Novel approach to the global analysis of proton form factors in elastic electron-proton scattering

C. McRae and P. G. Blunden ⬤

Department of Physics and Astronomy, University of Manitoba, Winnipeg, Manitoba, Canada R3T 2N2

We present a method for the global analysis of elastic electron-proton scattering when combining data sets from experiments with different overall normalization uncertainties. The method is a modification of one employed by the NNPDF collaboration in the fitting of parton distribution data. This method is an alternative to the 'penalty trick' method traditionally employed in global fits to proton electric and magnetic form factors, while avoiding the statistical biases inherent in that approach. We discuss issues that arise when extending the method to nonlinear models. For data with $Q^2 > 1\,\mathrm{GeV}^2$ we find relatively minor differences to traditional model fits when the normalization uncertainties from different experiments are correctly accounted for. We discuss implications of this method for the well-known discrepancy between the form factor ratio G_E/G_M extracted from the Rosenbluth and polarization transfer techniques.

I. INTRODUCTION

Elastic electron-proton scattering can be described by two functions of the squared four-momentum transfer Q^2: the Sachs electric and magnetic form factors $G_E(Q^2)$ and $G_M(Q^2)$, respectively. Experimentally there are two main techniques to determine the form factors: a longitudinal-transverse (LT) separation of unpolarized cross section data [1] and polarization transfer (PT) measurements of the 'polarization ratio' G_E/G_M. After taking into account standard radiative corrections, separate fits of the form factors to these data lead to fits that disagree significantly with one another as Q^2 increases. Model-dependent two-photon exchange (TPE) corrections to LT data are considered the most viable explanation for this discrepancy [2,3], but definitive experimental evidence at high Q^2 is lacking.

Under the one-photon exchange (OPE) assumption the differential cross section for elastic electron-proton scattering is proportional to the reduced cross section

$$\sigma_{\mathrm{red}} = \varepsilon\, G_E^2(Q^2) + \tau\, G_M^2(Q^2), \tag{1}$$

where $\tau = Q^2/(4M^2)$ and $\varepsilon^{-1} = 1 + 2(1 + \tau)\tan^2\theta/2$. Here, θ is the scattering angle in the laboratory frame, M is the proton mass, and the electron is taken to be massless.

The traditional Rosenbluth LT separation [1] of G_E and G_M uses measurements of σ_{red} at fixed Q^2 with a linear fit to the virtual photon polarization ε by varying the scattering angle. This produces form factors that are model-independent but are of limited range in beam energy and scattering angle. Alternatively, one can make a global fit of models of G_E and G_M to cross section data over a wide range of kinematical conditions [4]. This second method often requires combining data sets from different experiments, or different experimental setups within the same experiment, with different absolute normalizations and associated uncertainties.

A key question, which we address in this paper, is how to combine such measurements from different data sets in a statistically rigorous and meaningful way. To do so we introduce a method based on a modification of one employed by the NNPDF collaboration [5], which we call the iterated model fit (IMF). As a proof of concept, we compare this method to two recent analyses of cross section data that use the traditional 'penalty trick' method, which we will describe below. The first analysis is that of Gramolin and Nikolenko [6], who use a linear parametrization for σ_{red} in the model functions G_E^2 and G_M^2. They use three experimental data sets [7,8] with Q^2 in the range from 1 to 8.83 GeV^2. The second analysis is by the GMp12 collaboration [9], who use the same three data sets plus six additional data sets, including their own new precision data with Q^2 up to 15.75 GeV^2. They use a more flexible parametrization of model functions for G_M and $(\mu_p G_E/G_M)^2$, which leads to a nonlinear model for σ_{red}. We discuss how the IMF method can be adapted to such a nonlinear fitting procedure.

References [6,9] both apply updated and improved radiative corrections to the original experimental data, but do not include TPE corrections. Both groups still find significant disagreement between separate fits of the cross section and PT data. As it is the intention of this paper to compare methodologies rather than construct a state-of-the-art global fit to G_E and G_M, we leave it to future work to incorporate TPE contributions as well as cross section data with $Q^2 < 1\,\mathrm{GeV}^2$.

II. FITTING WITH NORMALIZATION UNCERTAINTY

The technique of fitting via minimization of a chi-square emerges out of the more fundamental desire to find the parameters of a model that, assuming the model is correct, are most likely to have produced the observed data (a maximum likelihood estimate). Thus if one is not careful when modifying the chi-square to be minimized, one will violate the underlying assumption that measurements are normally distributed (the chi-square is by definition the exponent of a Gaussian likelihood), leading to undesirable statistical biases in the fitted models. Of course, the Gaussian prior comes from a maximum entropy argument, namely that if one only knows the mean and variance of the data, and one believes the model is correct, then the underlying prior distribution that assumes the least about the data will be a Gaussian distribution. If one has reason to modify the prior distribution, then assuming an underlying Gaussian likelihood will also introduce a bias into the fitting procedure.

The standard method for dealing with normalization uncertainty is fitting via a modification of the traditional chi-square known colloquially as the 'penalty trick',

$$\chi_{\mathrm{p}}^2 = \sum_{i=1}^{\mathcal{N}} \left[\frac{(n_i - 1)^2}{(\Delta n_i)^2} + \sum_{j=1}^{N_i} \frac{(y_{ij} - \mathrm{M}_{ij}/n_i)^2}{(\Delta y_{ij})^2} \right]. \quad (2)$$

Here, a model M (in our case $\mathrm{M} = \sigma_{\mathrm{red}}$) is fitted, alongside floating normalizations (n_i), to data (y_{ij}) with both additive (Δy_{ij}) and multiplicative (Δn_i) uncertainties; i indexes the $\mathcal{N}$ experiments, and j indexes the N_i data points of each experiment (in more naive approaches, one may scale only the data by the n_i). This method fails to be Gaussian in the normalization parameters n_i, so we seek a more sound alternative.

In a seminal paper by Ball *et al.* [5], the NNPDF collaboration firstly demonstrate the statistical biases in the usual approaches for combining multiple experimental data sets with multiplicative uncertainties, and second, for fitting a single parameter, construct a Monte Carlo method that takes the normalization uncertainty into account without compromising the integrity of the fit. The most important steps in the derivation of the method are as follows. First, perform the usual error propagation for multiplying data with uncertainty:

$$(y_{ij} \pm \Delta y_{ij})(1 \pm \Delta n_i) = y_{ij} \pm \sqrt{(\Delta y_{ij})^2 + (y_{ij}\Delta n_i)^2}. \quad (3)$$

Second, recognizing that a well fitted model will resemble the data; in the chi-square of that error-propagated data, replace the instance of the data in the propagated error with a best guess for the model, $\hat{\mathrm{M}}$:

$$\chi_{t_0}^2 = \sum_{i=1}^{\mathcal{N}} \left[\sum_{j=1}^{N_i} \frac{(y_{ij} - \mathrm{M}_{ij})^2}{(\Delta y_{ij})^2 + (\hat{\mathrm{M}}_{ij}\Delta n_i)^2} \right]. \quad (4)$$

This $\chi_{t_0}^2$ has all the nice statistical properties we expect, and its likelihood remains Gaussian in both model and data. Therefore, if our assumption of a Gaussian prior is correct, the only source of bias in our fits will come from the explicit model-dependence. If our best guess coincides with the model

given from the optimization, the replacements are justified and we are done. Else, we use the optimized model as our ensuing guess, iterating until the model converges. This is the t_0 method of Ref. [5], which focuses on fitting single parameters at specific kinematics.

When one attempts to apply the t_0 method to data at multiple kinematics, it is ambiguous which kinematics the best-guess model should be evaluated at for off-diagonal elements of the constructed covariance matrix. We propose to generalize the method by appending the original data covariance matrix cov_i as follows for each of the i experiments:

$$\mathrm{cov}_{i,jk}^{\mathrm{IMF}} = \mathrm{cov}_{i,jk} + (\Delta n_i)^2 \hat{\mathrm{M}}_{ij}\hat{\mathrm{M}}_{ik}. \quad (5)$$

This generalizes the case of a single data point per experiment and of all data points being at the same kinematics [5]. With this, the remaining step used by NNPDF is left unaltered, and replica data are generated via a two step Monte Carlo procedure. First, for each datum one generates replica data y_{ij}^R by pulling from a normal distribution characterized by the point's central value and uncertainty. Second, a normalization factor n_i^R is pulled from a normal distribution centered around 1 with a standard deviation defined by the normalization uncertainty for the datum's experiment. Thus the final chi-square to minimize during our Monte Carlo replica iterated model fit is given in matrix form by

$$\chi_{\mathrm{IMF}}^2 = \sum_{i=1}^{\mathcal{N}} (n_i^R y_i^R - \mathrm{M}_i)^\mathsf{T} (\mathrm{cov}_i^{\mathrm{IMF}})^{-1} (n_i^R y_i^R - \mathrm{M}_i). \quad (6)$$

For $\mathcal{R}$ replica's generated we will have $\mathcal{R}$ best fit models and $\mathcal{R}$ sets of best fit parameters. These fits are averaged to produce an updated best fit model $\hat{\mathrm{M}}$ of the original data, which is then fed back into Eq. (5). This process iterates until we find convergence of the output model.

A. IMF vs. penalty trick (linear model)

Here, we compare fits via the penalty trick method to fits via the IMF method to understand how the ubiquitous method compares to the statistically rigorous method. Gramolin and Nikolenko [6] use a simple linear model for the square of the form factors to ensure the fitting procedure becomes a linear algebra problem in the parameters (intentionally utilizing the more naive version of the penalty trick). They take

$$G_E^2(Q^2; \boldsymbol{a}) = (1 - a_1\tau - a_2\tau^2 - a_3\tau^3)G_{\mathrm{dip}}^2,$$
$$G_M^2(Q^2; \boldsymbol{b})/\mu_p^2 = (1 - b_1\tau - b_2\tau^2 - b_3\tau^3)G_{\mathrm{dip}}^2, \quad (7)$$

where μ_p is the proton magnetic moment, and $G_{\mathrm{dip}}(Q^2)$ is a dipole form factor. The downside to using such a simple model is manifold. The squared form factors have nothing preventing them from becoming negative, the form factors do not fall off as predicted by quantum chromodynamics (QCD) scaling laws, and if one wishes to include PT data the fitting procedure becomes nonlinear. Nevertheless, it makes for a rather simple point of comparison. For the penalty trick method we solve the system of equations $\partial \chi_{\mathrm{p}}^2/\partial \alpha_i = 0$ for the nine parameters α_i, giving the results shown in Table I. The index of n_i refers to the Walker and Andivahis (8 GeV and 1.6 GeV) data sets itemized in Ref. [9].

TABLE I. Form factor parameters extracted from the penalty trick and iterated model fit (IMF) methods for the linear model of Eq. (7). We use the same three data sets as Ref. [6], but with updated radiative corrections from Ref. [9].

	Penalty trick	IMF
a_1	0.23 ± 0.22	0.17 ± 0.22
a_2	0.47 ± 0.44	0.55 ± 0.44
a_3	-0.35 ± 0.21	-0.38 ± 0.21
b_1	-0.407 ± 0.045	-0.413 ± 0.045
b_2	$0.373 \pm 0.045\$$	0.375 ± 0.046
b_3	-0.076 ± 0.013	-0.077 ± 0.013
n_1	1.008 ± 0.012	–
n_2	1.008 ± 0.012	–
n_3	0.963 ± 0.013	–
χ^2/d.o.f.	$28.3/39$	$28.1/42$

To implement the IMF method we begin with a guess of all parameters being 0, which sets the form factors to standard dipoles. Convergence required only two iterations of the IMF procedure with $\mathcal{R} = 1000$ replica data sets fitted at each iteration. The parameters in the column labeled IMF in Table I are the mean and standard deviation of the final six dimensional distribution of 1000 fitted parameters. In Fig. 1

FIG. 1. (a) The normalized $G_M^2(Q^2)$ for both the IMF method (solid line) and the penalty trick (dashed line), and (b) the squared form factor ratio $\mu_p^2 G_E^2/G_M^2$, plotted with PT data itemized in the Supplementary Material of Ref. [9] with $Q^2 > 1\,\text{GeV}^2$ (not included in the fit).

we can see plots of the G_M^2 and G_E^2/G_M^2 form factor fits, suitably normalized.

B. IMF vs. penalty trick (nonlinear model)

The application of the IMF method to nonlinear models creates a potential difficulty. Generically, the average of several functions that share the same functional form $F(x; \boldsymbol{\alpha})$ will only maintain that functional form when the function is linear in $\boldsymbol{\alpha}$. This ruins the iterative step because we do not have the same functional form at each iteration. There are two possible ways out of this. First, one could find the parameters $\boldsymbol{\alpha}$ for the chosen functional form that are closest to the average of the replica fits $\langle F_R(x) \rangle$ by minimizing the L^2 norm

$$\frac{\partial}{\partial \boldsymbol{\alpha}} \int_{x_{\min}}^{x_{\max}} dx \, (F(x, \boldsymbol{\alpha}) - \langle F_R(x) \rangle)^2 = \boldsymbol{0}. \tag{8}$$

This minimization is computationally expensive, but it ensures that the functions are similar as possible over the relevant range. Alternatively, one may simply keep taking average parameters and plugging them into the model, understanding that this is only an approximation for the aforementioned closest model to the average model, and hope for convergence, as convergence is all one needs for the results to be statistically sound. The second option works for the following model, and is used throughout the remainder of this paper.

A flexible parametrization for the form factors is the z-expansion model presented in Ref. [10]. The model is

$$G_E(z; \boldsymbol{a}) = \sum_{k=0}^{k_{\max}} a_k z^k, \quad G_M(z; \boldsymbol{b})/\mu_p = \sum_{k=0}^{k_{\max}} b_k z^k,$$

$$z(Q^2) = \frac{\sqrt{t_{\text{cut}} + Q^2} - \sqrt{t_{\text{cut}} - t_0}}{\sqrt{t_{\text{cut}} + Q^2} + \sqrt{t_{\text{cut}} - t_0}}. \tag{9}$$

In all fits we use the values $t_{\text{cut}} = 4m_\pi^2$ and $t_0 = t_{\text{cut}}(1 - \sqrt{1 + Q_{\max}^2/t_{\text{cut}}})$, but the fits are rather insensitive to these particular values. To ensure that $G_E(Q^2) \sim 1/Q^4$ as $Q^2 \to \infty$ there are constraints

$$\sum_{k=n}^{k_{\max}} \frac{k!}{(k-n)!} a_k = 0, \quad n = 0, 1, 2, 3, \tag{10}$$

and similarly for b_k. Additionally we require $G_E(0) = 1$ and $G_M(0) = \mu_p$. Choosing $k_{\max} = 8$ therefore leaves four free parameters for each form factor, which we designate to be those with $k = 0, 1, 2, 3$.

For the following fits we use the global cross section data from the GMp12 analysis [9] with $Q^2 < 16\,\text{GeV}^2$ (115 points, only six points are excluded). This compilation was built specifically from experiments for which past radiative corrections could be undone and then recalculated with updated corrections to ensure the analysis was consistent across all the data. The GMp12 analysis uses the penalty trick to account for the normalization uncertainty in their fits, but with a different parametrization than given in Eq. (9). Implementing the IMF method, and comparing against PT data, gives us another clean comparison between the fitting procedures.

TABLE II. Form factor parameters from the penalty trick and IMF methods for the nonlinear model of Eq. (9). Fitted normalizations for the nine data sets (not shown) range from 0.95 to 1.06, consistent with Ref. [9].

	Penalty trick	IMF
a_0	-0.847 ± 0.028	-0.847 ± 0.028
a_1	1.269 ± 0.065	1.269 ± 0.067
a_2	0.47 ± 0.34	0.45 ± 0.34
a_3	-2.36 ± 0.49	-2.31 ± 0.51
b_0	-0.821 ± 0.015	-0.822 ± 0.015
b_1	1.374 ± 0.058	1.377 ± 0.057
b_2	-0.513 ± 0.086	-0.516 ± 0.083
b_3	-1.082 ± 0.056	-1.085 ± 0.059
$\chi^2/\text{d.o.f.}$	$76.9/98$	$76.5/107$

Using the IMF method with 1500 replicas per iteration we find convergence within five iterations. In each iteration about 10% of replica form factor fits were discarded as they rapidly changed sign in the region $Q^2 < 1\,\text{GeV}^2$. This happens because the model is insensitive to the sign of the form factors, and we have not included any data in this region in the fit. The z-expansion model form factors for both the IMF and penalty trick methods are given in Table II and plotted in Fig. 2. These fits demonstrate that unless further PT data reveals an upturn, TPE corrections become increasingly important at large Q^2. The relative sizes of the error bands are not surprising. The Hessian method for approximating parameter errors in nonlin-
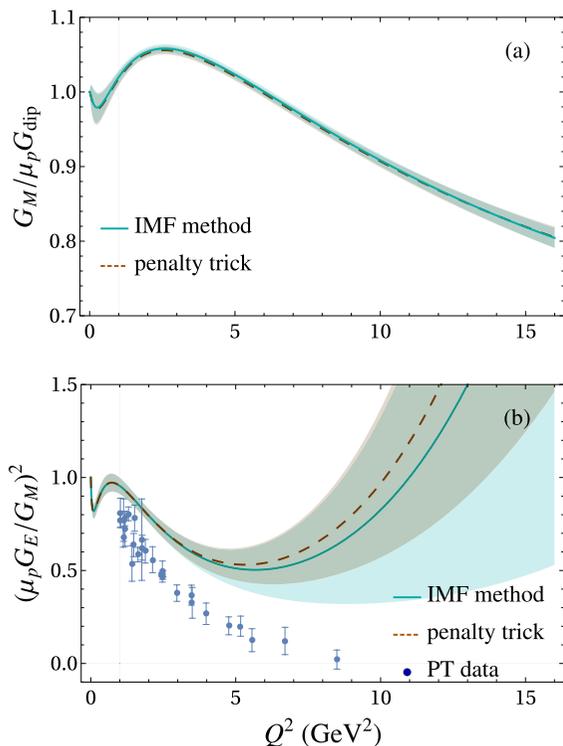
FIG. 2. The z-expansion model fits to the global cross section data from the nine data sets of Ref. [9] with $Q^2 < 16\,\text{GeV}^2$.

ear model fitting is known to be less accurate than bootstrap methods. We also note that the central value of the penalty trick fit pulls further away from the PT data than the IMF fit does. Thus it may be the case that the use of the penalty trick method accentuates the disagreement with the PT data.

III. THE IMPORTANCE OF CORRELATIONS

It is clear that the IMF method achieves fits of the same quality as the penalty trick fits without the need for the additional normalization parameters n_i, some of which differ substantially from 1. This demonstrates that these additional parameters are simply a means to approximate the correlated nature of the data, and are not physically meaningful on their own. Correlations enter via the normalization uncertainty due to the fact that if the true value of the normalization is varied the whole data set should vary in tandem.

This covariance of whole data sets requires extra attention when fitting said data, particularly when faced with the realities of fitting to data subsets. When fitting with only point-to-point errors, the addition or removal of a data point from the fit will only affect the fit locally near that point. However, when covariance is of the order of the point-to-point error, truncation and binning of the data set can significantly alter the best fit as the presence of covariance affects the fit function globally. In principle, the choice of inclusion or exclusion of a single data point can significantly tense or relax a fit. It is for this reason we do not perform LT separations invoking the IMF method, as separate linear fits to the relevant data subsets would be blind to the covariance of data at different Q^2 values.

With the same goal in mind, there is often a temptation to treat the normalization factors from the penalty trick as a pseudo-model-independent means to rescale data sets for subsequent analysis. This was done in Ref. [9] to facilitate an LT separation of the rescaled data, giving 'model-independent' form factors that could be compared to those obtained directly from the penalty trick fit. This approach ignores the other correlations between data, and therefore, in our view, does not achieve the goal of improving our knowledge of the true values of the form factors. Fitted normalizations are designed to pull the cross section data towards the model, and thus it is no surprise that an LT separation performed on rescaled data gives form factors that are very similar to the those obtained from the penalty trick fit.

All this furthers a point made by Bernauer *et al.* [4] that LT separations are not particularly helpful for fitting when compared to using cross section data. A global fit appears the only reasonable tool for form factor extraction when normalization uncertainty is present because the total covariance matrix can be incorporated during the fitting procedure. By contrast, an LT separation is forced to focus on only a subset of the data, ignoring said correlations.

IV. SUMMARY AND CONCLUSIONS

In this work a model-dependent but statistically unbiased global analysis of the proton's elastic form factors was presented. To perform these fits we introduced the IMF method, an extension of the statistically unbiased t_0 fitting method used

by the NNPDF collaboration [5]. We found that statistically unbiased fitting leads to minor improvements to fits of form factor models to the world cross section data without the need for additional normalization parameters for each data set.

The penalty trick, despite its flaws, seems to consistently provide similar fits of the proton's elastic form factors. This begs the question: why put in all the extra effort of the IMF method for a presumably small gain? D'Agostini [11] gives a real-world example of how the naive treatment of normalization uncertainties via the penalty trick led to a 'repulsion' of the fit from the data, while the chi-square per degree of freedom remained approximately constant. At the time this was remedied by updating the penalty trick to scale the uncertainties as well as the data, leading to Eq. (2). However, Ball *et al.* [5] show this version still fails to be free of statistical bias when combining multiple experimental data sets. This means there is nothing to prevent analogous misbehaviour of

fits with the updated penalty trick. Unfortunately, we know of no *a priori* criteria to determine when such misbehavior is likely to occur.

Both the low Q^2 global analysis by Bernauer *et al.* [4] and recent work on the proton radius [12] use the penalty trick in order to fit their global data sets. Encouraged by the success of the IMF method, we plan to reanalyze these fits to see if the statistically unbiased approach leads to any improvements in the extracted values and fits. Additionally, we plan to perform a combined global analysis of all readily available cross section and PT data, including TPE corrections, in a future work.

ACKNOWLEDGMENTS

[1] M. N. Rosenbluth, Phys. Rev. **79**, 615 (1950).

[2] P. G. Blunden, W. Melnitchouk, and J. A. Tjon, Phys. Rev. Lett. **91**, 142304 (2003).

[3] P. A. M. Guichon and M. Vanderhaeghen, Phys. Rev. Lett. **91**, 142303 (2003).

[4] J. C. Bernauer *et al.* (A1 Collaboration), Phys. Rev. C **90**, 015206 (2014).

[5] R. D. Ball, L. Del Debbio, S. Forte, A. Guffanti, J. I. Latorre, J. Rojo, and M. Ubiali (NNPDF Collaboration), J. High Energy Phys. 05 (2010) 075.

[6] A. V. Gramolin and D. M. Nikolenko, Phys. Rev. C **93**, 055201 (2016).

[7] R. C. Walker *et al.*, Phys. Rev. D **49**, 5671 (1994).

[8] L. Andivahis *et al.*, Phys. Rev. D **50**, 5491 (1994).

[9] M. E. Christy *et al.*, Phys. Rev. Lett. **128**, 102002 (2022).

[10] G. Lee, J. R. Arrington, and R. J. Hill, Phys. Rev. D **92**, 013013 (2015).

[11] G. D'Agostini, Nucl. Instrum. Methods Phys. Res. A **346**, 306 (1994).

[12] Y.-H. Lin, H.-W. Hammer, and Ulf-G. Meißner, Phys. Rev. Lett. **128**, 052002 (2022).