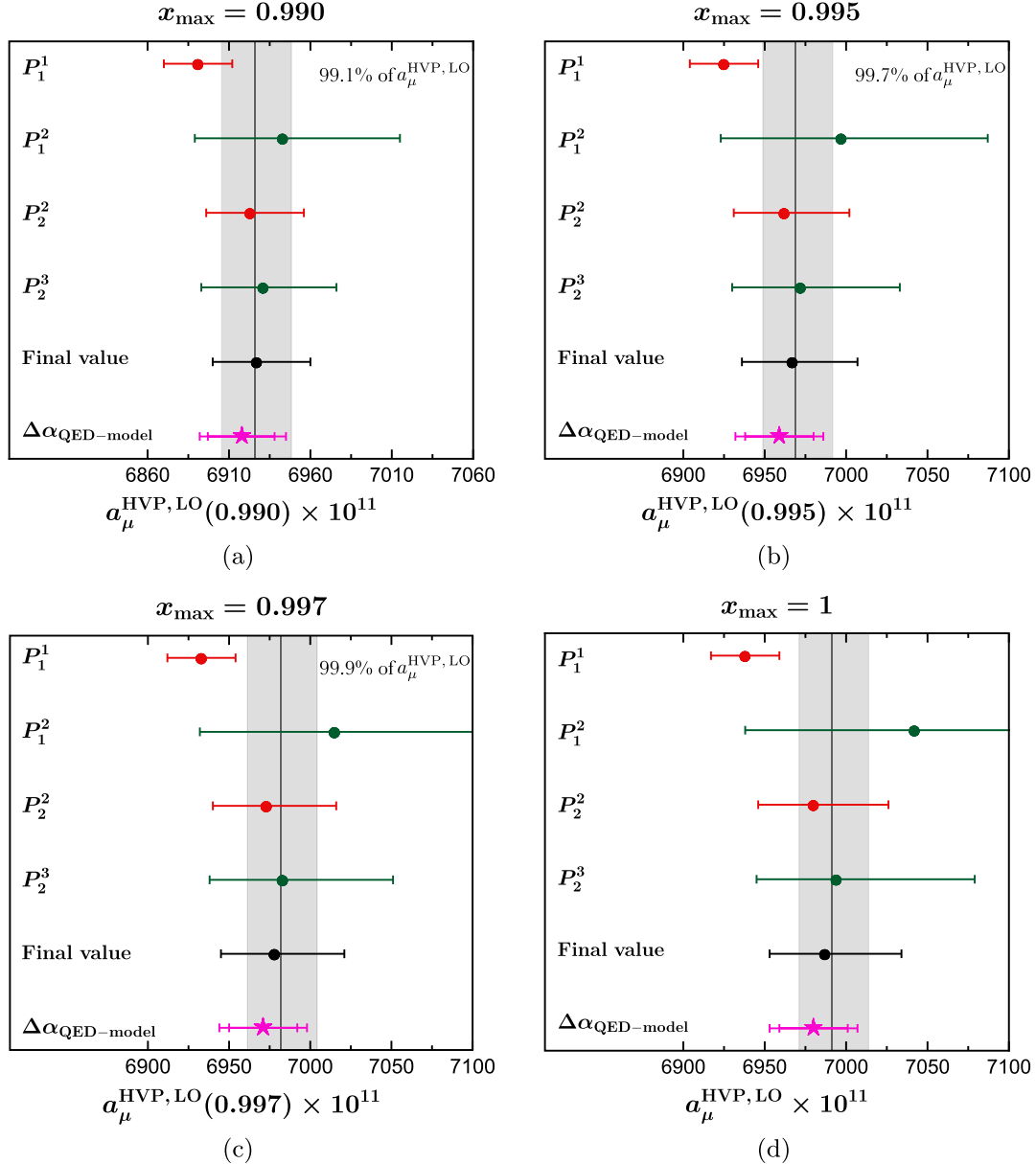


FIG. 7.  $a_\mu^{\text{HVP,LO}}(x_{\text{max}})$  from PAs fitted to the toy datasets for four different values of  $x_{\text{max}}$ : (a)  $x_{\text{max}} = 0.990$  ( $t_{\text{max}} = -1.1 \text{ GeV}^2$ ), (b)  $x_{\text{max}} = 0.995$  ( $t_{\text{max}} = -2.2 \text{ GeV}^2$ ), (c)  $x_{\text{max}} = 0.997$  ( $t_{\text{max}} = -3.7 \text{ GeV}^2$ ), and (d)  $x_{\text{max}} = 1$ . The PAs  $P_N^N$  are shown in red, while  $P_N^{N+1}$  appear in green. Final results obtained from the approximants appear as a black dot and results from the QED-inspired model of Eq. (34) as a pink star. The inner error bar in the QED model result represents the statistical uncertainty. The gray band gives  $a_\mu^{\text{HVP,LO}}(x_{\text{max}})$  with exact extrapolation using the model of Eq. (22).

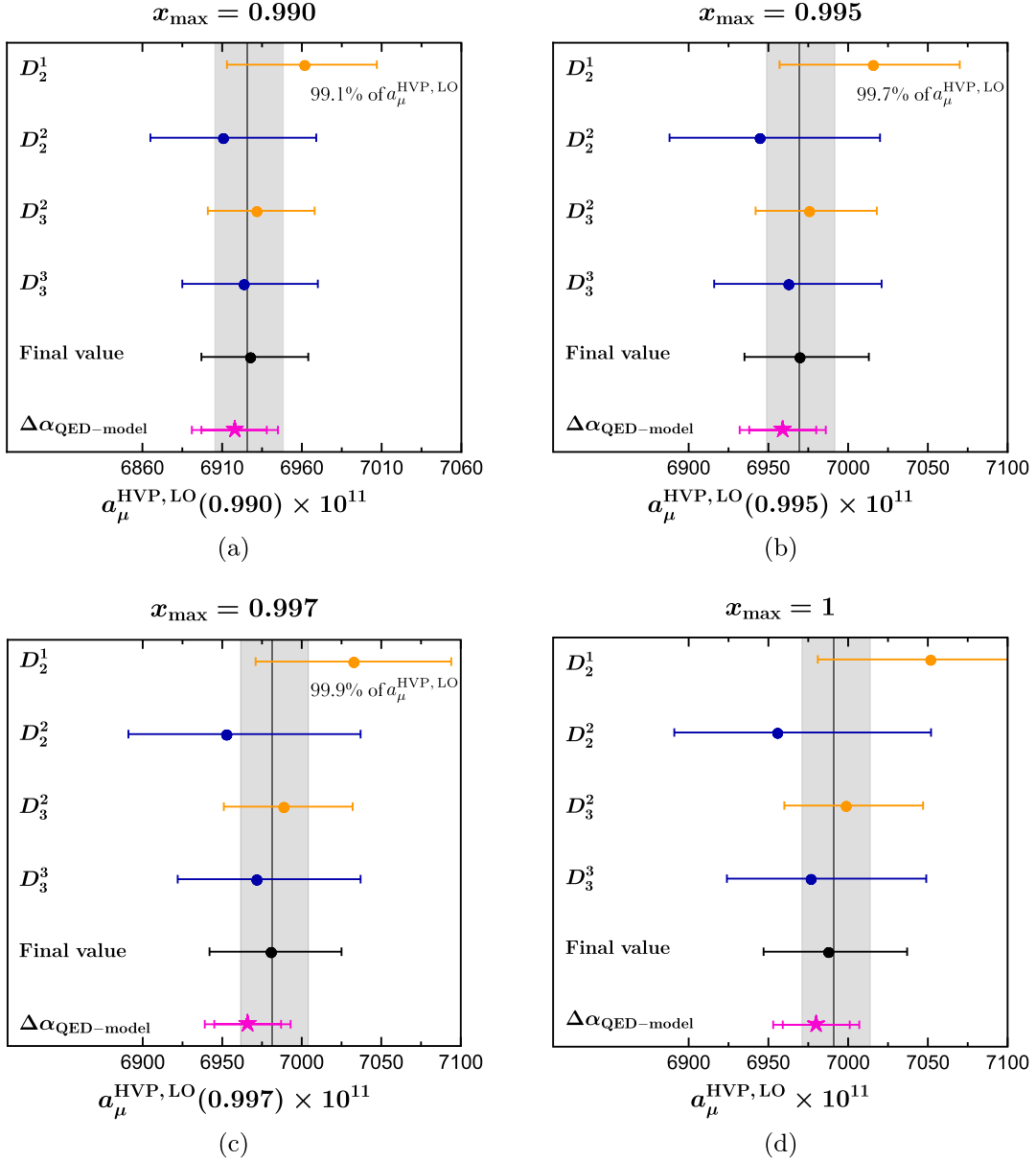


FIG. 8.  $a_\mu^{\text{HVP,LO}}(x_{\text{max}})$  from D-Logs fitted to the toy datasets for four different values of  $x_{\text{max}}$ : (a)  $x_{\text{max}} = 0.990$  ( $t_{\text{max}} = -1.1 \text{ GeV}^2$ ), (b)  $x_{\text{max}} = 0.995$  ( $t_{\text{max}} = -2.2 \text{ GeV}^2$ ), (c)  $x_{\text{max}} = 0.997$  ( $t_{\text{max}} = -3.7 \text{ GeV}^2$ ), and (d)  $x_{\text{max}} = 1$ . The D-Logs  $D_N^N$  are shown in blue, while  $D_{N+1}^N$  appear in yellow. Final results obtained from the approximants appear as a black dot and results from the QED-inspired model of Eq. (34) as a pink star. The inner error bar in the QED model result represents the statistical uncertainty. The gray band gives  $a_\mu^{\text{HVP,LO}}(x_{\text{max}})$  with exact extrapolation using the model of Eq. (22).

fitting each PA and D-Log to the 1000 toy datasets, together with the median of the reduced  $\chi^2$  (and its 68% CL). One can notice that the pattern of convergence of Eqs. (18) and (19) is obeyed for the central values obtained from the PAs and D-Logs, as already happened in the idealized case of Sec. VI. It is also possible to note from Table I that the uncertainty of  $a_\mu^{\text{HVP,LO}}$  from  $P_1^2$  is much larger than the other PAs' predictions. This can be understood by the fact that this approximant is not a Stieltjes function, as discussed in Sec. III—results from  $P_1^2$  can, therefore, be safely

discarded when confronted with the others. Finally, we observe that the uncertainty of  $a_\mu^{\text{HVP,LO}}$  increases from  $P_2^2$  to  $P_3^2$  as well as from  $D_3^2$  to  $D_3^3$  and we could not obtain meaningful fits for  $P_3^3$  and  $D_4^3$ . These are indications that we are at the limit of what can be done with the toy datasets given the size of the errors expected by the MUonE experiment.

The results from Table I show a somewhat large uncertainty. This uncertainty stems mostly from the extrapolation, i.e., deviations of different approximants outside the

TABLE II. Final results for  $a_\mu^{\text{HVP,LO}}(\pm\text{stat})(\pm\text{sys})$  in units of  $10^{11}$  for PAs, D-Logs, and the QED model, where stat and sys refer to the statistical and systematic uncertainties, respectively.

$x_{\text{max}}$	$a_{\mu,\text{PAs}}^{\text{HVP,LO}}$	$a_{\mu,\text{Dlogs}}^{\text{HVP,LO}}$	$a_{\mu,\text{QED-model}}^{\text{HVP,LO}}$	$a_{\mu,\text{datasets}}^{\text{HVP,LO}}$
0.990	$6927^{(+33)}_{(-27)}(\pm 4)$	$6928^{(+36)}_{(-31)}(\pm 4)$	$6918^{(+21)}_{(-20)}(\pm 4)$	$6926^{(+22)}_{(-20)}$
0.995	$6967^{(+40)}_{(-31)}(\pm 5)$	$6970^{(+42)}_{(-34)}(\pm 7)$	$6959(\pm 21)(\pm 17)$	$6969^{(+22)}_{(-20)}$
0.997	$6978^{(+43)}_{(-33)}(\pm 5)$	$6981^{(+43)}_{(-38)}(\pm 9)$	$6971(\pm 21)(\pm 17)$	$6982^{(+22)}_{(-20)}$
1.000	$6987^{(+46)}_{(-34)}(\pm 7)$	$6988^{(+48)}_{(-39)}(\pm 11)$	$6980(\pm 21)(\pm 17)$	$6991^{(+22)}_{(-20)}$

data region, corresponding to  $t \in (-\infty, -0.138]$  GeV<sup>2</sup>. To quantify the error from this extrapolation, we calculated the partial contributions  $a_\mu^{\text{HVP,LO}}(x_{\text{max}})$  up to  $x_{\text{max}}$ , defined in Eq. (10), with  $x_{\text{max}} \in [0.990, 1]$ . It is expected that by restricting the extrapolation to a smaller interval, the final errors will be smaller. For example, with  $x_{\text{max}} = 0.990$  one can cover 99.1% of the value of  $a_\mu^{\text{HVP,LO}}$ , which may be an acceptable trade-off if the uncertainties are significantly reduced. The remaining 0.9% would have to be obtained externally by matching to perturbative QCD or lattice data if available at similar precision, for example.

The  $a_\mu^{\text{HVP,LO}}(x_{\text{max}})$  values estimated for  $x_{\text{max}} = \{0.990, 0.995, 0.997, 1\}$  are illustrated for the PAs in Fig. 7 and for the D-Logs in Fig. 8. The pattern of convergence is evident for all  $x_{\text{max}}$  in Figs. 7 and 8, where the gray band is the expected model value of  $a_\mu^{\text{HVP,LO}}(x_{\text{max}})$  given by Eq. (22). The red points in Fig. 7 represent the  $P_N^N$  sequence and the green ones  $P_N^{N+1}$ . In Fig. 8, the blue points show the  $D_N^N$  sequence, while results for  $D_{N+1}^N$  appear in yellow. One can observe in the results of these figures that the uncertainty steadily increases with  $x_{\text{max}}$ , which reflects the dispersion due to extrapolation outside the data region.

We can then obtain a final value for  $a_\mu^{\text{HVP,LO}}(x_{\text{max}})$ . Since with both PA and D-Logs the approximants from the two sequences are expected to bound the true value, it is natural to use the average of the highest-order approximants, in this case  $P_2^2$  and  $P_3^3$ , on the one hand, and  $D_3^3$  and  $D_3^3$  on the other, as the central value for the final estimate. We will consider the final statistical uncertainty to be the smallest between each pair of approximants. This can be considered conservative since we do not reduce the final error, as would be done in a weighted average, due to the expected strong correlations between the two results. It is important to mention that we book the error stemming from the extrapolation—which is the dominant source of error—as part of the statistical uncertainty since, ultimately, this error is rooted in the statistical fluctuations of the data. Finally, for the systematic uncertainty, we use half the interval spanned by the central values from the two highest-order approximants. Our final estimates for PAs and D-Logs for the different values of  $x_{\text{max}}$  are collected in Table II—second and third columns, respectively—where “stat” and “sys” refer to the statistical (in the sense explained above)

and systematic uncertainties. These final estimates are also illustrated in Figs. 7 and 8 by the black dots.

As one can notice, our final results are in excellent agreement with the expectation from the model we used (fifth column of Table II), with the central values for  $a_\mu^{\text{HVP,LO}}$  off by at most 0.06% in the case of PAs and 0.05% in the case of D-Logs. Uncertainties are dominated by statistics and the extrapolation (which we book as stat). A reduction in the uncertainty of about 25% in both methods is achieved by computing the integral up to  $x = 0.990$ , which covers 99.1% of  $a_\mu^{\text{HVP,LO}}$ . This is very likely an acceptable compromise since estimating the remaining 0.9% from  $e^+e^-$  data or perturbative QCD would not increase the uncertainties in any significant way. Finally, we observe that the systematic uncertainties, which are small, do not change significantly with  $x_{\text{max}}$ .

For comparison, we also employ the QED-inspired model of Eq. (34) to perform the fits to the toy datasets. For every dataset, a value for the parameters  $K$  and  $M$  is determined, as well as a prediction for  $a_\mu^{\text{HVP,LO}}(x_{\text{max}})$ , obtained by using the fitting function to compute the integral of Eq. (10). We again quote the median as the central value with an uncertainty obtained within its 68% CL. The predictions for the different  $x_{\text{max}}$  are indicated in Table II, where the systematic error is determined as 0.24% of the central value, as explained in Sec. VI. These values can be seen in Figs. 7 and 8 represented by the pink star with the inner error band representing solely the statistical uncertainty and the larger error band representing the total uncertainty, from the sum in quadrature of the two error sources in Table II. The QED-inspired model, which has only two free parameters, clearly outperforms  $P_1^1$ , which also has only two free parameters. This is certainly a result of the additional structure of the model, which contains, for example, a logarithmic cut. On the other hand, the result from the model underestimates the true value, which reflects a systematic uncertainty associated with the model dependency, something already observed in Sec. VI. The model dependency can be inferred from the fitted values of  $K$  and  $M$ , which read  $6871^{+43}_{-38}$  GeV<sup>-2</sup> and  $0.060^{+0.004}_{-0.003}$  GeV<sup>2</sup>, respectively. In particular, the value of  $M$  is incompatible with  $m_\pi^2$ , as would be expected from the model in Eq. (22). In comparison with the final result from

the PAs and D-Logs, the model displays a smaller uncertainty, which stems from the fact that it has only two free parameters, but is further away from the true value. In this respect, the use of PAs and D-Logs helps to reduce the final systematic uncertainty, being, in addition, completely model independent.

### VIII. CONCLUSION

In this work, we described the use of Padé approximants and D-Log Padé approximants to fit and extrapolate toy data reflecting the expected results of the future MUonE experiment. This is a model-independent strategy to extract  $a_\mu^{\text{HVP,LO}}$  that relies on general knowledge about the fundamental properties of the hadronic contribution to the running of the fine-structure constant, which is a Stieltjes function in the variable  $t$  [57]. This type of functions are ruled by a determinant condition, given in Eq. (17), which generates constraints for their Taylor series coefficients. This is a key factor, since the convergence of rational approximants built from the Taylor series of Stieltjes functions is guaranteed by theorems [43,44] in a very specific pattern, where the PAs and D-Logs belonging to the diagonal sequence approach  $\Delta\alpha_{\text{had}}(t)$  from below, while the PAs of the sequence  $P_N^{N+1}(t)$  and the D-Logs of the sequence  $D_{N+1}^N(t)$  do so from above, as indicated in Eqs. (18) and (19).

In our case, the approximants are used as fitting functions that would be employed to fit the future MUonE data. Our fitting functions were constructed as follows. First, the approximants were formally built with the canonical procedure, i.e., by matching to the Taylor series of  $\Delta\alpha_{\text{had}}(t)$  around  $t=0$ , with generic Taylor coefficients that would be fixed by the data. Then, the function in the variable  $x$ , which is more suited for fits in the MUonE framework, was obtained by the change of variables of Eq. (9). These new functions in  $x$  are then used to fit the toy datasets in order to determine  $a_\mu^{\text{HVP,LO}}$ , by employing the results from the approximant in the integral of Eq. (7) and, as a by-product, to estimate the Taylor coefficients of  $\Delta\alpha_{\text{had}}(t)$ . The toy datasets used in this work were generated from the model for the Euclidean correlator proposed by Greynat and de Rafael in Ref. [40], briefly motivated in Sec. IV. We believe this simple model to be sufficient for our purposes since it captures the main features of  $\Delta\alpha_{\text{had}}(t)$  and is a Stieltjes function.

First we showed, as proof of concept, how sequences of approximants bound the true value of  $a_\mu^{\text{HVP,LO}}$  in idealized scenarios with no fluctuations in the datasets. We then turned to tests of our method in the realistic case, where the toy datasets are generated from the model with fluctuations that reflect the expected uncertainties to be obtained from the MUonE experiment. We produced 1000 datasets and fitted them with the PAs  $P_1^1, P_1^2, P_2^2$ , and  $P_2^3$ , and the D-Logs  $D_2^1, D_2^2, D_3^2$ , and  $D_3^3$ . The constraints imposed on the Taylor series coefficients, derived from the determinant condition

of Eq. (17) and given in Eq. (21), were imposed on the fits. We analyzed then the estimate of  $a_\mu^{\text{HVP,LO}}(x_{\text{max}})$  for each approximant up to  $x_{\text{max}} = 0.990, 0.995, 0.997$ , and 1. Taking advantage of the expected convergence pattern, i.e., the fact that different sequences bound the true value as in Eqs. (18) and (19), we obtained the final estimate of  $a_\mu^{\text{HVP,LO}}(x_{\text{max}})$ . Our final results are obtained from the average between the two highest-order approximants of each sequence, concretely, the PAs  $P_2^2$  and  $P_2^3$  and the D-Logs  $D_3^2$  and  $D_3^3$ . The systematic uncertainty of our final prediction can be estimated and it was calculated as half of the difference between these two approximants, while the statistical error was taken as the smallest error among the highest-order approximants of each sequence.

For all the values of  $x_{\text{max}}$  employed, our final estimates are fully compatible with what was expected from the underlying model, as one can see in Figs. 7 and 8 for the PAs and D-Logs, respectively. Our final central values differ from the expected ones by less than 0.06% in the PAs case and 0.05% for D-Logs. The final uncertainty is, in all cases, dominated by the uncertainty stemming from the extrapolation. We observe, however, that when the region in which we extrapolate the results is expanded, the dispersion between different fits grows, leading to larger final uncertainties. This can certainly be expected since the extrapolation that is performed is far from trivial and relies on data in a very limited region of the variable  $t$ . In this respect, an extrapolation in a somewhat limited interval, with  $x_{\text{max}} = 0.990$  for example, which covers 99.1% of the  $a_\mu^{\text{HVP,LO}}$  integrand, significantly reduces the final uncertainty. Of course, in this case one needs to resort to external information to complete the determination of  $a_\mu^{\text{HVP,LO}}$ .

We also compared our final prediction with the ones obtained from the QED-inspired model of Eq. (34) used in preliminary studies of the MUonE proposal [37]. This model leads to smaller uncertainties than those obtained from our procedure, in part because it has only two parameters, but its central value is further away from the true value, as seen in Figs. 7 and 8. This larger systematic uncertainty reflects a model dependency that can hardly be avoided with functions of this type.

In summary, in this work we showed that the systematic use of PAs and D-Logs as model-independent fitting functions to the future MUonE data can provide a powerful framework for the extraction of  $a_\mu^{\text{HVP,LO}}$ . Our explorations show that this method is superior to the use of a single fixed fitting function, which may carry a model dependence and an associated systematic uncertainty that would be difficult to estimate on the basis of real experimental data. The nicest feature of the method is the fact that we expect different sequences to bound the true value, which renders the average of results from these two sequences superior to the estimate arising from a single approximant. Therefore, the PAs and D-Logs provide the basis for a model-independent and systematic method, relying only on the analytical structure of the

two-point correlator underlying  $a_\mu^{\text{HVP,LO}}$ , that is able to yield a result with a competitive, although conservative, uncertainty.

### ACKNOWLEDGMENTS

The authors thank Carlo Carloni Calame, Giovanni Abbiendi, and Graziano Venanzoni for providing us with the data errors expected in the MUonE experiment. We thank Jérôme Charles for comments on a previous version of this manuscript. P. M. thanks IFSC, at University of Sao Paulo, and C. Y. L. thanks IFAE, at Universitat Autònoma de Barcelona, for hospitality. The work of C. Y. L. was financed by the São Paulo Research Foundation (FAPESP) Grants No. 2020/15532-1 and No. 2022/02328-2. The work of D. B. was supported by FAPESP Grant No. 2021/06756-6 and by CNPq Grant No. 308979/2021-4. The work of P. M. has been supported by the Ministerio de Ciencia e Innovación under Grant No. PID2020–112965 GB-I00 and by the Secretaria d'Universitats i Recerca del Departament d'Empresa i Coneixement de la Generalitat de Catalunya under Grant No. 2021 SGR 00649. I. F. A. E. is partially funded by the CERCA program of the Generalitat de Catalunya.

### APPENDIX: FITTING FUNCTIONS CONSTRUCTED USING PAs AND D-LOGS

In this appendix we give the explicit expressions of the approximants used in this work as a function of  $x$  and in terms of the unknown Taylor coefficients  $a_n$  of  $\Delta\alpha_{\text{had}}$ .

We start with the sequence  $P_N^N$ , where the first PA is  $P_1^1$ . As seen in Sec. VI, the fitting function is

$$P_1^1(x) = -\frac{b_1 m_\mu^2 x^2}{1-x + b_2 m_\mu^2 x^2}, \quad (\text{A1})$$

where the parameters are

$$\begin{aligned} b_1 &= a_1, & b_2 &= a_2, & b_3 &= a_3, \\ b_4 &= \frac{a_2 a_5 - a_3 a_4}{a_3^2 - a_2 a_4}, & b_5 &= \frac{a_4^2 - a_3 a_5}{a_3^2 - a_2 a_4}. \end{aligned} \quad (\text{A7})$$

In addition, the constraints employed in the fits are  $b_3 < b_2 < b_1 < 0$ ,  $b_4 < 0$ , and  $b_5 > 0$ . It is important to stress that all the constraints shown in this appendix are model independent, since they follow from the fact that  $\Delta\alpha_{\text{had}}(t)$  is a Stieltjes function.

The D-Logs are constructed from the Taylor series as was described in Sec. VI and reparametrized in terms of the  $x$  variable. All the functions employed for fitting purposes are detailed in Table III. Parameters  $r_n$  refer to the branch point

with  $b_1 = a_1$  and  $b_2 = a_2/a_1$ . The constraints employed are  $b_1 < 0$  and  $b_2 > 1$ . The next approximant in this sequence is  $P_2^2$ , whose final expression is

$$P_2^2(x) = \frac{b_1 m_\mu^2 x^2 (x-1) + (b_2 - b_1 b_3) m_\mu^4 x^4}{(1-x)^2 + b_3 m_\mu^2 x^2 (1-x) + b_4 m_\mu^4 x^4}, \quad (\text{A2})$$

where the fit parameters are now

$$\begin{aligned} b_1 &= a_1, & b_2 &= a_2, & b_3 &= \frac{a_2 a_3 - a_1 a_4}{a_2^2 - a_1 a_3}, \\ b_4 &= \frac{a_3^2 - a_2 a_4}{a_2^2 - a_1 a_3}. \end{aligned} \quad (\text{A3})$$

From the structure of  $\Delta\alpha_{\text{had}}$  and its series representation in Eq. (20), we know that  $b_1 < 0$  and  $b_2 < b_1$ . Analyzing the Stieltjes determinants of Eq. (17), we get the additional relations:  $b_3 > 0$  and  $b_4 > 0$ . Since the other approximants of the sequence  $P_N^N$  were not applied to the realistic datasets, we will refrain from showing their expressions here.

The  $P_N^{N+1}$  sequence starts with  $P_1^2$ , which reads, as a function of  $x$

$$P_1^2(x) = -\frac{b_1 m_\mu^2 x^2 (1-x) + (b_1 b_3 - b_2) m_\mu^4 x^4}{(1-x)^2 + b_3 m_\mu^2 x^2 (1-x)}. \quad (\text{A4})$$

The  $b_n$  parameters in this case, together with their limits, are

$$b_1 = a_1 < 0, \quad b_2 = a_2 < b_1, \quad b_3 = \frac{a_3}{a_2} > 1. \quad (\text{A5})$$

The last approximant used is  $P_2^3$ , which is given by

$$P_2^3(x) = -\frac{b_1 m_\mu^2 x^2 (1-x)^2 - (b_2 + b_1 b_4) m_\mu^4 x^4 (1-x) + (b_3 + b_1 b_5 + b_2 b_4) m_\mu^6 x^6}{(1-x)^3 - b_4 m_\mu^2 x^2 (1-x)^2 + b_5 m_\mu^4 x^4 (1-x)}, \quad (\text{A6})$$

in the  $t$  variable,  $\gamma_n$  refers to the multiplicity for the corresponding cut,  $\beta$  is an exponential factor that appears only in the diagonal sequence, and  $f_0$  is a normalization factor.

TABLE III. D-Log fitting functions as a function of  $t$  or  $x$ .

	$D_M^N(t)$	$D_M^N(x)$
$D_2^1$	$\frac{-f_0 t}{(r_1 - t)^{\gamma_1}}$	$\frac{f_0 m_\mu^2 x^2 (1-x)^{-1+\gamma_1}}{(r_1 - r_1 x + m_\mu^2 x^2)^{\gamma_1}}$
$D_2^2$	$\frac{-f_0 t e^{\beta t}}{(r_1 - t)^{\gamma_1}}$	$\frac{f_0 m_\mu^2 x^2 (1-x)^{-1+\gamma_1}}{(r_1 - r_1 x + m_\mu^2 x^2)^{\gamma_1}} e^{\beta \frac{m_\mu^2 x^2}{(1-x)}}$
$D_2^3$	$\frac{-f_0 t}{(r_1 - t)^{\gamma_1} (r_2 - t)^{\gamma_2}}$	$\frac{f_0 m_\mu^2 x^2 (1-x)^{-1+\gamma_1+\gamma_2}}{(r_1 - r_1 x + m_\mu^2 x^2)^{\gamma_1} (r_2 - r_2 x + m_\mu^2 x^2)^{\gamma_2}}$
$D_3^3$	$\frac{-f_0 t e^{\beta t}}{(r_1 - t)^{\gamma_1} (r_2 - t)^{\gamma_2}}$	$\frac{f_0 m_\mu^2 x^2 (1-x)^{-1+\gamma_1+\gamma_2}}{(r_1 - r_1 x + m_\mu^2 x^2)^{\gamma_1} (r_2 - r_2 x + m_\mu^2 x^2)^{\gamma_2}} e^{\beta \frac{m_\mu^2 x^2}{(1-x)}}$

- [1] Muon  $g-2$  Collaboration, Measurement of the positive muon anomalous magnetic moment to 0.46 ppm, *Phys. Rev. Lett.* **126**, 141801 (2021).
- [2] Muon  $g-2$  Collaboration, Measurement of the positive muon anomalous magnetic moment to 0.20 ppm, *Phys. Rev. Lett.* **131**, 161802 (2023).
- [3] Muon  $g-2$  Collaboration, Final report of the muon E821 muon anomalous magnetic moment measurement at BNL, *Phys. Rev. D* **73**, 072003 (2006).
- [4] T. Aoyama *et al.*, The anomalous magnetic moment of the muon in the Standard Model, *Phys. Rep.* **887**, 1 (2020).
- [5] T. Aoyama, M. Hayakawa, T. Kinoshita, and M. Nio, Complete tenth-order QED contribution to the muon  $g-2$ , *Phys. Rev. Lett.* **109**, 111808 (2012).
- [6] T. Aoyama, T. Kinoshita, and M. Nio, Theory of the anomalous magnetic moment of the electron, *Atoms* **7**, 28 (2019).
- [7] A. Czarnecki, W. J. Marciano, and A. Vainshtein, Refinements in electroweak contributions to the muon anomalous magnetic moment, *Phys. Rev. D* **67**, 073006 (2003).
- [8] C. Gnendiger, D. Stöckinger, and H. Stöckinger-Kim, The electroweak contributions to  $(g-2)_\mu$  after the Higgs boson mass measurement, *Phys. Rev. D* **88**, 053005 (2013).
- [9] M. Davier, A. Hoecker, B. Malaescu, and Z. Zhang, Reevaluation of the hadronic vacuum polarisation contributions to the Standard Model predictions of the muon  $g-2$  and  $\alpha(m_Z^2)$  using newest hadronic cross section data, *Eur. Phys. J. C* **77**, 827 (2017).
- [10] A. Keshavarzi, D. Nomura, and T. Teubner, Muon  $g-2$  and  $\alpha(M_Z^2)$ : A new data-based analysis, *Phys. Rev. D* **97**, 114025 (2018).
- [11] M. Davier, A. Hoecker, B. Malaescu, and Z. Zhang, A new evaluation of the hadronic vacuum polarisation contributions to the muon anomalous magnetic moment and to  $\alpha(m_Z^2)$ , *Eur. Phys. J. C* **80**, 241 (2020).
- [12] A. Keshavarzi, D. Nomura, and T. Teubner,  $g-2$  of charged leptons,  $\alpha(M_Z^2)$ , and the hyperfine splitting of muonium, *Phys. Rev. D* **101**, 014029 (2020).
- [13] G. Colangelo, M. Hoferichter, and P. Stoffer, Two-pion contribution to hadronic vacuum polarization, *J. High Energy Phys.* **02** (2019) 006.
- [14] M. Hoferichter, B.-L. Hoid, and B. Kubis, Three-pion contribution to hadronic vacuum polarization, *J. High Energy Phys.* **08** (2019) 137.
- [15] B.-L. Hoid, M. Hoferichter, and B. Kubis, Hadronic vacuum polarization and vector-meson resonance parameters from  $e^+e^- \rightarrow \pi^0\gamma$ , *Eur. Phys. J. C* **80**, 988 (2020).
- [16] A. Kurz, T. Liu, P. Marquard, and M. Steinhauser, Hadronic contribution to the muon anomalous magnetic moment to next-to-next-to-leading order, *Phys. Lett. B* **734**, 144 (2014).
- [17] K. Melnikov and A. Vainshtein, Hadronic light-by-light scattering contribution to the muon anomalous magnetic moment reexamined, *Phys. Rev. D* **70**, 113006 (2004).
- [18] P. Masjuan and P. Sanchez-Puertas, Pseudoscalar-pole contribution to the  $(g_\mu-2)$ : A rational approach, *Phys. Rev. D* **95**, 054026 (2017).
- [19] G. Colangelo, M. Hoferichter, M. Procura, and P. Stoffer, Rescattering effects in the hadronic-light-by-light contribution to the anomalous magnetic moment of the muon, *Phys. Rev. Lett.* **118**, 232001 (2017).
- [20] G. Colangelo, M. Hoferichter, M. Procura, and P. Stoffer, Dispersion relation for hadronic light-by-light scattering: Two-pion contributions, *J. High Energy Phys.* **04** (2017) 161.
- [21] M. Hoferichter, B.-L. Hoid, B. Kubis, S. Leupold, and S. P. Schneider, Pion-pole contribution to hadronic light-by-light scattering in the anomalous magnetic moment of the muon, *Phys. Rev. Lett.* **121**, 112002 (2018).
- [22] M. Hoferichter, B.-L. Hoid, B. Kubis, S. Leupold, and S. P. Schneider, Dispersion relation for hadronic light-by-light scattering: Pion pole, *J. High Energy Phys.* **10** (2018) 141.
- [23] A. Gérardin, H. B. Meyer, and A. Nyffeler, Lattice calculation of the pion transition form factor with  $N_f = 2 + 1$  Wilson quarks, *Phys. Rev. D* **100**, 034520 (2019).
- [24] J. Bijnens, N. Hermansson-Truedsson, and A. Rodríguez-Sánchez, Short-distance constraints for the HLbL contribution to the muon anomalous magnetic moment, *Phys. Lett. B* **798**, 134994 (2019).
- [25] G. Colangelo, F. Hagelstein, M. Hoferichter, L. Laub, and P. Stoffer, Short-distance constraints on hadronic light-by-light scattering in the anomalous magnetic moment of the muon, *Phys. Rev. D* **101**, 051501 (2020).
- [26] G. Colangelo, F. Hagelstein, M. Hoferichter, L. Laub, and P. Stoffer, Longitudinal short-distance constraints for the hadronic light-by-light contribution to  $(g-2)_\mu$  with large- $N_c$  Regge models, *J. High Energy Phys.* **03** (2020) 101.
- [27] G. Colangelo, M. Hoferichter, A. Nyffeler, M. Passera, and P. Stoffer, Remarks on higher-order hadronic corrections to the muon  $g-2$ , *Phys. Lett. B* **735**, 90 (2014).
- [28] T. Blum, N. Christ, M. Hayakawa, T. Izubuchi, L. Jin, C. Jung, and C. Lehner, Hadronic light-by-light scattering contribution to the muon anomalous magnetic moment from lattice QCD, *Phys. Rev. Lett.* **124**, 132002 (2020).
- [29] S. Borsanyi *et al.*, Leading hadronic contribution to the muon magnetic moment from lattice QCD, *Nature (London)* **593**, 51 (2021).
- [30] M. Hansen, A. Lupo, and N. Tantalo, Extraction of spectral densities from lattice correlators, *Phys. Rev. D* **99**, 094508 (2019).
- [31] G. Baïlas, S. Hashimoto, and T. Ishikawa, Reconstruction of smeared spectral function from Euclidean correlation functions, *Prog. Theor. Exp. Phys.* **2020**, 043B07 (2020).
- [32] D. Boito, M. Golterman, K. Maltman, and S. Peris, Spectral-weight sum rules for the hadronic vacuum polarization, *Phys. Rev. D* **107**, 034512 (2023).
- [33] J. A. Miranda and P. Roig, New  $\tau$ -based evaluation of the hadronic contribution to the vacuum polarization piece of the muon anomalous magnetic moment, *Phys. Rev. D* **102**, 114017 (2020).
- [34] P. Masjuan, A. Miranda, and P. Roig,  $\tau$  data-driven evaluation of Euclidean windows for the hadronic vacuum polarization, *Phys. Lett. B* **850**, 138492 (2024).
- [35] C. M. Carloni Calame, M. Passera, L. Trentadue, and G. Venanzoni, A new approach to evaluate the leading hadronic corrections to the muon  $g-2$ , *Phys. Lett. B* **746**, 325 (2015).
- [36] G. Abbiendi *et al.*, Measuring the leading hadronic contribution to the muon  $g-2$  via  $\mu e$  scattering, *Eur. Phys. J. C* **77**, 139 (2017).

- [37] G. Abbiendi, Letter of Intent: The MUonE project, Technical Report No. CERN-SPSC-2019-026; SPSC-I-252, CERN, Geneva, 2019.
- [38] G. Abbiendi, Status of the MUonE experiment, *Phys. Scr.* **97**, 054007 (2022).
- [39] F. Ignatov, R. N. Pilato, T. Teubner, and G. Venanzoni, An alternative evaluation of the leading-order hadronic contribution to the muon  $g-2$  with MUonE, *Phys. Lett. B* **848**, 138344 (2024).
- [40] D. Greynat and E. de Rafael, Hadronic vacuum polarization and the MUonE proposal, *J. High Energy Phys.* **05** (2022) 084.
- [41] P. Masjuan and S. Peris, Padé Theory applied to the vacuum polarization of a heavy quark, *Phys. Lett. B* **686**, 307 (2010).
- [42] M. Golterman, K. Maltman, and S. Peris, New strategy for the lattice evaluation of the leading order hadronic contribution to  $(g-2)_\mu$ , *Phys. Rev. D* **90**, 074508 (2014).
- [43] G. A. Baker, *Essentials of Padé Approximants* (Academic Press, New York, 1975).
- [44] G. A. Baker and P. Graves-Morris, *Padé Approximants*, 2nd ed. (Cambridge University Press, Cambridge, England, 1996).
- [45] P. Masjuan Queralto, Rational approximations in quantum chromodynamics, Ph.D. thesis, Universitat Autònoma de Barcelona, 2010.
- [46] O. Costin and G. V. Dunne, Conformal and uniformizing maps in Borel analysis, *Eur. Phys. J. Special Topics* **230**, 2679 (2021).
- [47] D. Boito, C. Y. London, and P. Masjuan, Higher-order QCD corrections to  $H \rightarrow b\bar{b}$  from rational approximants, *J. High Energy Phys.* **01** (2022) 054.
- [48] D. Boito, P. Masjuan, and F. Oliani, Higher-order QCD corrections to hadronic  $\tau$  decays from Padé approximants, *J. High Energy Phys.* **08** (2018) 075.
- [49] G. Ballerini *et al.*, A feasibility test run for the MUonE project, *Nucl. Instrum. Methods Phys. Res., Sect. A* **936**, 636 (2019).
- [50] G. Abbiendi *et al.*, Results on multiple Coulomb scattering from 12 and 20 GeV electrons on carbon targets, *J. Instrum.* **15**, 01 (2020).
- [51] G. Abbiendi *et al.*, A study of muon-electron elastic scattering in a test beam, *J. Instrum.* **16**, P06005 (2021).
- [52] C. Bouchiat and L. Michel, La résonance dans la diffusion méson  $\pi$ —méson  $\pi$  et le moment magnétique anormal du méson  $\mu$ , *J. Phys. Radium* **22**, 121 (1961).
- [53] S. J. Brodsky and E. De Rafael, Suggested boson-lepton pair couplings and the anomalous magnetic moment of the muon, *Phys. Rev.* **168**, 1620 (1968).
- [54] B. E. Lautrup and E. De Rafael, Calculation of the sixth-order contribution from the fourth-order vacuum polarization to the difference of the anomalous magnetic moments of muon and electron, *Phys. Rev.* **174**, 1835 (1968).
- [55] B. e. Lautrup, A. Peterman, and E. de Rafael, Recent developments in the comparison between theory and experiments in quantum electrodynamics, *Phys. Rep.* **3**, 193 (1972).
- [56] D. Bernecker and H. B. Meyer, Vector correlators in lattice QCD: Methods and applications, *Eur. Phys. J. A* **47**, 148 (2011).
- [57] C. Aubin, T. Blum, M. Golterman, and S. Peris, Model-independent parametrization of the hadronic vacuum polarization and  $g-2$  for the muon on the lattice, *Phys. Rev. D* **86**, 054509 (2012).
- [58] T. Carleman, Über die Approximation analytischer Funktionen durch lineare Aggregate von vorgegebenen Potenzen, *Ark. Mat. Astr. Fys.* **17**, 1 (1923).
- [59] P. Masjuan, S. Peris, and J. J. Sanz-Cillero, Vector meson dominance as a first step in a systematic approximation: The pion vector form-factor, *Phys. Rev. D* **78**, 074028 (2008).
- [60] P. Masjuan,  $\gamma^*\gamma \rightarrow \pi^0$  transition form factor at low-energies from a model-independent approach, *Phys. Rev. D* **86**, 094021 (2012).
- [61] P. Masjuan and J. J. Sanz-Cillero, Padé approximants and resonance poles, *Eur. Phys. J. C* **73**, 2594 (2013).
- [62] R. Escribano, P. Masjuan, and P. Sanchez-Puertas,  $\eta$  and  $\eta'$  transition form factors from rational approximants, *Phys. Rev. D* **89**, 034014 (2014).
- [63] R. Escribano, S. González-Solís, P. Masjuan, and P. Sanchez-Puertas,  $\eta'$  transition form factor from space- and timelike experimental data, *Phys. Rev. D* **94**, 054033 (2016).