Reexamining the proton-radius problem using constrained Gaussian processes

Shuang Zhou, P. Giulani, J. Piekarewicz, Anirban Bhattacharya, and Debdeep Pati

Phys. Rev. C **99**, 055202 — Published 14 May 2019

DOI: 10.1103/PhysRevC.99.055202

Revisiting the proton-radius problem using constrained Gaussian processes

Shuang Zhou,[1] P. Giulani,[2] J. Piekarewicz,[2] Anirban Bhattacharya,[1] and Debdeep Pati[1]

[1]*Department of Statistics, Texas A&M University, College Station, TX 77843*
[2]*Department of Physics, Florida State University, Tallahassee, FL 32306*

Background: The "proton radius puzzle" refers to an eight-year old problem that highlights major inconsistencies in the extraction of the charge radius of the proton from muonic Lamb-shift experiments as compared against experiments using elastic electron scattering. For the latter approach, the determination of the charge radius involves an extrapolation of the experimental form factor to zero momentum transfer.

Purpose: To estimate the proton radius, a novel and powerful nonparametric method based on a constrained Gaussian process is introduced. The constrained Gaussian process models the electric form factor as a function of the momentum transfer.

Methods: Within a Bayesian paradigm, we develop a model flexible enough to fit the data without any parametric assumptions on the form factor. The Bayesian estimation is guided by imposing only two physical constraints on the form factor: (a) its value at zero momentum transfer (normalization) and (b) its overall shape, assumed to be a monotonically decreasing function of the momentum transfer. Variants of these assumptions are explored to assess their impact.

Results: By adopting both constraints and incorporating the whole range of experimental data available we extracted a charge radius of $r_p = 0.845 \pm 0.001\,\mathrm{fm}$, consistent with the muonic experiment. Nevertheless, we show that within our model the extracted radius depends on both the assumed constraints and the range of experimental data used to fit the Gaussian process. For example, if only low momentum-transfer data is used, relaxing the normalization constraint provides a value compatible with the larger electronic value.

Conclusions: We have presented a novel technique to estimate the proton radius from electron scattering data based on a constrained Gaussian process. We demonstrated that the impact of imposing sensible physical constraints on the form factor is substantial. Also critical is the range of the experimental data used in the extrapolation. We are hopeful that as the technique gets refined, together with the anticipated new results from the PRad experiment, we will get closer to a resolution of the puzzle.

I. INTRODUCTION

Nuclear Physics is an extremely broad field of science whose mission is to understand all manifestations of nuclear phenomena [1]. Regardless of whether probing individual nucleons, atomic nuclei, or neutron stars, a common theme across this vast landscape is the characterization of these objects in terms of their mass and radius. Indeed, shortly after the discovery of the neutron in 1932, Gamow, Weizsäcker, Bethe, and Bacher formulated the "liquid-drop" model to estimate the masses of atomic nuclei [2, 3]. Since then, remarkable advances in experimental techniques have been exploited to determine nucleon and nuclear masses with unprecedented precision; for example, the rest mass of the proton is known to a few parts part in a billion [4]. Similarly, starting with the pioneering work of Hofstadter in the late 1950's [5] and continuing to this day [6–8], elastic electron scattering has provided the most accurate and detailed picture of the distribution of charge in nuclear systems. Although not as impressive as in the case of nuclear masses, the charge radii of atomic nuclei has nevertheless been determined with extreme precision; for example, the charge radius of ^{208}Pb is known to about two parts in 10,000 [8] (or $R_{\rm ch}^{208}\!=\!5.5012(13)\,{\rm fm}$). Given such an impressive track record, it came as a shocking surprise that the accepted 2010-CODATA value for the charge radius of the proton obtained from electronic hydrogen and electron scattering was in stark disagreement with a new result obtained from the Lamb shift in muonic hydrogen [9]. This unforeseen conflict with the structure of the proton has given rise to the "proton radius puzzle" [10–12],

The value of the charge radius of the proton $r_p\!=\!0.84087(39)\,{\rm fm}$ determined from muonic hydrogen [9, 10] differs significantly (by $\sim$4% or nearly 7σ) from the recommended CODATA value of $r_p = 0.8775(51)\,{\rm fm}$. Note that the CODATA value is obtained by combining the results from both electron scattering and atomic spectroscopy [4, 10, 12]. The muonic measurement is so remarkably precise because the muon—with a mass that is more than 200 times larger than the electron mass and thus a Bohr radius 200 times smaller—is a much more sensitive probe of the internal structure of the proton. Of great relevance to the proton puzzle is the recent measurement of the 2S-4P transition frequency in electronic hydrogen that suggests a smaller proton radius of $r_p\!=\!0.8335(95)\,{\rm fm}$—in agreement with the result from muonic hydrogen [13]. Although significant, it remains to be understood why the present extraction differs from the large number of spectroscopic measurements carried out in electronic hydrogen throughout the years.

As in the case of earlier physics puzzles—notably the "solar neutrino problem"—one attempts to explain the discrepancy by exploring three non-mutually-exclusive options: (a) the experiment (at least one of them) is in error, (b) theoretical models used in the extraction of the proton radius are the culprit (see for example [14] and references contained therein), or (c) there is new physics that affects the muon differently than the electron. Indeed, hints of possible violations to lepton universality are manifested in the anomalous magnetic moment $(g\!-\!2)$ of the muon [15] and in certain decays of the B-meson into either a pair of electrons or a pair of muons [16].

In an effort to resolve the "proton radius puzzle" a suite of experiments in both spectroscopy and lepton-proton scattering are being commissioned. Spectroscopy of both electronic and muonic atoms, as already initiated by Beyer *et al.* [13], will continue with a measurement of a variety of transitions to improve both the value of the Rydberg constant and the charge radius of the proton; note that the Rydberg constant and r_p are known to be highly correlated. Lepton scattering experiments are planned at both the Thomas Jefferson National Accelerator Facility (JLab) and at the Paul Scherrer Institute (PSI). The proton radius experiment (PRad) at JLab has already collected data in the momentum-transfer range of $Q^2\!=\!(10^{-4}\!-\!10^{-1})\,{\rm GeV}^2$ [17], a wide-enough region to allow for comparisons against the most recent Mainz data [18], but also to extend the Mainz data to significantly lower values of Q^2. Finally, the Muon Proton Scattering Experiment (MUSE) will fill a much-needed gap by determining r_p from the scattering of both positive and negative muons of the proton. These experiments will be conducted concurrently with electron scattering measurements in an effort to minimize systematic uncertainties [19].

Within this broad context our contribution is rather modest, as our main goal is to address how best to extract the charge radius of the proton form existing electron scattering data. The view adopted here is that the puzzle lays not in the experimental data, but rather in the extraction of the proton radius from the scattering data. The proton charge radius is related to the slope of the electric form factor of the proton $G_{\rm E}(Q^2)$ at the origin, *i.e.*, at $Q^2\!=\!0$ (see Sec.II). Despite heroic efforts at both Mainz [18] and JLab [17] to determine $G_{\rm E}(Q^2)$ at extremely low values of Q^2, a subtle *extrapolation* to $Q^2 = 0$ is unavoidable. Given the current data available, the value one can obtain for the proton radius from the extrapolation is quite sensitive to the model used to describe the form factor. In a first attempt at mitigating such uncontrolled extrapolations, Higinbotham and collaborators have brought to bear the power of statistical methods into the solution of the problem [20]; see also [21]. They have concluded that "statistically justified linear extrapolations of the extremely-low-Q^2 data produce a proton charge radius which is consistent with the muonic results and is systematically smaller than the one extracted using higher-order extrapolation functions". However, recent analyses of electron scattering data that suggest smaller proton radii consistent with the muonic Lamb shift have been called into question [22]. Moreover, much controversy has been generated around the optimal ("parametric") model that should be used to fit the electric charge form factor of the proton—ranging from monopole, to dipole, to polynomial fits, to Pade' approximants, among many others. There has been also several efforts in performing extractions that rely on analytical properties of the form factor, see for example [23, 24].

In an effort to eliminate the reliance on specific functional forms, we introduce a method that does not assume a particular parametric form for the form factor. Such a nonparametric approach aims to "let data speak for itself" without introducing any preconceived biases. Although the nonparametric approach does not assume a particular form for the form factor, several constraints justified by physical considerations are imposed. In essence, a *nonparametric Bayesian* curve fitting procedure that incorporates various physical constraints is used to provide robust predictions and uncertainty estimates for the charge radius of the proton. In our analysis we use the 1422 data points from the Mainz collaboration [25].

The paper has been organized as follows. In Sec. II, we introduce some of the basic concepts necessary to understand the measurement of the electric form factor of the proton. After such brief introduction, we explain the critical concepts behind our nonparametric approach, including the selection of the basis functions and the Gaussian process used for their calibration. The electron-scattering data analysis is presented Sec. III. We offer our conclusions and some perspective for future improvements in Sec. IV. Finally, for pedagogical reasons, the analysis of synthetic data and several details about the implementation of the model and on the analysis of the real data are presented in the supplemental material [26].

II. FORMALISM

We start this section with a brief introduction to elastic electron scattering with particular emphasis on the determination of the electric form factor of the proton from the scattering data. Then, we proceed in significant more detail to describe the formalism associated with the determination of the charge radius of the proton by extrapolating the experimental data to zero momentum transfer.

A. Electron scattering

In the one-photon exchange approximation, the most general expression for the elastic electron-proton cross section consistent with Lorentz and parity invariance is encoded in two Lorentz-scalar functions: the electric G_E and magnetic G_M form factors of the proton. That is,

$$\frac{d\sigma}{d\Omega} = \left(\frac{d\sigma}{d\Omega}\right)_{\text{Mott}} \left(\frac{G_E^2(Q^2) + \tau G_M^2(Q^2)}{1+\tau} + 2\tau G_M^2(Q^2)\tan^2(\theta/2)\right), \tag{1}$$

where the square of the four-momentum transfer is given by:

$$Q^2 \equiv -(p'-p)^2 = 4EE'\sin^2(\theta/2). \tag{2}$$

Note that E (E') is the initial (final) energy of the electron, θ is the scattering angle (all in the laboratory frame), $\tau \equiv Q^2/4M^2$, and M is the mass of the proton. The internal structure of the proton is imprinted in the two form factors, with the electric one describing (in a non-relativistic picture) the distribution of charge and the magnetic one the distribution of magnetization. Finally, the Mott cross section introduced in Eq.(1) represents the scattering of a massless electron from a spinless and structureless point charge. That is,

$$\left(\frac{d\sigma}{d\Omega}\right)_{\text{Mott}} = \frac{4\alpha^2}{Q^4}\frac{E'^3}{E}\cos^2(\theta/2) = \frac{\alpha^2}{4E^2\sin^4(\theta/2)}\frac{E'}{E}\cos^2(\theta/2), \tag{3}$$

where α is the fine structure constant.

At low-energy –or long wavelength– electrons are unable to resolve the internal structure of the proton and are therefore only sensitive to its entire charge. As the momentum transfer increases and the wavelength becomes commensurate with the proton size, finer details may now be resolved. In particular, the charge radius of the proton is defined as:

$$r_p^2 \equiv \langle r_{\scriptscriptstyle E}^2 \rangle = -6\frac{dG_E}{dQ^2}\bigg|_{Q^2=0}. \tag{4}$$

For a recent discussion on the subtleties of the definition of the proton radius, see [27].

Once our target, the slope at zero of the electric form factor, has been defined, we introduce the following constraints that are the cornerstone of the nonparametric approach:

$$G_E(Q^2\!=\!0) = 1, \tag{5a}$$

$$G'_E(Q^2) \equiv \frac{dG_E}{dQ^2} < 0, \tag{5b}$$

$$G''_E(Q^2) \equiv \frac{d^2G_E}{d(Q^2)^2} > 0. \tag{5c}$$

The first equation (5a) is model independent since it is directly related to the charge of the proton. The other two equations (5b)-(5c), which we will call the shape constraints, are not guaranteed by the definitions above, but rather, are deduced from the analytic properties of the form factor, see for example [28, 29]. Most of the parametric models that have been used for the form factor respect these shape constraints for the parameters estimated, see for example [21].

B. Approximating G_E: basis construction

Having introduced the electric form factor of the proton we now proceed to build a flexible nonparametric model that will allow us to extrapolate $G_E(Q^2)$ to $Q^2\!=\!0$. Our main goal is to incorporate the general constraints given in Eq. (5) into the estimation procedure without making parametric assumptions on the functional form of $G_E(Q^2)$. The available experimental data will guide the shape of such nonparametric curve, ultimately allowing us to estimate r_p. In this section we describe the fully constrained model while in Sec. III we explore the impact of relaxing the constraints.

To facilitate the implementation of the nonparametric approach, we select a maximum value of Q^2, Q^2_{max}, up to where the analysis is performed, a selection that has been shown to impact the estimation of r_p. Once the momentum-transfer range has been selected, $0 \leq Q^2 \leq Q^2_{\text{max}}$, we define the dimensionless scaled variable x as $x = Q^2/Q^2_{\text{max}}$. Following this definition the range of x is the unit interval $x \in [0, 1]$.

We start by defining a working grid formed by a collection of $N\!+\!1$ equally spaced points $x_j\!=\!j/N$ in the closed interval $[0, 1]$, in which j runs from 1 to N. We adopt the notation of [30] to define a set of basis functions:

$$h_j(x) = \begin{cases} 1 - N|x - x_j|, & \text{if } |x - x_j| \leq 1/N; \\ 0, & \text{otherwise.} \end{cases} \tag{6}$$

and its corresponding integrals as follows:

$$\psi_j(x) = \int_0^x h_j(t)\, dt, \tag{7}$$

$$\phi_j(x) = \int_0^x dt \int_0^t h_j(s)\, ds. \tag{8}$$

Figure 1 shows the form of these functions for $j = 0$. We characterize our regression model in terms of $(N\!+\!3)$ free parameters ξ_j that will be obtained from a suitable fit to the experimental data. That is,

$$f(x) \approx f_\xi(x) \equiv \xi_1 + \xi_2\, x + \sum_{j=0}^{N} \xi_{j+3}\, \phi_j(x), \tag{9}$$

where $f(x)$ is the form factor in the rescaled variable x. Under this scheme, the terms ξ_1 and ξ_2 relate to the value of the modeled function $f(x)$ and its derivative at zero, respectively, while the terms ξ_{j+3} relate to the values of the second derivative of the modeled function on the grid points, but during the regression process all of them remain as free coefficients.

We explain in detail in Sec. I of the supplemental material [26] the construction of this approximation for $f(x)$ and the relation between the terms ξ_{j+3} and the values of its second derivative. Moreover, this approximation has a nice physical underpinning. If we regard $f(t)$ as the one-dimensional trajectory of a particle as a function of time t, then the approximation:

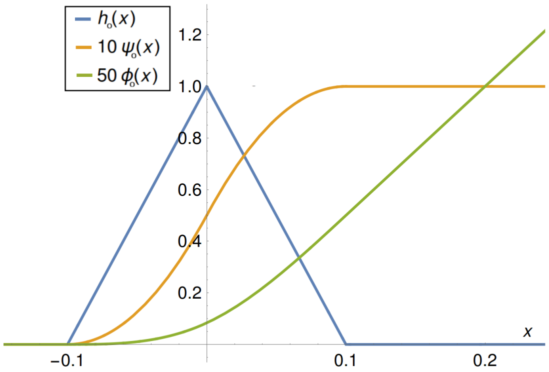

FIG. 1. Functions $h_0(x)$ (blue), $\psi_0(x)$ (orange) and $\phi_0(x)$ (green) for $N = 10$. The functions ψ_0 and ϕ_0 have been rescaled by a factor of 10 and 50 respectively.

$$f(t) \approx f(0) + t f'(0) + \sum_{j=0}^{N} f''(t_j)\, \phi_j(t), \tag{10}$$

may be explained as follows. At time $t = 0$ the particle starts at a position $f(0)$ with an initial velocity $f'(0)$. As time evolves, corrections to the straight-line trajectory are implemented by the different ϕ_j in proportion to $f''(t_j)$, that can be thought as “acceleration spikes” that stir the particle into the correct trajectory.

C. Incorporating physical constraints

The great virtue of the nonparametric approach adopted here is that no assumption is made about the functional form of the electric form factor. However, if the calibration parameters ξ_j defined in Eq. (9) are left unrestricted, the resulting model for $f_\xi(x)$ is likely to violate the physical constraints imposed in Eq. (5). In the notation assumed in this section these constraints are given by: (a) $f_\xi(0) = 1$, (b) $f'_\xi(x) < 0$, and (c) $f''_\xi(x) > 0$, where (b) and (c) hold for all x in $[0, 1]$. In this section we discuss the model formulation with all the constraints. In order to satisfy these constraints the model parameters must obey the following linear relations:

$$\xi_1 = 1, \tag{11a}$$

$$\xi_2 + \sum_{j=0}^{N} c_j\, \xi_{j+3} \le 0, \tag{11b}$$

$$\xi_{j+3} \ge 0, \text{ for } j = 0, 1, \ldots, N. \tag{11c}$$

where $c_j = \psi_j(1)$ is the area under the triangle formed by the function $h_j(x)$, except for the first one c_0 and last one c_N which are equal to half the area of the triangle. The proof of this statement is shown in Appendix A.

In order to incorporate the constraints in Eq. (11) let us call $\mathcal{C}_\xi$ the set of all the ξ_j that satisfies the relationships (11). In formal notation that would be:

$$\mathcal{C}_\xi \equiv \left\{ \xi \in \mathbb{R}^{N+3} :\ \xi_1 = 1,\ \xi_2 + \sum_{j=0}^{N} c_j\ \xi_{j+3} \le 0,\ \xi_{j+3} \ge 0,\ j = 0, \ldots, N \right\}. \tag{12}$$

The proton radius introduced in Eq. (4) is expressed directly in terms of ξ_2 as:

$$r_p = \frac{\sqrt{-6\xi_2}}{Q_{\max}}, \tag{13}$$

where $Q_{\max}$ enters to account for the rescaling of Q^2 into the dimensionless variable $x = Q^2/Q^2_{\max}$.

Note that the value of ξ_1 is fixed at 1 and r_p only depends on the value of ξ_2 in the constraint set $\mathcal{C}_\xi$. We provided a detailed discussion on a partially constrained model with the condition $\xi_1 = 1$ removed in Sec. II on the supplemental material [26]. The rest of the discussion in the following sections obeys a fully constrained model.

D. Probabilistic model for fully constrained function estimation

The observed experimental data consists of n pairs of the form (x_i, g_i), where $x_i = Q_i^2/Q^2_{\max}$ and g_i is equal to the form factor $G_E(Q_i^2)$ up to some experimental noise. Specifically, one assumes that the n experimental measurements g_i have normally distributed experimental errors ϵ_i. That is, $g_i = G_E(Q_i^2) + \epsilon_i$, where we assume that each ϵ_i is a normally distributed variable with zero mean and standard deviation σ, an unknown hyperparameter that we learn from the data.

Let $Y = (y_1, \ldots, y_n)^{\mathrm{T}}$ with $y_i := g_i - \xi_1 = g_i - 1$ (the subtraction of the independent term ξ_1 is made in order to build an homogeneous matrix equation), and set $\varepsilon = (\epsilon_1, \ldots, \epsilon_n)^{\mathrm{T}}$. Also, define a basis matrix Φ (a $n \times (N+2)$ matrix) with ith row $(x_i, \phi_0(x_i), \ldots, \phi_N(x_i))$. With these ingredients, we express our model in vectorized notation as:

$$Y = \Phi\xi + \varepsilon, \quad \varepsilon \sim \mathcal{N}_n(0, \sigma^2 \mathrm{I}_n), \quad \xi \in \mathcal{C}_\xi, \tag{14}$$

where $\mathcal{C}_\xi$ is defined in Eq. (12) and I_n denotes the unity matrix of size n. The notation $v \sim \mathcal{N}_n(\mu, \Sigma)$ means that the random variable v follows a multivariate Gaussian distribution with mean μ and covariance matrix Σ.

We operate in a Bayesian framework [31] and express pre-experimental uncertainty in ξ through a prior distribution $P(\xi)$. The prior for ξ is combined with the data likelihood $P(Y|\xi)$ to obtain the posterior distribution for ξ given the observed values Y:

$$P(\xi|Y) = \frac{P(Y|\xi)P(\xi)}{P(Y)}. \tag{15}$$

This posterior distribution of the parameters $P(\xi|Y)$ can then be used to make inference on r_p including point estimates and uncertainty quantification through credible intervals. Since we assume Gaussian distributed noise ε_i for the observational points y_i, our likelihood term $P(Y|\xi)$ will be of the form $Y \sim \mathcal{N}_n(\Phi\xi, \sigma^2 \mathrm{I}_n)$, which represents an exponential decay in the square of the difference between our observed data and our model prediction, usually denoted by χ^2 and defined as: $\chi^2 = \sum_1^n (Y_i - f_\xi(x_i))^2$. The choice of a suitable prior $P(\xi)$ is critical for a valid inference on r_p. A flexible representation for f can be reproduced through the coefficients ξ which is in turn relatable to f through its derivatives (see Sec. I on the supplemental material [26]). In the unconstrained setting, a natural choice of prior for ξ can be induced through a Gaussian process prior on f. On the other hand, the prior for ξ should be supported on the restricted space $\mathcal{C}_\xi$ so that any prior draw obeys the constraints for ξ. We combine these two features to propose a flexible constrained Gaussian prior for ξ and describe this procedure in the following subsection.

E. Prior specification: Constrained Gaussian Process

A Gaussian process (GP) [32] is a distribution of functions on the functions space such that the collection of random variables obtained by evaluating the random function at a finite set of points is multivariate Gaussian (see [32]). We use the notation $f \sim \mathrm{GP}(\mu, \tau^2 K)$ to denote that the function f follows a Gaussian process with mean function μ and covariance function $\tau^2 K$, where τ is a hyperparameter controlling the overall scale of K. This means that any finite collection of points $f_1(x_1), ...f_N(x_N)$ at locations $x_1, ..., x_N$ has a joint normal Gaussian distribution given by:

$$\Big(f_1(x_1), ...f_N(x_N)\Big) \sim \mathcal{N}(\mu, \Sigma), \tag{16}$$

where $\mu = \big(\mu(x_1), \ldots, \mu(x_N)\big)$ and $\Sigma_{ij} = \tau^2 K(x_i, x_j)$. The mean function $\mu(x)$ controls around *where* the points are distributed, and the covariance function $K(x, x')$ controls *how* the points are distributed, τ being an expected size for the deviations. The covariance function $K(x, x')$ would allow us to introduce a notion of correlation between different four momenta Q^2 in the form factor, allowing our estimation for the radius (related to the slope at $Q^2 = 0$) to borrow information for the entire Q^2 range.

The model parameters ξ_j are related to first and second derivatives of the form factor G_E, or equivalently to its rescaled version f at the various grid points x_j. Since Gaussian processes are closed under linear operations, such as taking derivatives [32], they represent an optimal choice in estimating the form factor. If $f \sim \mathrm{GP}(0, \tau^2 K)$[1], then any finite number of observations $f(x_1), ..., f(x_N)$ follow the distribution specified by Eq. (16). Therefore, a collection of

---

[1] The selection $\mu(x) = 0$ is done to avoid centering the GP around any parametric form.

random variables that involves derivatives $f'(0), f''(x_0)..., f''(x_N)$ also follow a Gaussian distribution with a covariance matrix Γ that involves up to four mixed partial derivatives of the covariance function $K(x,x')$; see Theorem 2.2.2 in [33]. That is,

$$\Gamma = \begin{bmatrix} \frac{\partial^2 K}{\partial x \partial x'}(0,0) & \frac{\partial^3 K}{\partial x \partial x'^2}(0,x_0) & \cdots & \frac{\partial^3 K}{\partial x \partial x'^2}(0,x_N) \\ \frac{\partial^3 K}{\partial x^2 \partial x'}(x_0,0) & \frac{\partial^4 K}{\partial x^2 \partial x'^2}(x_0,x_0) & \cdots & \frac{\partial^4 K}{\partial x^2 \partial x'^2}(x_0,x_N) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^3 K}{\partial x^2 \partial x'}(x_N,0) & \frac{\partial^4 K}{\partial x^2 \partial x'^2}(x_N,x_0) & \cdots & \frac{\partial^4 K}{\partial x^2 \partial x'^2}(x_N,x_N) \end{bmatrix}_{(N+2)\times(N+2)}. \tag{17}$$

For illustration purposes consider the first row of the matrix Γ. It specifies how the derivative of the function at zero, ξ_2, correlates with all the other ξ_j. The correlation between ξ_2 and the other ξ_j for $j > 2$ is controlled by the mixed partial third derivative of K at x_j.

If the model parameters ξ_j are left unconstrained, then a natural prior, induced from a GP prior on the unknown function f, would be a finite-dimensional Gaussian prior $\xi \sim \mathcal{N}_{N+2}(0, \tau^2\,\Gamma)$ with Γ as in Eq. (17). However, since the various shape constraints on the function impose a corresponding set of constraints on the model parameters, we adopted a truncated Gaussian prior on ξ:

$$p(\xi) = \frac{1}{M_\xi}\,(2\pi)^{-(N+2)/2}\,|\Gamma|^{-1/2}\,(\tau^2)^{-(N+2)/2}\,e^{-\frac{\xi^T\Gamma^{-1}\xi}{2\tau^2}}\,\mathbb{1}_{\mathcal{C}_\xi}(\xi), \tag{18}$$

where the "indicator function" $\mathbb{1}_{\mathcal{C}_\xi}(\xi)$ filters the ξ_j such that only the allowed combinations are those that satisfy the constraints listed in Eq. (11): $\mathbb{1}_{\mathcal{C}_\xi}(\xi) = 1$ if $\xi \in C_\xi$, and $\mathbb{1}_{\mathcal{C}_\xi}(\xi) = 0$ otherwise. In the above expression M_ξ is a constant of proportionality required to make $p(\xi)$ a density distribution, *i.e.*, $p(\xi)$ must integrate to one. As is commonly done [34] we have placed an (improper) objective prior on τ^2. We shall denote $p(\xi)$ by $\mathcal{N}_{N+2}(0, \tau^2\,\Gamma)\mathbb{1}_{\mathcal{C}_\xi}(\xi)$ and refer to it as the constrained Gaussian Process (cGP) prior for ξ.

To fully specify the cGP prior we still need to define the covariance function $K(x,x')$ that determines the matrix Γ. Following common practice, we chose K to be a stationary Matérn kernel [32] with smoothness parameter $\nu = 5/2$ and length-scale $\ell > 0$. Such a kernel only depends on the relative distance between the coordinates $r \equiv |x - x'|$ and can be written in closed form as follows:

$$\mathrm{K}(x,x') \equiv \mathrm{k}_{\nu=5/2,\ell}(r) = \left(1 + \frac{\sqrt{5}\,r}{\ell} + \frac{5r^2}{3\ell^2}\right)\exp\left(-\frac{\sqrt{5}\,r}{\ell}\right). \tag{19}$$

In our analysis we also explored the values $\nu = 3$ and $\nu = 7/2$. The more general definition for the Matérn kernel is shown in Sec. III of the supplemental material [26]. The optimal value for the correlation length ℓ is chosen by a cross-validation scheme outlined in Sec.IV and Sec.V of the supplemental material [26].

F. Posterior sampling and inference

Our objective is now to obtain a reasonable sized distribution of values for the coefficients ξ_j that will allow us to report a distribution for the infered proton radius r_p. Given the complex nature of the model space associated with the allowed values of ξ, an analytic expression of M_ξ is not available. However, we show in the Appendix A that M_ξ does not depend on the unknown parameter τ. Hence, provided Γ is fixed, one can exploit this fact and use a Markov Chain Monte Carlo (MCMC) algorithm to sample the posterior distribution. The model along with priors on various components are represented in a hierarchical fashion as follows:

$$\begin{aligned} &Y \mid \xi, \sigma^2, \tau^2 \sim \mathcal{N}_n(\Phi\xi, \sigma^2 \mathrm{I}_n), \\ &\xi \sim \mathcal{N}_{N+2}(\xi; 0, \tau^2\,\Gamma)\,\mathbb{1}_{\mathcal{C}_\xi}(\xi), \quad p(\tau^2) \propto \frac{1}{\tau^2}, \quad p(\sigma^2) \propto \frac{1}{\sigma^2}, \end{aligned} \tag{20}$$

in which we have made the common non-informative prior choice for the observational noise standard deviation σ^2. For the hierarchical model above, the joint posterior distribution of the model parameters is given by:

$$P(\xi, \tau^2, \sigma^2 \mid Y) \propto \left\{(\sigma^2)^{-n/2}\,e^{-\frac{\|Y-\Phi\xi\|^2}{2\sigma^2}}\right\}\left\{(\tau^2)^{-(N+2)/2} e^{-\xi^{\mathrm{T}}\Gamma^{-1}\xi/(2\tau^2)}\,\mathbb{1}_{\mathcal{C}_\xi}(\xi)\right\}\,(\tau^2)^{-1}\,(\sigma^2)^{-1}. \tag{21}$$

The final normalizing constant of the posterior distribution is intractable and hence we resort to MCMC [31] to sample from the posterior distribution of the model parameters. More specifically, we use Gibbs sampling to iteratively sample from the full conditional distribution of (i) $\xi \mid \tau^2, \sigma^2, Y$ [2], (ii) $\tau^2 \mid \xi, \sigma^2, Y$, and (iii) $\sigma^2 \mid \xi, \tau^2, Y$. The conditional posterior of ξ in (i) is a truncated multivariate normal distribution which is sampled using the method proposed in [35]. The conditional posteriors of σ^2 and τ^2 in (ii) and (iii) are inverse-gamma (IG) distributions and hence easy to sample from. The details of the algorithm are provided in Sec. III of the supplemental material [26].

After discarding initial burn-in samples, let $\xi_j^{(1)}, \ldots, \xi_j^{(T)}$ be T successive iterate values of ξ_j from the Gibbs sampling algorithm, for $j = 2, \ldots, N+3$. Our point estimate for r_p based on the posterior samples is:

$$\widehat{r}_p = T^{-1} \sum_{t=1}^{T} \frac{\sqrt{-6\xi_2^{(t)}}}{Q_{\max}}. \tag{22}$$

The 68% confidence interval (one sigma) for r_p is also computable from our sampling algorithm.

III. ELECTRON-SCATTERING DATA ANALYSIS

Prior to dealing with the real data set we analyzed pseudo data generated by a known dipole function to test the performance of our method and to observe the impact that the different hyperparameters and data range had on the extracted radius (see Sec.IV of the supplemental material [26]). In this section, we present our analysis of the real electron-proton scattering data obtained from Mainz [36–38].

We tried four different variants of our model by imposing or relaxing the different constraints on ξ in Eq. 11. These are:

1. **cGP:** denotes the proposed constrained GP model as described in Eq. (14). The curve is restricted to be convex and the value at $Q^2 = 0$ is fixed at 1 ($\xi_1 = 1$).

2. **c$_0$GP:** denotes the model in Eq. (14) with the only constraint being (5a), the value at zero ($\xi_1 = 1$). The parameters $\xi_2, \ldots, \xi_{N+3}$ are left unconstrained in this model and therefore the curve is not necessarily monotonic and convex.

3. **c$_1$GP:** denotes the model with only shape constraints (5b) and (5c), which implies that the function is non-increasing and convex, but the value at zero is not fixed (ξ_1 is left unconstrained).

4. **uGP:** denotes the completely unconstrained GP, all the parameters $\xi_1, \xi_2, \ldots, \xi_{N+3}$ are free.

We note that although the condition $G_E(0) = 1$ (5a) is ultimately related to the charge of the proton, systematic errors can have an appreciable impact on the fulfillment of this constraint in the experimental data. It has become a customary practice (see for example [21]) to represent the observed values as $f(Q^2) = n_0 G_E(Q^2)$, where n_0 is a floating normalization parameter, $f(Q^2)$ are the observed values and $G_E(Q^2)$ is the true proton form factor. We can identify in our framework the choice $n_0 = 1$ with the requirement that our model estimate for the form factor has the fixed value of 1 at $Q^2 = 0$ (cGP and c0GP). Instead, leaving n_0 as an adjustable parameter corresponds to the relaxation of the constraint at zero (c1GP and uGP).

We conducted the real analysis in two regimes: low $Q^2 < 1.36$ fm^{-2} (the first 500 data points) and high $Q^2 < 25.12$ fm^{-2} (the full data set). The low regime was chosen based on the results in the pseudo-data analysis in which we observed that in this range the models gave a more accurate estimate of the slope of the assumed Dipole function. On the other hand, even though in the high Q^2 regime we observed some biasing toward lower estimates of the slope, we considered also the full data analysis. It is well known that due to the difficulty of measuring the form factor for smaller values of the momentum, the experimental data might be significantly biased for $Q^2 \approx 0$ and also the noise structure could not satisfy the assumptions we made on the pseudo data analysis: it could not be independent and identically distributed and all the points might not share the same variance. Thus, incorporating the whole range of values could help the analysis to overcome that experimental bias. Finally, having the two extrema (low and high regimes) is beneficial for comparison.

The analysis started with conducting pilot experiments with subsets of the data of size $n = 250$ randomly selected from the range of the potential values (Q^2) for the high regime, and with the full 500 points in the low regime. The

---

[2] Recall that in Bayesian notation $\xi \mid \tau^2, \sigma^2, Y$ means the conditional posterior distribution of ξ given τ^2, σ^2, and Y.

pilot experiments provided us with a better idea of the roles of the different hyperparameters of our model in the real data set, N, ℓ and ν, before eventually analyzing it. Recall that the Q^2 values are rescaled to $[0, 1]$ before the analysis. Overall we used 500 MCMC iterations after discarding a burning of 100 samples to form the posterior summary estimates of the radius.

TABLE I. High and low regime posterior estimates of the radius and one standard deviation (Δr_p) for cGP, c$_0$GP, c$_1$GP and uGP with $N = \{n/4, n\}$ and $\nu = 2.5$.

		High Regime n=(1422)		Low Regime (n=500)	
Model	N	r_p (fm)	Δr_p (fm)	r_p (fm)	Δr_p (fm)
cGP	n/4	0.844	0.002	0.853	0.002
	n	0.845	0.001	0.855	0.002
c$_0$GP	n/4	0.836	0.005	0.841	0.008
	n	0.845	0.004	0.840	0.006
c$_1$GP	n/4	0.842	0.004	0.873	0.006
	n	0.831	0.003	0.872	0.005
uGP	n/4	0.847	0.011	0.857	0.017
	n	0.858	0.008	0.862	0.01

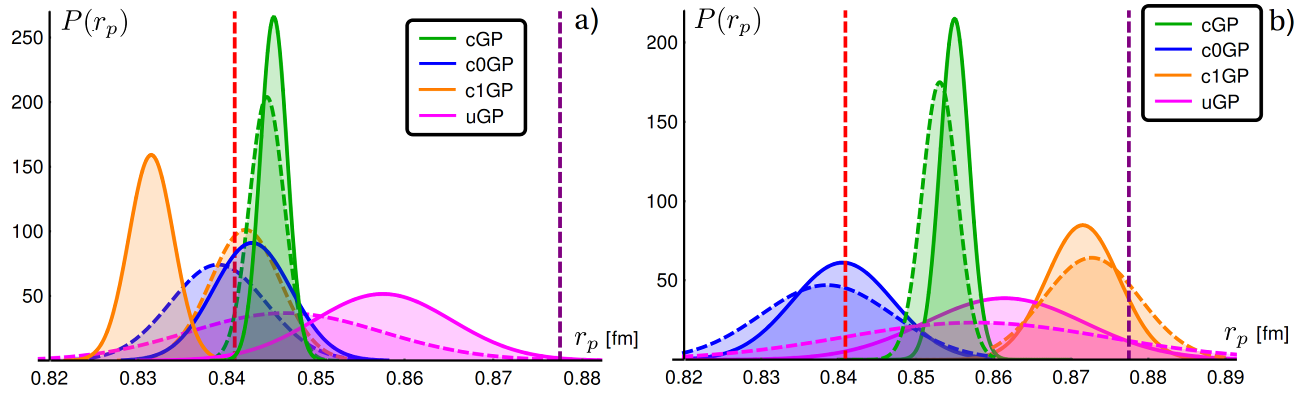

FIG. 2. Estimated density plots of MCMC samples of radius r_p for cGP, c$_0$GP, c$_1$GP and uGP with $N = n/4$ (dotted line), n (solid line) and $\nu = 2.5$ for the high Q^2 regime (a) and for the low Q^2 regime (b). The vertical dashed lines stand for the muonic result of 0.84 fm (Red) and the recommended CODATA value of 0.88 fm (Purple).

We conducted a cross validation procedure to select the optimal scale-length parameter ℓ for each regime, the details of which are shown in Sec.V of the supplemental material [26]. Our analysis guided us to choose $\ell_{\text{opt}} = 0.5$ for the full data set and $\ell_{\text{opt}} = 10$ on the low Q^2 set.

Having chosen the correlation length we performed the MCMC iterations for the four models, selecting the number of grid points $N = n/4$ and $N = n$ in order to compare results. Table I shows the posterior medians of r_p of the four models and the one sigma value in the high and low regime respectively for $\nu = 2.5$ while Fig. 2 shows the density plots (posterior distribution $P(r_p)$) in the high regime (a) and low regime (b).

We observed that, as a general trend, as the constraints are removed the model becomes more sensitive to the data range used and the choice of the number of grid points N and ν (the analysis and results for $\nu = 3$ and $\nu = 3.5$ are given in the supplemental material Sec. V [26]). For example, in the high regime for all values of N and ν cGP estimations of the radius are in all cases around 0.843 fm, while on the other extremum the unconstrained model uGP estimations range between 0.76 (for $\nu = 3$) and 0.86 fm. Incorporating the constraints also strongly affects the one sigma deviation of each model: cGP one sigma intervals were between 0.001 and 0.005 fm wide, while uGP one sigma intervals were as wide as 0.014 fm.

For the choice $\nu = 2.5$ we can see in Figure 2 that as N increases the estimate of c$_1$GP moves to a lower value of r_p while the estimates of all the other models increase to a higher value of r_p. This effect is less prominent in the low regime and overall cGP is the most robust with respect to changing the number of grid points. In all the cases, as N increases the variability in the estimation reduces (the estimated σ is slightly smaller than those for $N = n/4$), giving more precise results. As we observed in Table I, in going from high regime to low regime all the models, with the

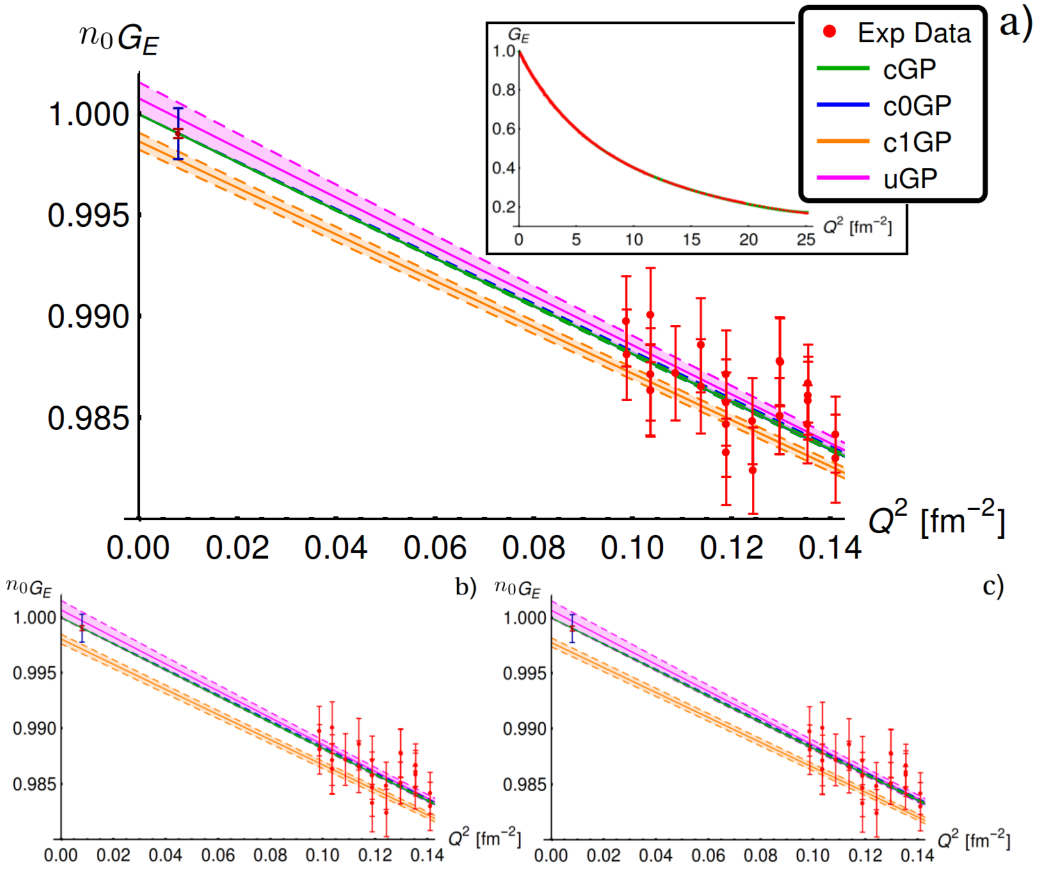

FIG. 3. Function fit with $\nu = 2.5$ (a), 3 (b), 3.5 (c) and $N = n$ in the high regime. The inset plot in (a) shows the overall fit of the models for $\nu = 2.5$ to the entire data range. The solid curves denote the model predictions while the shaded intervals bounded by dotted lines represent the 95% confidence intervals for the predictions. The red dots denote the experimental data obtained from Mainz with their respective error bars. The red and blue points near the origin at $Q^2 = 0.008$ fm^{-2} represent the lower value the new PRad experiment will be able to measure, with two different estimates for the projected uncertainty [17] and arbitrary $G_E(Q^2)$ value.

exception of c0GP, gave a larger estimate of the radius, c1GP being the one that showed the biggest change. uGP is the only model that includes both 0.84 and 0.88 fm in its support in both regimes.

We show in the supplemental material Sec.V [26] more in detail each individual posterior histogram of the MCMC samples from GP models for the radius and the MCMC samples of $n_0 G_E(0)$ from c$_1$GP and uGP. Recall that n_0 is defined as a floating normalization factor, while $G_E(0) = 1$ is a guaranteed property by the definition of G_E. We observed that the sample centers of $n_0 G_E(0)$ deviate from 1 by a very small amount ($|n_0 G_E(0) - 1| \lesssim 0.0014$) for both models in both regimes. It is remarkable how such a small deviation in the case of c$_1$GP can make such drastic changes when r_p results are compared with the fully constrained model cGP. For example, in the low regime for $N = n/4$ cGP estimates $r_p = 0.853$ fm while c1GP, having a value of 1.0014 at zero, estimates $r_p = 0.873$, a result that highlights the impact that a floating normalization can have on the extraction of the radius.

Fig. 3 and 4 show the function fits for the high and low regime respectively, with $\nu = 2.5$, 3, 3.5 for $N = n$. The overall fit is good for all the methods in both regimes, the real differences appear as $Q^2 \to 0$. For this reason we show the full fit in each regime only for $\nu = 2.5$ in the inset of the respective top plot, the full fits for the other values of ν being visually indistinguishable.

Overall we found relatively small variability in the function fits across different values of ν in both regimes, not enough to change the estimation of the radius by more than 0.01 fm within any of the models. Due to the constraint at the origin, both posterior medians of cGP and c0GP agree as $Q^2 \to 0$ with very narrow credible intervals, while c$_1$GP and uGP are either below or above and start going close to the other GP models estimates as Q^2 grows. As expected, the shape constraints help reduce the variability of the models, which is evidenced by the smaller credible intervals of c1GP in comparison with uGP, specially in the low regime. In the low regime, it seems that without the

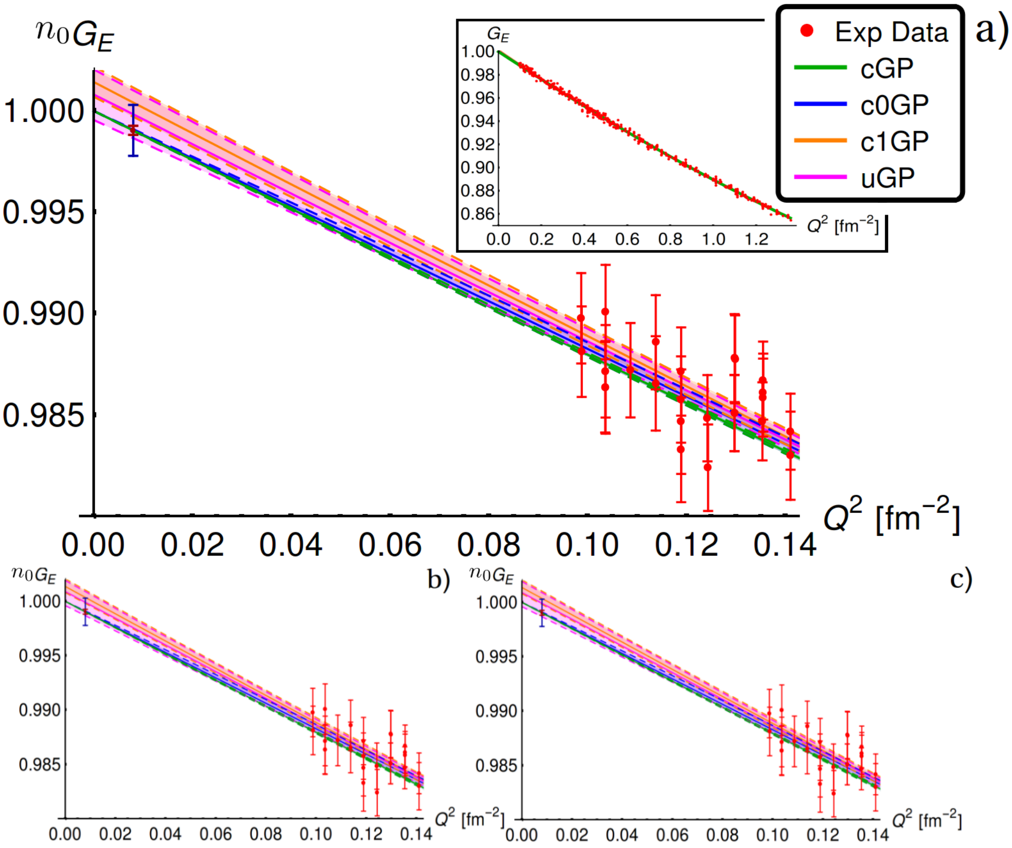

FIG. 4. Function fit with $\nu = 2.5$ (a), 3 (b) 3.5 (c) and $N = n$ in the low regime. The inset plot of (a) shows the overall fit of the models for $\nu = 2.5$ to the entire data range. The solid curves denote the model predictions while the shaded intervals bounded by dotted lines represent the 95% confidence intervals for the predictions. The red dots denote the experimental data obtained from Mainz with their respective error bars. The red and blue points near the origin at $Q^2 = 0.008$ fm^{-2} represent the lower value the new PRad experiment will be able to measure, with two different estimates for the projected uncertainty [17] and arbitrary $G_E(Q^2)$ value.

normalization constraint the extrapolations are likely to attain values at $Q^2 = 0$ larger than 1, which in turn pushes the estimate of the radius to larger values, as can be also seen in Fig. 2. In the low regime, as a general trend, we observed wider credible intervals for all the models.

The blue and red points near $Q^2 = 0.008$ fm^{-2} displayed in Fig. 3 and 4 for an arbitrary $G_E(Q^2)$ value represent the lowest momentum that will be measured by the new PRad experiment [17]. The blue and red error bars are two different estimates of the projected uncertainty the measurement will have. In the case of our proposed model, it seems that the blue uncertainty could allow us to discard either c1GP or uGP, while the red uncertainty would allow us to discard up to three of the model selected, clearly imposing a stringent constraint in the final estimation of the radius.

In summary, the main result of the present work is the extraction of a proton charge radius of $r_p = 0.845 \pm 0.001$ fm, consistent with the muonic Lamb-shift experiment. This result was obtained by enforcing both the normalization and shape constraints (see Eq.(6)). However, we caution the reader that the extrapolation to zero momentum transfer is subtle and sensitive to these constraints as well as to the Q^2 range adopted in the analysis. A more exhaustive analysis on the role played by the underlying assumptions in an effort to provide more robust predictions is left for a future work.

IV. CONCLUSIONS

The charge radius of the proton is a fundamental parameter that has attracted enormous attention during the last decade because of a discrepancy between two experimental methods. The value of the charge radius of the proton $r_p = 0.84087(39)\,\mathrm{fm}$ determined from muonic hydrogen [9, 10] differs significantly (by $\sim$4% or nearly 7σ) from the recommended CODATA value of $r_p = 0.8775(51)\,\mathrm{fm}$ obtained from decades of experiments in electron scattering and atomic spectroscopy. Many possible solutions to the "proton puzzle" have been proposed ranging from errors in the experimental data or in its interpretation all the way to new physics associated to a violation in lepton universality. Within this wide context our contribution is rather modest. In our view, the proton puzzle lays not in the experimental data, but rather in the extraction of the proton radius from the scattering data. To extract the charge radius from the electron scattering data set, one must extrapolate from the measured values of the electric form factor at a finite momentum transfer Q^2 all the way to $Q^2 = 0$. How to properly extrapolate to $Q^2 = 0$ has been the source of much controversy and innumerable debates. Many of these debates center around the optimal functional form (*e.g.*, monopole, dipole, polynomial, Padé, *etc.*) that should be adopted to carry out the extrapolation and on how best to determine the parameters associated to such functions. In this paper we also sought for an optimal extraction of the proton radius from the scattering data. However, in contrast to most of these approaches and in an effort to eliminate any reliance on specific functional forms, we have introduced a nonparametric method that does not assume any particular functional form for the form factor. Rather, we adopted a method that is flexible enough to "let the data speak for itself" and that solely relies on two physical constraints inherent to the form factor: (a) $G_E(Q^2\!=\!0)\!=\!1$ and (b) $G_E(Q^2)$ is a monotonically decreasing function of the momentum transfer. Note that this last constraint implies that $G'_E(Q^2)\!<\!0$ and $G''_E(Q^2)\!>\!0$ for all values of Q^2. These constraints are adopted in our study and their individual effects on the estimation of r_p are explored. The modeled form factor was expanded in terms of a suitable set of basis functions with coefficients restricted exclusively by the shape constraints. To determine the optimal coefficients, the experimental data was divided into two Q^2 regions: (i) low $Q^2 \leq 1.36\ \mathrm{fm}^{-2}$ and (ii) high $Q^2 \leq 25.12\ \mathrm{fm}^{-2}$. For each of these regions, the optimal hyperparameters –the correlation length ℓ, the smoothness parameter ν, and the number of grid points N– were obtained by monitoring the performance of the algorithm against the 20% of the data that was left out from the calibration, an analysis we present in the supplemental material Sec.V [26]. The actual implementation of the algorithm was carried out via MCMC sampling of the posterior distribution using Bayesian inference.

To test the robustness and reliability of the approach we started by confronting our results against (known) synthetically-generated data with random Gaussian errors in low, medium and high regime (the complete analysis is presented in the supplemental material Sec.IV [26]). For the case in which both shape constraints were incorporated (labeled in the main text as cGP) we obtained an accurate and precise determination of the proton radius in both the low and medium Q^2 regions. In the high Q^2 region where the entire synthetic data set was used, we observed a systematic shift towards lower values of the (known) radius. We believe that this issue may be associated to the method chosen to determine the hyperparameters. We plan to devote more attention to this matter in a future work.

In the case of the real experimental data from Mainz, we also found that the extraction of the proton radius is sensitive to the range of Q^2 values considered in the analysis. In the case of the high Q^2 region where the entire experimental data set is incorporated, the CODATA value of $r_p = 0.878\,\mathrm{fm}$ is disfavored regardless of the adopted constraints. If both constraints are incorporated (cGP) we extract a charge radius of $r_p = 0.845 \pm 0.001\,\mathrm{fm}$, which is one of the central results of this work. The value is even lower if we assume a floating normalization ($\mathrm{c_1GP}$): $r_p\!=\!0.831 \pm 0.003$ fm. We note that we also considered a scenario of largely academic interest in which no constraints were incorporated. As expected, the unconstrained model (uGP) returned posterior distributions that were wide enough to be consistent with both the muonic hydrogen and CODATA values. We conclude that if the entire Mainz data set is included, our analysis favors the smaller value of the proton radius, as suggested by the muonic Lamb shift.

However, if the low Q^2 region is used to inform the posterior distribution, we obtained mixed results. First, when both shape constraints are included, we obtain a proton radius of $r_p = 0.855 \pm 0.002$ fm—that falls almost in the middle of the two experimental values. If now one of the constraints is removed the behavior is radically different. Removing the normalization constraint in favor of a floating normalization ($\mathrm{c_1GP}$) shifts the posterior distribution to a large enough value of r_p to make it consistent with the CODATA estimate. Note that the value at zero of c1GP is 1.0014, not far away from 1, and yet that is enough to produce a radius 0.02 fm bigger than the fully constrained model cGP. In contrast, leaving the normalization fixed at $G_E(Q^2\!=\!0)\!=\!1$ but relaxing the demand for $G_E(Q^2)$ to be a monotonically decreasing function of Q^2 results in a value for r_p consistent with muonic result. In this regard, we anticipate that the PRad analysis will play a critical role in helping resolve this ambiguity. However, based solely on the present analysis focused on the low Q^2 region (where the behavior of the form factor is nearly linear) our results are inconclusive as far as resolving the proton puzzle.

In the future, we propose to improve our model in order to overcome a possible bias in the analysis of the high Q^2 region, an objective that could be accomplished by developing a better procedure for estimating the hyperparameters.

As this technique is still in development, we would like to test it on more synthetic data sets, similar in spirit to the framework developed by Yan et al [21]. We trust that lessons learned from their project will help us improve the robustness of our nonparametric model.

Yet, even if the resolution of the proton puzzle is found elsewhere, the advances along this direction would have not been in vain. The proton puzzle as well as many other developments have allowed us to realize the importance of *enhancing the interaction between nuclear experiment and theory through information and statistics* [39]. We are entering into a new era in which statistical insights will become essential and uncertainty quantification will be demanded.

ACKNOWLEDGMENTS

We are enormously grateful to Prof. Douglas Higinbotham for his unconditional help, guidance and lightning fast email responses. This material is based upon work supported by the U.S. Department of Energy Office of Science, Office of Nuclear Physics Awards Number DE-FG02-92ER40750. Dr. Bhattacharya acknowledges NSF CAREER (DMS 1653404), NSF DMS 1613156 and National Cancer Institute's R01 CA 158113, and Dr. Pati acknowledges NSF DMS 1613156 for supporting this research.

Appendix A: Theoretical guarantees for the constraints on f_ξ and on the independence of $\mathcal{C}_\xi$ on τ

In this section we show the equivalence between the shape constraints on our model f_ξ and the inequalities on the coefficients ξ. Denote by $\mathcal{C}_f$ the function subspace of all the f_ξ defined in Eq (9) that obey the required constraints: $f_\xi(0) = 1$, $f'_\xi(x) < 0$ and $f''_\xi(x) > 0$.

We show below that the constraints that define $\mathcal{C}_f$ can be *equivalently* represented as linear restrictions on ξ. We state Proposition 1 which provides an explicit characterization of the stated linear constraints.

Proposition 1 *$f_\xi \in \mathcal{C}_f$ if and only if $\xi \in \mathcal{C}_\xi$.*

Recall $\mathcal{C}_\xi$ is defined as:

$$\mathcal{C}_\xi \equiv \left\{\xi \in \mathbb{R}^{N+3} :\ \xi_1 = 1,\ \xi_2 + \sum_{j=0}^{N} c_j\ \xi_{j+3} \leq 0,\ \xi_{j+3} \geq 0,\ j = 0, \ldots, N\right\}, \tag{A1}$$

Proof 1 *We first check the convexity constraint, by taking second order derivative we have $f''_\xi(x) = \sum_{j=0}^{N} \xi_{j+3} h_j(x)$, by the non-negativity of h_j for all $x \in [0,1]$ and any $j = 0, \ldots, N$, the set $\{f''_\xi(x) \geq 0, \forall x \in [0,1]\}$ is equivalent to $\{\xi_{j+3} \geq 0, j = 0, \ldots, N\}$. To impose the non-increasing constraint, we need to check the following:*

$$f'_\xi(x) = \xi_2 + \sum_{j=0}^{N} \xi_{j+3}\psi_j(x) \leq 0, \forall x \in [0,1].$$

Observe that this is equivalent to

$$\xi_2 \leq -\max_{x\in[0,1]} \left(\sum_{j=0}^{N} \xi_{j+3}\psi_j(x)\right) = -\sum_{j=0}^{N} c_j \xi_{j+3}. \tag{A2}$$

(A2) *follows since ψ_j defined in (7) is a non-decreasing function of x and $\max_{x\in[0,1]} \psi_j(x) = \psi_j(1) =: c_j$ for $j = 0, \ldots, N$. This concludes the proof of the proposition.* ♢

Now we proceed to show why he normalizing constant M_ξ of the truncated prior distribution of ξ is independent of τ.

Proposition 2 *The normalizing constant M_ξ associated with the truncated prior distribution of ξ is a constant in $[0,1]$ that does not depend on τ^2.*

Proof 2 *By definition*

$$M_\xi = \int_{C_\xi} (\tau^2)^{-(N+2)/2} (|\Gamma|)^{(-1/2)} e^{-\frac{1}{2\tau^2}\xi^T\Gamma^{-1}\xi} d\xi.$$

By change of variable $\xi' = \xi/\tau$, observe that the truncated region $C_{\xi'}$ is the same as C_ξ as long as $\tau > 0$. Hence, $M_\xi \in [0,1]$ does not depend on τ. ♢

---

[1] *Reaching for the Horizon; The 2015 Long Range Plan for Nuclear Science* (2015).
[2] C. F. von Weizsäcker, Z. Physik **96**, 431 (1935).
[3] H. A. Bethe and R. F. Bacher, Rev. Mod. Phys. **8**, 82 (1936).
[4] P. J. Mohr, D. B. Newell, and B. N. Taylor, Rev. Mod. Phys. **88**, 035009 (2016).
[5] R. Hofstadter, Rev. Mod. Phys. **28**, 214 (1956).
[6] H. De Vries, C. W. De Jager, and C. De Vries, Atom. Data Nucl. Data Tabl. **36**, 495 (1987).
[7] G. Fricke, C. Bernhardt, K. Heilig, L. A. Schaller, L. Schellenberg, E. B. Shera, and C. W. de Jager, Atom. Data and Nucl. Data Tables **60**, 177 (1995).
[8] I. Angeli and K. Marinova, At. Data Nucl. Data Tables **99**, 69 (2013).
[9] R. Pohl *et al.*, Nature **466**, 213 (2010).
[10] R. Pohl, R. Gilman, G. A. Miller, and K. Pachucki, Ann. Rev. Nucl. Part. Sci. **63**, 175 (2013).

[11] J. C. Bernauer and R. Pohl, Sci. Am. **310**, 18 (2014).
[12] C. E. Carlson, Prog. Part. Nucl. Phys. **82**, 59 (2015).
[13] A. Beyer *et al.*, Science **358**, 79 (2017).
[14] D. Robson, Int. J. Mod. Phys. **E23**, 1450090 (2015).
[15] G. W. Bennett *et al.* (Muon g-2), Phys. Rev. **D73**, 072003 (2006).
[16] R. Aaij *et al.* (LHCb Collaboration), Phys. Rev. Lett. **113**, 151601 (2014).
[17] A. H. Gasparian (PRad), *Proceedings, 14th International Conference on Meson-Nucleon Physics and the Structure of the Nucleon (MENU 2016): Kyoto, Japan, July 25-30, 2016*, JPS Conf. Proc. **13**, 020052 (2017).
[18] J. C. Bernauer *et al.* (A1), Phys. Rev. Lett. **105**, 242001 (2010).
[19] R. Gilman *et al.* (MUSE), (2013), arXiv:1303.2160 [nucl-ex].
[20] D. W. Higinbotham, A. A. Kabir, V. Lin, D. Meekins, B. Norum, and B. Sawatzky, Phys. Rev. **C93**, 055207 (2016).
[21] X. Yan, D. W. Higinbotham, D. Dutta, H. Gao, A. Gasparian, M. A. Khandaker, N. Liyanage, E. Pasyuk, C. Peng, and W. Xiong, (2018), arXiv:1803.01629 [nucl-ex].
[22] J. C. Bernauer and M. O. Distler, in *ECT* Workshop on The Proton Radius Puzzle Trento, Italy, June 20-24, 2016* (2016) arXiv:1606.02159 [nucl-th].
[23] R. J. Hill and G. Paz, Physical Review D **82**, 113005 (2010).
[24] G. Lee, J. R. Arrington, and R. J. Hill, Physical Review D **92**, 013013 (2015).
[25] Data files with the data from the Mainz collaboration [36–38] obtained from D. Higinbotham in a private communication.
[26] See Supplemental Material at [URL will be inserted by publisher].
[27] G. A. Miller, Physical Review C **99**, 035202 (2019).
[28] J. Alarcón and C. Weiss, arXiv preprint arXiv:1803.09748 (2018).
[29] J. Alarcón and C. Weiss, Physical Review C **97**, 055203 (2018).
[30] H. Maatouk and X. Bay, Mathematical Geosciences **49**, 557 (2017).
[31] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian data analysis* (CRC press, 2013).
[32] C. Rasmussen and C. Williams, *Gaussian Processes for Machine Learning*, Adaptative computation and machine learning series (University Press Group Limited, 2006).
[33] R. J. Adler, *The geometry of random fields*, Vol. 62 (Siam, 1981).
[34] H. Jeffreys, Proc. R. Soc. Lond. A **186**, 453 (1946).
[35] Z. Botev, Journal of the Royal Statistical Society: Series B (Statistical Methodology) **79**, 125 (2017).
[36] J. C. Bernauer and A. Collaboration, in *AIP Conference Proceedings*, Vol. 1388 (AIP, 2011) pp. 128–134.
[37] J. Bernauer, M. Distler, J. Friedrich, T. Walcher, P. Achenbach, C. A. Gayoso, R. Böhm, D. Bosnar, L. Debenjak, L. Doria, *et al.*, Physical Review C **90**, 015206 (2014).
[38] J. Bernauer, P. Achenbach, C. A. Gayoso, R. Böhm, D. Bosnar, L. Debenjak, M. Distler, L. Doria, A. Esser, H. Fonvieille, *et al.*, Physical Review Letters **107**, 119102 (2011).
[39] D. G. Ireland and W. Nazarewicz, J. Phys. **G42**, 030301 (2015).