

# Hadronic structure in high-energy collisions

Karol Kovařík


*Institut für Theoretische Physik, Westfälische Wilhelms-Universität Münster,  
Wilhelm-Klemm Straße 9, D-48149 Münster, Germany*

Pavel M. Nadolsky\*

*Department of Physics, Southern Methodist University, Dallas, Texas 75275-0181, USA*

Davison E. Soper†

*Institute of Theoretical Science, University of Oregon, Eugene, Oregon 97403-5203, USA*

 (published 4 November 2020)

Parton distribution functions (PDFs) describe the structure of hadrons as composed of quarks and gluons. They are needed to make predictions for short-distance processes in high-energy collisions and are determined by fitting to cross-section data. Definitions of the PDFs and their relations to high-energy cross sections are reviewed. The focus is on the PDFs in protons, but PDFs in nuclei are also discussed. The standard statistical treatment needed to fit the PDFs to data using the Hessian method is reviewed in some detail. Tests are discussed that critically examine whether the needed assumptions are indeed valid. Also presented are some ideas of what one can do in case tests indicate that the assumptions fail.

DOI: [10.1103/RevModPhys.92.045003](https://doi.org/10.1103/RevModPhys.92.045003)

## CONTENTS

I. Introduction	1	E. Dependence on the theory parameters	21
A. The parton model	2	F. Distribution of the parameters	23
B. Cross sections and factorization	3	G. Calculation of a cross section	24
C. Practical issues in the theory	4	1. Formulas to compute PDF uncertainties	25
D. The strong coupling	5	H. Expectation value and variance of $\chi^2$	25
E. Heavy-quark masses	5	IV. Tests of Performance of the Fit	26
F. Special kinematic regions	5	A. Testing with resampled data	27
G. Fitting	6	B. Dependence on the number of PDF parameters	28
II. Review of Theory	7	C. Dependence on the PDF functional form	29
A. Definition of parton distribution functions	7	D. Closure test	30
1. Momenta	7	E. Test of the nuisance parameters	30
2. Parton distributions in canonical field theory	8	F. Test of data residuals	31
3. Gauge invariance	9	G. Value of $\chi^2$ from an individual experiment	31
4. Renormalization	9	H. Test of $\chi^2$ from individual experiments	32
5. Gluons	10	I. Test of consistency between experiments	35
6. Evolution equation	10	J. A more conservative way to adjust the errors	38
B. Infrared safety and factorization	10	K. Summary of measures of goodness of fit	39
1. Kinematics	11	L. Global and dynamic tolerance	40
2. Infrared safety	11	V. Parton Distributions for Heavy Ions	41
3. Factorization	12	A. Universality of nuclear PDFs	41
C. Treatment of heavy quarks	14	B. Parametrizing the A dependence	43
1. Heavy-quark masses	15	C. Comparisons of nuclear PDFs	44
2. Fitted charm quarks	15	VI. Conclusions	47
D. Deeply inelastic scattering	16	Acknowledgments	47
III. Statistical Inference in Fitting the Parton Distributions	16	Appendix: Transformation for $\chi^2(D, a, \lambda)$	48
A. Bayes's theorem	17	References	48
B. Gaussian model for the distribution of data	18		
C. Determining theory parameters using $\chi^2$	19		
D. Another definition of $\chi^2$	19		

\*nadolsky@smu.edu

†soper@uoregon.edu

## I. INTRODUCTION

Since 2018, the Large Hadron Collider (LHC) has taken a large data sample of proton-proton collisions and has been using these data to precisely measure the properties of the Higgs boson and search for physics beyond the standard model. The ATLAS and CMS experiments took about

150 fb<sup>-1</sup> each at the center-of-mass energy of  $\sqrt{s} = 13$  TeV and more than 22 fb<sup>-1</sup> of data at  $\sqrt{s} = 8$  TeV. In addition, the LHCb experiment has accumulated more than 9 fb<sup>-1</sup> of data at various energies, including data at extreme rapidities. In addition to proton-proton collisions, all experiments at the LHC, and, in particular, the ALICE experiment, are taking data in proton-lead and lead-lead collisions.

All predictions at the LHC are dependent on knowledge of the quark and gluon content of the proton. The probability distributions of the constituents of the nucleon, collectively called *partons*, need to be known to make predictions in the theoretical framework of perturbative quantum chromodynamics (PQCD). The concept of a parton originated in a model by Bjorken (1969) and Feynman (1972), who referred to quasi-independent pointlike constituents observed inside hadrons that undergo deeply inelastic scattering. As the formalism of PQCD developed, partons were shown to be excitations of elementary quantum fields of spins 1/2 and 1 (quarks, antiquarks, gluons, and even photons) in a hadron undergoing a hard scattering process.

Our understanding of the structure of the proton is being updated continuously using the wealth of LHC and other world data. Knowledge of the structure of the nucleon needed for a large class of theoretical predictions in PQCD is encoded in the collinear parton distribution functions (PDFs). The PDFs have been determined from data going back to the early 1980s (Gluck, Hoffmann, and Reya, 1982; Duke and Owens, 1984; Eichten *et al.*, 1984). They are now determined using the method of the *global QCD analysis* (Martin, Roberts, and Stirling, 1988, 1989; Harriman *et al.*, 1990; Morfin and Tung, 1991; Owens, 1991) from experimental measurements at colliders such as HERA, the Tevatron, and the LHC, and in fixed-target experiments. The PDFs are provided in a number of practically useful forms by several collaborations, including ABM (Alekhin *et al.*, 2017), HERAPDF (Abramowicz *et al.*, 2015), CT (Dulat *et al.*, 2016), CTEQ-JLab (Accardi, Brady *et al.*, 2016), MMHT (Harland-Lang *et al.*, 2015), and NNPDF (Ball *et al.*, 2017). The modern PDF parametrizations are provided with families of “error PDF sets” (Giele and Keller, 1998; Giele, Keller, and Kosower, 2001; Pumplin *et al.*, 2001) that allow the user to assess the total uncertainty on the PDFs arising from a variety of experimental and theoretical errors. Methods for statistical combination of PDF ensembles from various groups exist (Gao and Nadolsky, 2014; Carrazza, Forte *et al.*, 2015; Carrazza, Latorre *et al.*, 2015), and comprehensive guidelines on uses of PDFs at the LHC were published by the PDF4LHC Working Group (Butterworth *et al.*, 2016). To date close to 700 PDF ensembles from various groups have been distributed as numerical tables in the standard *Les Houches Accord PDF* (LHAPDF) format (Giele *et al.*, 2002; Whalley, Bourilkov, and Group, 2005; Bourilkov, Group, and Whalley, 2006; Buckley *et al.*, 2015) from a public online repository.<sup>1</sup>

The parton distributions in nuclei have been analyzed by several collaborations: EPPS (Eskola *et al.*, 2017), nCTEQ (Kovářik *et al.*, 2016), DSSZ (de Florian *et al.*, 2012), HKN

(Hirai, Kumano, and Nagai, 2007) KA (Khanpour and Atashbar Tehrani, 2016), and NNPDF (Abdul Khalek, Ethier, and Rojo, 2019). They are provided either as parametrizations of the nuclear PDFs themselves or as the nuclear-correction factor applied to a predefined reference proton PDF.

Increasingly precise requirements are imposed on the determination of PDFs and their uncertainties during the high-luminosity (HL) phase of the LHC operation to precisely measure Higgs boson couplings and electroweak parameters (de Florian *et al.*, 2016), and to maximize the HL-LHC reach in a variety of tests of the standard model and new physics searches (ATLAS and CMS Collaborations, 2019).

The purpose of this review is to discuss select topics related to the theoretical definition, determination, and usage of PDFs in modern applications. We concentrate on methodological aspects of the PDF analysis that will be of growing importance in the near-future LHC era. We primarily focus on theoretical and statistical aspects of the determination of PDFs in the nucleon and nuclei, especially on proper theoretical definitions, statistical inference of the PDF parametrizations from the experimental data, and factorization for heavy nuclei. This review supplements the discussions of phenomenological applications of PDFs by Forte and Watt (2013) and Gao, Harland-Lang, and Rojo (2018), as well as extensive comparisons of PDFs from various collaborations and QCD predictions based on these PDFs (Alekhin *et al.*, 2011; Watt and Thorne, 2012; Accardi *et al.*, 2016; Butterworth *et al.*, 2016). Introductory texts on the fundamentals of QCD factorization, global PDF analysis, and collider applications of PDFs were given, e.g., by Brock *et al.* (1995), Collins (2013), and Campbell, Huston, and Krauss (2017).

We begin with the parton model and its relation to QCD, the field theory of the strong interactions.

## A. The parton model

The principal aim of contemporary particle physics is to test the current theory, the standard model, and to look for evidence of new physics that is not included in the standard model. The standard model is a renormalizable quantum field theory with fields for leptons ( $e$ ,  $\mu$ ,  $\tau$ ), their associated neutrinos, quarks ( $d, u, s, c, b, t$ ), vector bosons ( $\gamma$ ,  $W^\pm$ ,  $Z$ ,  $g$ ), and a scalar boson field, the Higgs field. The part of the theory involving the photon,  $W^\pm$ ,  $Z$ , and Higgs fields involves spontaneous symmetry breaking and is quite subtle. The part of the theory involving the gluon  $g$  constitutes the theory of the strong interactions, QCD. The gluon couples to quarks, but not to leptons. We cannot offer a review of the standard model, but we assume that the reader is familiar with it; see, for instance, Peskin and Schroeder (1995), Srednicki (2007), Aitchison and Hey (2012), and Schwartz (2014).

One important feature of the standard model, and quantum field theory in general, is that the couplings that appear at the vertices of Feynman diagrams representing the theory are best considered to be dependent on a squared momentum scale  $\mu^2$ . For instance, the QCD coupling constant  $g_s^2/(4\pi) \equiv \alpha_s$  becomes  $\alpha_s(\mu^2)$ . The dependence on  $\mu^2$  is derived from the theory, even though the value of  $\alpha_s(\mu_0^2)$  at a scale like  $\mu_0^2 = M_Z^2$  is a free parameter of the theory. Which value of  $\mu^2$  is

<sup>1</sup>See <https://lhpdf.hepforge.org>.

useful in addressing a particular physical problem depends on the typical momentum scale of the problem.

The electroweak coupling  $\alpha(\mu^2)$  increases with  $\mu^2$ , but for physically relevant momentum scales  $\alpha(\mu^2)$  is small. Thus, scattering cross sections for electroweak interactions can be usefully computed as a perturbation series in the small parameter  $\alpha(\mu^2)$ .

The QCD coupling constant  $\alpha_s(\mu^2)$  decreases with  $\mu^2$ . It is large when  $\mu$  is of the order of 1 GeV or less, and it reduces to  $\alpha_s(\mu^2) \approx 0.1$  at  $\mu = 100$  GeV. This scale dependence of  $\alpha_s(\mu^2)$  tells us that perturbation theory should be useful for describing the parts of a physical scattering process with hadrons that involves only large momentum scales.

However, we run into a complication. How can we use experiments involving protons to investigate the standard model? For instance, we know from experiment that we can make Higgs bosons in proton-proton collisions, but how can we understand the Higgs production process quantitatively? The Higgs boson production cross section depends on the internal structure of the initial-state protons. The strong coupling  $\alpha_s(\mu^2)$  is large for  $\mu$  of the order of the proton mass  $m_p$  or smaller. This means that we cannot expect perturbation theory to be useful for calculating the structure of the proton as a bound state of quarks, antiquarks, and gluons.

To understand the problem and its tentative solution, consider deeply inelastic electron scattering from a proton (DIS). The DIS process is reviewed in Sec. II.D. Suppose that an electron scatters from a proton of momentum  $P_A$  by exchanging a photon with momentum  $q^\beta$ .<sup>2</sup> We define  $Q^2 \equiv -q^2$ . Since the 4-momentum  $q^\beta$  is spacelike, we have  $Q^2 > 0$ . We demand that  $Q^2$  be much larger than 1 GeV<sup>2</sup>. Then there is some hope for a perturbative approach that uses an expansion in powers of  $\alpha_s(Q^2)$ . The lowest-order Feynman diagram, corresponding to  $t$ -channel electromagnetic scattering  $e + \overset{(-)}{q} \rightarrow e + \overset{(-)}{q}$  of an electron on a quark or antiquark, is pretty simple. But how do we relate the initial-state quark field to the proton?

Here is an approach we can follow. We use a “brick-wall” reference frame in which  $\vec{q}$  lies entirely in the  $-z$  direction,  $q^0 = 0$ , and in which  $\vec{P}_A$  lies entirely in the  $+z$  direction. Then  $|q^z| = Q$ , and we can consider the hard quark-photon interaction to be localized in a time interval  $\Delta t_H = 1/Q$ . In DIS, we also demand that  $P_A \cdot q$  be large, of the order of  $Q^2$ . Then the proton momentum is large, with  $P_A^z = (P_A \cdot q)/Q$ . This means that the proton is highly boosted, with a boost factor  $e^\omega = (2P_A \cdot q)/(m_p Q)$ . In the proton rest frame, we can suppose that the time period between successive quark-gluon interactions is of the order of  $\Delta t_S^{\text{rest}} \sim 1/m_p$ . In the brick-wall reference frame, the typical time for these soft internal interactions is  $\Delta t_S \sim e^\omega/m_p = 2P_A \cdot q/(m_p^2 Q)$  as a consequence of relativistic time dilation. The time interval  $\Delta t_S$  is much longer than the hard-interaction timescale:  $\Delta t_S/\Delta t_H \sim 2P_A \cdot q/m_p^2$ . This argument implies that the quark-gluon interactions inside the proton are largely frozen

during the time interval  $\Delta t_H$  while the hard scattering interaction with the virtual photon is taking place. The struck quark is effectively free.

This gives us the parton model of Bjorken (1969) and Feynman (1972). The word *parton* as used now refers to a quark, antiquark, gluon, and sometimes a photon. Originally, partons meant just the constituents of the proton, whatever those constituents might be.

It is easy to use the parton model to describe the cross section for DIS. We assume that, inside a fast-moving proton, there are partons of various types  $a$ . A parton can carry a momentum  $k^\beta$  that is a fraction  $\xi$  of the momentum of the proton:  $k^\beta = \xi P_A^\beta$ . Denote by  $f_{a/p}(\xi)d\xi$  the probability of finding such a parton with momentum fraction between  $\xi$  and  $\xi + d\xi$ . Let  $d\hat{\sigma}/[dQ^2 d(P_A \cdot q)]$  be the cross section for the electron to scatter from a free quark with momentum  $k = \xi P_A$ , calculated in lowest-order perturbation theory. Then the cross section for this scattering from a proton should be

$$\frac{d\sigma}{dQ^2 d(P_A \cdot q)} = \sum_a \int d\xi f_{a/p}(\xi) \frac{d\hat{\sigma}}{dQ^2 d(P_A \cdot q)}. \quad (1)$$

This picture is simple and intuitive. An analogous picture covers processes such as Higgs boson production in high-energy proton-proton collisions.

The parton model as just described does not survive scrutiny when we include gauge field interactions at higher orders of  $\alpha_s$ : one encounters contradictions as soon as one tries to calculate  $d\hat{\sigma}/[dQ^2 d(P_A \cdot q)]$  beyond the leading order in QCD. The basic problem is that QCD is a quantum field theory, in which interactions among the quarks and gluons occur at all time-scales or distance scales, including timescales much smaller than the  $\Delta t_S$  scale that one would naively associate with the interior motions inside a highly boosted proton.

Nevertheless, one can turn the parton model picture into a theoretically consistent framework that includes higher-order radiative contributions. One first needs to carefully define what one means by a parton distribution function  $f_{a/p}(\xi)$ . With a careful definition, the parton distribution functions become  $f_{a/p}(\xi, \mu^2)$ , with a dependence on a scale  $\mu^2$ . Then cross sections for DIS and for many processes in hadron-hadron scattering have a property known as *factorization*, which means that they satisfy a formula similar to Eq. (1).

## B. Cross sections and factorization

In this review, we review how the PDFs are systematically defined in the QCD theory and determined by applying statistical inference to experimental observations sensitive to the PDFs. We also review the general formalism to estimate the uncertainty on the PDFs that results from fitting the experimental data.

It is most common to define and determine the PDFs in the nucleon in the  $\overline{\text{MS}}$  factorization scheme discussed in some detail in Sec. II.A. Intuitively, these functions  $f_{a/A}(\xi, \mu^2)$  represent the probability of finding a parton of type  $a$  (a gluon or a particular flavor of quark or antiquark) in a hadron of type  $A$ , for example, a proton, as a function of the fraction  $\xi$  of the momentum of the hadron that is carried by the parton.

<sup>2</sup>We denote Lorentz indices with Greek letters and use the  $\{+, -, -, -\}$  sign convention for the Minkowski metric tensor.

The argument  $\mu^2$  in  $f$  indicates the momentum scale at which the parton distribution function applies. The  $\mu^2$  dependence is given by the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) evolution equations (Gribov and Lipatov, 1972; Altarelli and Parisi, 1977; Dokshitzer, 1977), which we describe in Sec. II.A.6. With the aid of these evolution equations, the functions  $f_{a/A}(\xi, \mu^2)$  can be determined from the functions  $f_{a/A}(\xi, \mu_0^2)$  at a scale  $\mu_0^2$  that is typically chosen to be around 1 GeV<sup>2</sup>. The functions at scale  $\mu_0^2$  cannot be calculated in perturbation theory.

The PDFs at the starting scale  $\mu_0^2$  are determined from experimental data. Consider first a cross section  $\sigma[F]$ , defined by integrating the completely differential cross section for any number of final-state particles, multiplied by functions  $F$  that describe what is measured in the final state. For instance, if we observe a single weakly interacting particle, the measurement function  $F$  might be simply a product of delta functions that specify the energy and direction of momentum of the particle. We thus integrate over the momenta of particles that are not measured, giving us an inclusive cross section. The observable  $\sigma[F]$  must be “infrared safe,” as described later in Sec. II.B.2. For lepton-hadron scattering, the cross section  $\sigma[F]$  is related to parton distributions by

$$\sigma[F] \approx \sum_a \int d\xi f_{a/A}(\xi, \mu^2) \hat{\sigma}[F]. \quad (2)$$

For cross sections at hadron colliders, a parton distribution function is needed for each of two colliding hadrons:

$$\sigma[F] \approx \sum_{a,b} \iint d\xi_a d\xi_b f_{a/A}(\xi_a, \mu^2) f_{b/B}(\xi_b, \mu^2) \hat{\sigma}[F]. \quad (3)$$

We review this formula in more detail in Sec. II.B.3.

To determine the PDFs at the starting scale  $\mu_0^2$ , one selects observables that are sensitive to different combinations of parton distributions. The parton distributions at scale  $\mu_0^2$  are parametrized by a sufficiently flexible functional form. The observables are first calculated using the parton distributions, where the free parameters are given some initial values and are compared to the data. The parameters are then adjusted until the theoretical predictions describe the data well.

### C. Practical issues in the theory

Following this simple strategy requires in reality a detailed understanding of many facets of perturbative QCD. First, as the precision of the determination of parton distributions needs to match the experimental precision, fitted observables are calculated at the next-to-next-to-leading-order (NNLO) in perturbative QCD for nucleon PDFs. The NNLO accuracy in the global fits corresponds to computing the hard cross sections by including perturbative radiative contributions suppressed by up to 2 powers of  $\alpha_s$ . Such predictions at NNLO for hard processes suitable for determination of PDFs are increasingly available (Zijlstra and van Neerven, 1991; van Neerven and Zijlstra, 1991; Buza and van Neerven, 1997; Moch, Vermaseren, and Vogt, 2005; Vermaseren, Vogt, and

Moch, 2005; Catani and Grazzini, 2007; Catani *et al.*, 2009; Campbell and Ellis, 2010; Gavin *et al.*, 2011, 2013; Li and Petriello, 2012; Berger *et al.*, 2016; Gehrmann-De Ridder *et al.*, 2016; Boughezal *et al.*, 2017; Currie, Glover, and Pires, 2017; Currie *et al.*, 2017; Gehrmann-De Ridder *et al.*, 2018. There are also partial results at N<sup>3</sup>LO, such as Vermaseren, Vogt, and Moch, 2005; Moch and Rogal, 2007; Bierenbaum, Blumlein, and Klein, 2009; Moch, Vermaseren, and Vogt, 2009; Ablinger *et al.*, 2011, 2015). Theoretical predictions used for nuclear PDFs are still typically at next-to-leading order (NLO), even though first NNLO nuclear PDF analyses exist.

Incorporating the theoretical calculations into the fit requires a careful selection of observables that are theoretically well defined (infrared safe) and can be calculated up to the required order. Because of the nature of higher-order calculations, the numerical evaluation can be time consuming. This shortcoming is usually solved either by using precomputed tables of computationally slow point-by-point NNLO corrections applied to fast NLO calculations, or increasingly by using fast gridding techniques, such as the ones implemented in the fastNLO (Wobisch *et al.*, 2011), APPLGRID (Carli *et al.*, 2010), aMCFast (Bertone *et al.*, 2014), and NNPDF FastKernel (Forte *et al.*, 2010) programs. The comparably accurate and fast DGLAP evolution of PDFs up to NNLO accuracy is implemented in a number of public codes: PEGASUS (Vogt, 2005), HOPPET (Salam and Rojo, 2009), QCDNUM (Botje, 2011), and APFEL (Bertone, Carrazza, and Rojo, 2014).

Even after implementing the measured observables and the corresponding DGLAP evolution for the PDFs at (N)NLO, one still has to address a number of issues that become important as the PDF analysis is pushed toward higher precision. On the experimental side, the NNLO PDFs are increasingly constrained by high-luminosity measurements, in which the statistical experimental errors are small, and adequate implementation of many (sometimes hundreds) of correlated systematic uncertainties is necessary. Commonly followed procedures for implementation of systematic uncertainties in the PDF fits are reviewed in Appendix A of Ball *et al.* (2013b). We discuss the treatment of systematic uncertainties in some detail in Sec. III.

From the side of theory, subtle radiative contributions, such as NLO electroweak or higher-twist contributions, are comparable to NNLO QCD contributions in some fitted observables. The photon constituents contribute at a fraction-of-percent level to the total momentum of the proton. The associated parton distribution for the photon can be computed accurately using the structure functions and nucleon form factors from lepton-hadron (in)elastic scattering as the input (Manohar *et al.*, 2016, 2017). The resulting LUXqed parametrization of the photon PDF was already implemented by Bertone, Carrazza *et al.* (2017) and Nathvani *et al.* (2018). An alternative is to fit a phenomenological parametrization of the photon PDF at the initial scale of evolution together with the rest of the PDFs. Such phenomenological parametrizations constrained just by the global fit (Ball, Bertone *et al.*, 2013; Schmidt *et al.*, 2016; Giuli *et al.*, 2017) are less precise than the LUXqed form.

Even at NNLO, the residual theoretical uncertainties due to missing higher-order contributions in  $\alpha_s$  may have an impact on PDFs. Such theory uncertainties are partly correlated in a generally unknown way across experimental data points. In addition to the traditional estimation of higher-order contributions by the variation of factorization and renormalization scales, recently, more elaborate methods for estimation of higher-order uncertainties have been explored with an eye on applications in the PDF fits, such as those given by [Olness and Soper \(2010\)](#), [Gao \(2011\)](#), [Forte, Isgro, and Vita \(2014\)](#), [Abdul Khalek \*et al.\* \(2019a, 2019b\)](#), [Cacciari and Houdeau \(2011\)](#), and [Harland-Lang and Thorne \(2019\)](#); see the discussion at Eq. (79).

#### D. The strong coupling

Another associated issue is the treatment of the strong coupling  $\alpha_s(\mu_R^2)$ . The strong coupling depends on a scale  $\mu_R^2$ , called the renormalization scale. For a review, see any text on QCD, for example, [Collins \(2013\)](#). Since  $\alpha_s(\mu_R^2)$  obeys a renormalization group equation, its value at any  $\mu_R^2$  can be determined from its value at a fixed scale  $\mu_{R0}^2$ . Normally, one sets  $\mu_{R0}$  as the mass  $M_Z$  of the Z boson. All QCD observables depend on  $\alpha_s(M_Z^2)$ , and so do the fitted PDFs. Conventionally, the world-average value of  $\alpha_s(M_Z^2)$  ([Tanabashi \*et al.\*, 2018](#)) is derived from a combination of experimental measurements, with the tightest constraints imposed by the QCD observables that do not depend on the PDFs, notably, hadroproduction in electron-positron collisions, hadronic  $\tau$  decays, and quarkonia masses.

Other useful constraints on  $\alpha_s$  are imposed by a variety of hadron-scattering observables (lepton-hadron DIS, jet and  $t\bar{t}$  production, etc.) that are simultaneously sensitive to PDFs, but the constraints of this class are generally weaker and more susceptible to systematic effects. As some hadronic observables of the latter class are also included in the global fit to constrain the PDFs, in principle, these observables can determine both  $\alpha_s$  and the PDFs at the same time. Consequently, several treatments of  $\alpha_s$  exist in the current PDF analyses. Most PDF groups publish some global fits that determine  $\alpha_s(M_Z^2)$  and the PDFs simultaneously. They typically find that the best-fit  $\alpha_s(M_Z^2)$  is consistent within the world average of  $\alpha_s(M_Z^2)$ , but with a considerably larger uncertainty than in the world average. ABM fits are representative of this approach ([Alekhin \*et al.\*, 2017](#)), as are the dedicated studies performed in the global framework ([Ball, Carrazza \*et al.\*, 2018](#); [Thorne, Harland-Lang, and Martin, 2018](#)).

As an example, the best-fit  $\alpha_s(M_Z^2) = 0.1166 \pm 0.0027$  in CT18 at NNLO ([Hou \*et al.\*, 2019](#)) is consistent with the world average  $\alpha_s(M_Z^2) = 0.1181 \pm 0.0011$  ([Tanabashi \*et al.\*, 2018](#)), as well as with  $\alpha_s(M_Z^2) \approx 0.1176$  in [Thorne, Harland-Lang, and Martin \(2018\)](#) and a somewhat higher  $\alpha_s(M_Z^2) = 0.1185 \pm 0.0012$  in [Ball, Carrazza \*et al.\* \(2018\)](#). The CT18 value results as some trade-off between the DIS experiments (notably, fixed-target DIS), which collectively prefer a somewhat lower  $\alpha_s(M_Z^2) \approx 0.115$ , and jet + top and Drell-Yan experiments, which collectively prefer a higher  $\alpha_s(M_Z^2) \approx 0.119$  ([Hou \*et al.\*, 2019](#)). The quoted uncertainty on  $\alpha_s(M_Z^2)$  at

68% probability level varies among the PDF-fitting groups in 2020 from about 0.001 to 0.0025 depending on the adopted definitions for the uncertainty (with the CT18 uncertainty the most conservative).

On the other hand, it is often advantageous to perform the PDF fits and determine the PDF uncertainty at a fixed world-average value of  $\alpha_s(M_Z^2)$ , then estimate the  $\alpha_s$  uncertainty of the fits by using a few PDF fits with alternative  $\alpha_s$  values. If the uncertainties obey a Gaussian probability distribution, it can be rigorously demonstrated that, to compute the total PDF +  $\alpha_s$  uncertainty that includes all correlations, it suffices to add the resulting PDF and  $\alpha_s$  uncertainties in quadrature ([Lai, Huston \*et al.\*, 2010](#)). The empirical probability distributions in the PDF fits are indeed sufficiently close to being Gaussian, so this prescription for computing the PDF +  $\alpha_s$  uncertainty is adopted by a majority of recent fits. For example, the PDF4LHC group recommends ([Butterworth \*et al.\*, 2016](#)) calculating the PDF +  $\alpha_s$  uncertainty at the 68% confidence level by adding in quadrature the PDF uncertainty computed using 30 (100) PDF4LHC15 error sets for the world-average  $\alpha_s(M_Z^2) = 0.1180$ , and the  $\alpha_s$  uncertainty computed from two best-fit PDF sets for  $\alpha_s(M_Z^2) = 0.1165$  and 0.1195.

#### E. Heavy-quark masses

Another issue that needs to be addressed in global fits of PDFs is the treatment of massive charm and bottom quarks. Mass effects play an important role in describing, for example, subprocesses with charm (anti)quarks in DIS. There are several approaches to treating the mass of the quark, such as the zero-mass variable-flavor-number (ZM-VFN) scheme ([Collins, Wilczek, and Zee, 1978](#); [Collins and Tung, 1986](#)) or the general-mass variable-flavor-number (GM-VFN) scheme ([Aivazis \*et al.\*, 1994](#); [Collins, 1998](#); [Thorne and Roberts, 1998](#); [Krämer, Olness, and Soper, 2000](#); [Thorne, 2006](#); [Forte \*et al.\*, 2010](#)). In Sec. II.C, we provide a pedagogical introduction to the extensive topic of the treatment of masses for heavy quarks. For a more thorough review of the heavy-quark schemes see [Accardi \*et al.\* \(2016\)](#) and [Butterworth \*et al.\* \(2016\)](#).

#### F. Special kinematic regions

Some kinematic regions require special treatment if their respective experimental data are to be included in a global fit of PDFs. One such region is where the typical momentum fraction  $\xi$  is large but the momentum scale of the process is not too large. Then the nonzero mass of the proton (or, in general, the target), normally neglected, may need to be taken into account. These target-mass corrections were discussed in detail by [Schienbein \*et al.\* \(2008\)](#). In the same region, deuterium nuclear corrections also have to be considered when data in this specific kinematic region are taken on deuterium rather than proton targets ([Accardi \*et al.\*, 2010, 2011](#)).

The other kinematic region in need of careful treatment is the one where the momentum transfer  $Q^2$  is low, typically below 4 GeV<sup>2</sup>, at moderate or large  $\xi$ . Here power corrections can become important and can be taken into account, for example, as in [Martin \*et al.\* \(2004\)](#), [Alekhin, Blumlein, and Moch \(2012\)](#), and [Thorne \(2014\)](#).

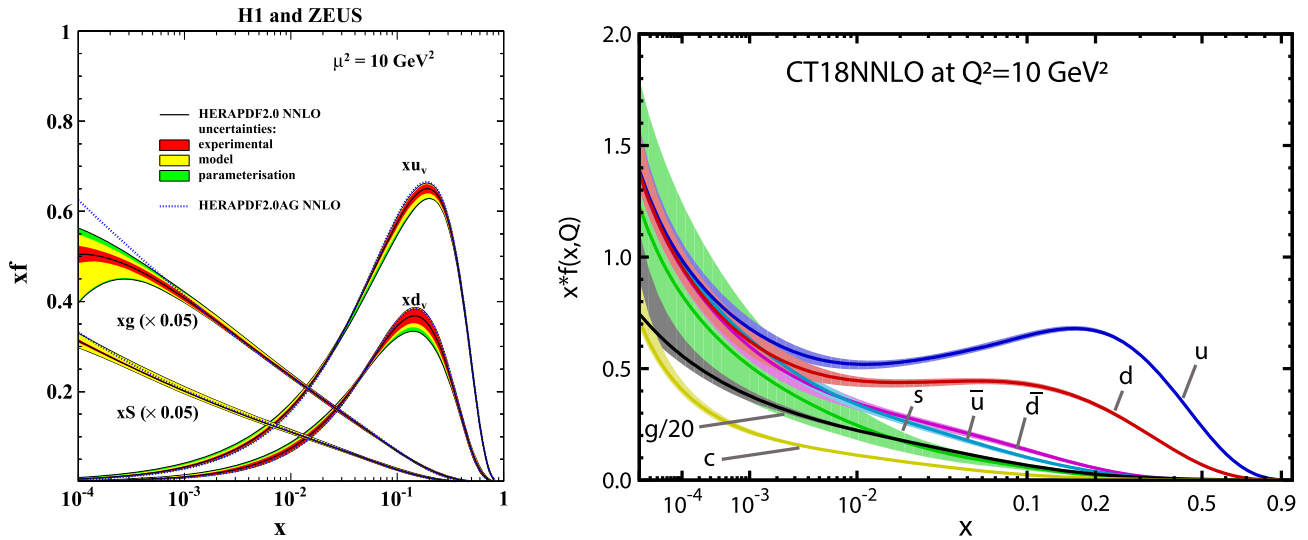


FIG. 1. Left panel: the parton distribution functions  $xu_v \equiv x(u - \bar{u})$ ,  $xd_v \equiv x(d - \bar{d})$ ,  $xS \equiv 2x(\bar{u} + \bar{d} + \bar{s} + \bar{c})$ , and  $xg$  of HERAPDF2.0 NNLO at  $\mu^2 = 10 \text{ GeV}^2$ . The experimental, theoretical model, and parameterization uncertainties are shown separately. From [Abramowicz et al., 2015](#). Right panel: the PDF uncertainty bands for CT18 NNLO PDFs ([Hou et al., 2019](#)) at  $\mu^2 = 10 \text{ GeV}^2$ .

Another such region arises in DIS at small  $x$  and  $Q$ , roughly satisfying  $Q^2 < A_{\text{cut}}/x^\lambda$  with  $A_{\text{cut}} \sim 0.5\text{--}1.5 \text{ GeV}^2$ , and  $\lambda \sim 0.3$  ([Golec-Biernat and Wusthoff, 1999](#); [Caola, Forte, and Rojo, 2010](#)). (Here  $x = x_{bj}$ . See Sec. II.D.) This is the limit where summation of small- $x$  logarithms becomes necessary, and indeed, a slowdown in perturbative convergence of inclusive DIS cross sections and resulting small- $x$  PDFs is observed even at NNLO in the affected HERA region ([Abramowicz et al., 2015](#)). The DIS data in this region provide valuable constraints on the small- $x$  behavior of the gluon PDF. The small- $x$  instability in DIS can be cured to a certain extent by inclusion of power-suppressed (higher-twist) contributions ([Harland-Lang et al., 2016](#)) or, effectively in the HERA region at  $Q^2 > 4 \text{ GeV}^2$ , by using an  $x$ -dependent factorization scale  $\mu^2$  in NNLO DIS cross sections in some of the PDF sets (CT18X and CT18Z) published by [Hou et al. \(2019\)](#). Summation of small- $x$  logarithms, matched to NNLO, was successfully implemented in NNPDF ([Ball, Bertone et al., 2018](#)) and xFitter analyses ([Abdolmaleki et al., 2018](#)). It results in an even better description of the accessible small- $x$  region. Either NNLO + NLLx summation, as in [Ball, Bertone et al. \(2018\)](#), or the choice of a special  $x$ -dependent factorization scale in the fixed-order NNLO cross section, as in [Hou et al. \(2019\)](#), thus leads to a better description of the small- $x$  subsample of the HERA DIS data. The resulting PDFs obtained after these changes tend to have elevated gluon and strangeness components at  $x < 10^{-2}$  as a result of slower  $Q$  dependence of the DIS cross sections at small  $x$  than would be predicted at a fixed order with a standard scale  $\mu = Q$ . LHC predictions based on such modified PDFs, such as CT18Z, may lie outside of the nominal error bands of the PDF set with default choices, such as CT18.

## G. Fitting

After addressing all necessary features of theory predictions such as the ones just spelled out, one compares the theory

predictions to the experimental data. The process of fitting the theoretical predictions to data by adjusting the PDFs is the main focus of this review. The reason is that a proper determination of PDF uncertainties is important for an analysis of the high-luminosity LHC data, as the PDF uncertainty will soon dominate systematic uncertainties on the theory side in key tests of electroweak symmetry breaking, including the measurements of Higgs couplings and mass of the charged weak boson ([de Florian et al., 2016](#); [ATLAS and CMS Collaborations, 2019](#)). The statistical framework of the PDF fits is fundamentally more complex than the one in the electroweak precision fits: while the parametric model of the electroweak fits is uniquely determined by the standard model Lagrangian, the parametric model for the parton distribution functions may change within some limits to optimize agreement between QCD theory and data.

Consequently, the PDF uncertainty comprises four categories of contributions:

- (1) *Experimental uncertainties*, including statistical and correlated and uncorrelated systematic uncertainties of each experimental dataset.
- (2) *Theoretical uncertainties*, including the absent higher-order and power-suppressed radiative contributions, as well as uncertainties in using parton showering programs to correct the data in order to compare to fixed-order perturbative cross sections.
- (3) *Parameterization uncertainties* associated with the choice of the PDF functional form.
- (4) *Methodological uncertainties*, such as those associated with the selection of experimental datasets, fitting procedures, and goodness-of-fit criteria.

As an illustration, the left panel of Fig. 1 shows the HERAPDF2.0 parameterizations determined exclusively from the fits to DIS data. The PDF uncertainty corresponding to the PDF solutions covering 68% of the cumulative probability comprises the experimental, theoretical model, and parameterization components that were estimated for a select fitting

methodology (Abramowicz *et al.*, 2015). Other groups may not separate all four of the previously listed components in the total PDF uncertainty. In the right panel of Fig. 1, the CT18 NNLO PDF uncertainty bands are evaluated for 68% cumulative probability according to a two-tier goodness-of-fit criterion (Lai, Guzzi *et al.*, 2010) that accounts both for the agreement with the totality of fitted data and with individual experimental datasets. The CT18 analysis includes a variety of datasets on DIS, vector boson, jet, and  $t\bar{t}$  production. While this diversity of data allows one to resolve differences between PDFs of various flavors and probe a broader range of PDF parametrization forms, in practice, some incompatibilities (*tensions*) between constraints on the PDFs from various experiments are introduced and need to be either eliminated or accounted for in the PDF uncertainty estimate. (The CT18 and HERAPDF2.0 PDFs are fitted to 3690 and 1130 data points, respectively. About 350 different parametrization forms have been tried in the CT18 analysis, contributing to the spread of the PDF uncertainty.) The width of the CT18 error bands thus depends on a two-level *tolerance* convention (Lai, Guzzi *et al.*, 2010; Pumplin *et al.*, 2002) that is adjusted so as to reflect PDF variations associated with some disagreements between experiments, parametrization, and theoretical uncertainties. We notice, for example, that the CT18 error bands for some poorly constrained flavors, notably, the strangeness PDF  $x s(x, \mu^2)$  at small  $x$  (green band), may be broader than the respective HERAPDF2.0 error bands at the same probability level, despite having more experimental data included in the CT18 analysis compared to HERAPDF. The wider error bands reflect, for a large part, the spread in the acceptable PDFs estimated using the CT18 flexible parametrization forms, but also some inflation of the experimental uncertainty to reflect the imperfect agreement among experiments. Section IV.H shows how to examine several experiments for their agreement.

To find most likely solutions for PDFs and establish the respective uncertainties, one must answer a fundamental question: how good, actually, is each PDF fit? We explore this question and advocate using a strong set of goodness-of-fit criteria that go beyond the weak criterion based on just the value of the goodness-of-fit function  $\chi^2$ .

All PDF fitters employ some version of minimization of the goodness-of-fit function

$$\chi^2(a) = \sum_{ij} [D_i - T_i(a)][D_j - T_j(a)] C_{ij}, \quad (4)$$

where  $D_i$  are the data values,  $T_i(a)$  are the corresponding theory predictions, which depend on free PDF parameters  $a$ , and  $C_{ij}$  is the covariance matrix. The goodness-of-fit function is used to assess the quality of the theoretical description of the data and to estimate the uncertainty in the determination of the fit parameters  $a$ . The statistical foundations that motivate the use of the goodness-of-fit functions are discussed in Sec. III. While it is most common to find the global minimum of  $\chi^2(a)$  numerically, many insights about the global fits can be gleaned analytically in the so-called Hessian approach to fitting the PDFs, first developed by Pumplin *et al.* (2001) and Stump *et al.* (2001) and refined since. We discuss this

approach in great detail to illustrate various aspects of the fits. It relies on the observation that the PDFs approximately obey the multivariate Gaussian probability distribution in the well-constrained kinematic regions, which in turn allows one to derive the key outcomes of the PDF analysis in a closed algebraic form. For example, the Hessian method is commonly used at the end of the fit to quantify the uncertainty on the resulting PDFs. There is a powerful alternative approach for finding the best-fit PDFs and determining their uncertainties using stochastic (Monte Carlo) sampling of PDFs (Giele and Keller, 1998; Giele, Keller, and Kosower, 2001) and PDF parametrizations by neural networks (Forte *et al.*, 2002). Although our results are demonstrated in the Hessian approximation, they also elucidate the numerical outcomes of the Monte Carlo sampling PDF analyses such as the one used by NNPDF. They also apply to the approximate techniques for updating the published PDF ensembles with information from new data by statistical reweighting of PDF replicas in the Monte Carlo (Giele and Keller, 1998; Ball *et al.*, 2011, 2012; Sato, Owens, and Prosper, 2014) or Hessian (Watt and Thorne, 2012; Paukkunen and Zurita, 2014; Schmidt, Pumplin, and Yuan, 2018) representations.

Section IV is devoted to the discussion of tests that one can perform to determine the extent to which the fitting procedure is consistent with the statistical hypotheses used in the procedure. This leads to a discussion of the strong goodness-of-fit set of criteria.

## II. REVIEW OF THEORY

In this section, we provide a brief overview of the theory of PDFs and their relation to cross sections. We start with the definition of PDFs as matrix elements of quantum field operators. Then we discuss the factorization property of QCD, which allows us to relate certain kinds of cross sections to PDFs and perturbatively calculated quantities. Finally, we turn to the treatment of heavy quarks in these relations, although we treat this complex subject only briefly.

### A. Definition of parton distribution functions

Here we give definitions for PDFs as matrix elements in a proton (or other hadron) of certain operators. Instead of simply stating the definitions, we motivate them from basic field theory, following the reasoning of Collins and Soper (1982). For more details, see Collins (2013).<sup>3</sup>

#### 1. Momenta

Consider a proton with momentum  $P$  along the  $+z$  direction. We define the  $+$  and  $-$  components of vectors using  $v^\pm = (v^0 \pm v^3)/\sqrt{2}$ . Then  $P$  has the components

$$(P^+, P^-, P^\perp) = \left( P^+, \frac{m_p^2}{2P^+}, \mathbf{0} \right). \quad (5)$$

<sup>3</sup>Our conventions follow the Particle Data Group (Tanabashi *et al.*, 2018) and Collins (2013). In particular, we choose the sign of the strong coupling  $g$  so that the quark-gluon vertex is  $-ig\gamma^\mu t^a$ . This is the opposite of the choice of Collins and Soper (1982).

It is helpful to think of  $P^+$  as large, 9.2 TeV for the LHC, but the size of  $P^+$  does not matter for the definition of the PDFs.

When in the following exposition we introduce a function  $F$  of a four-vector  $v$ , we use two alternative notations,  $F(v)$  and  $F(v^+, v^-, \mathbf{v}_\perp)$ .

We seek to define PDFs  $f_{a/p}(\xi, \mu^2)$ , which can be interpreted as giving the probability density for finding a parton of flavor  $a$  (a quark, antiquark, or gluon), which carries a fraction  $\xi$  of  $P^+$ , in the proton with momentum  $P$ . This function depends on the momentum squared scale  $\mu^2$  at which one imagines measuring the presence of the parton.

## 2. Parton distributions in canonical field theory

To get started with the definition of PDFs, consider an operator  $b(\xi P^+, \mathbf{k}_\perp, s, c; i)$  that destroys a quark of flavor  $i$  having helicity  $s$ , color  $c$ ,  $+$ -momentum  $\xi P^+$ , and transverse momentum  $\mathbf{k}_\perp$ . This quark then carries a fraction  $\xi$  of the  $+$ -momentum  $P^+$  of the proton. The adjoint operator  $b^\dagger(\xi P^+, \mathbf{k}_\perp, s, c; i)$  then creates a quark with the same quantum numbers. We normalize the creation and destruction operators to have anticommutation relations

$$\begin{aligned} & [b(\xi' P^+, \mathbf{k}'_\perp, s', c'; i), b^\dagger(\xi P^+, \mathbf{k}_\perp, s, c; i)]_+ \\ &= (2\pi)^3 2\xi P^+ \delta(\xi' P^+ - \xi P^+) \delta(\mathbf{k}'_\perp - \mathbf{k}_\perp) \delta_{s's} \delta_{c'c}. \end{aligned} \quad (6)$$

Additionally, we suppose that the vacuum state  $|0\rangle$  has no quarks in it, so

$$b(\xi P^+, \mathbf{k}_\perp, s, c; i)|0\rangle = 0. \quad (7)$$

With the quark creation and destruction operator at hand, we can construct the following operator that counts the number of quarks in a region of  $\xi$  and  $\mathbf{k}_\perp$ :

$$\begin{aligned} & \rho(\xi P^+, \mathbf{k}_\perp; i) \\ &= \frac{1}{(2\pi)^3 2\xi} \sum_{s,c} b^\dagger(\xi P^+, \mathbf{k}_\perp, s, c; i) b(\xi P^+, \mathbf{k}_\perp, s, c; i). \end{aligned} \quad (8)$$

The reader can verify that if  $|\Psi\rangle$  is obtained by applying quark creation operators to the vacuum then the integral of  $\rho$  over a momentum-space volume  $\mathcal{V}_3$  counts the number  $N(\mathcal{V}_3)$  of quarks in  $\mathcal{V}_3$ :

$$\int_{\mathcal{V}_3} d\xi d\mathbf{k}_\perp \rho(\xi P^+, \mathbf{k}_\perp; i) |\Psi\rangle = N(\mathcal{V}_3) |\Psi\rangle. \quad (9)$$

We want to define a parton density, the number of partons  $f_{i/p}(\xi)$  of flavor  $i$  per unit  $d\xi$  in a proton. We can take the following matrix element of  $\rho$  in a proton state to define this:

$$f_{i/p}^{(0)}(\xi) \langle P|P\rangle = \int d\mathbf{k}_\perp \langle P| \rho(\xi P^+, \mathbf{k}_\perp; i) |P\rangle. \quad (10)$$

Here, for simplicity, we consider the proton to be spinless, but one can substitute a spin average  $(1/2) \sum_{s_p} \langle P', s_p | \cdots | P, s_p \rangle$ . As previously noted, we take the proton momentum  $P$  to be along the  $z$  axis so that  $\mathbf{P}_\perp = 0$ . However,  $P^+$  is arbitrary. We

give  $f$  a superscript (0) to indicate that this is a preliminary version of the needed definition.

To make this definition more useful, we can relate the quark creation and destruction operators to the quark field operator  $\psi_i(x)$ . For this purpose, we use the version of QCD quantized on planes of equal  $x^+ = (x^0 + x^3)/\sqrt{2}$  instead of planes of equal time  $t = x^0$  (Kogut and Soper, 1970; Bjorken, Kogut, and Soper, 1971). The fields obey canonical commutation relations on planes of equal  $x^+$ . To make this work, we use the gauge  $A^+(x) = 0$  for the gluon field. With this way of writing the theory, the two components of the four-component Dirac field projected by  $P_{\text{dy}} = (1/2)\gamma^- \gamma^+$  (such that  $P_{\text{dy}}^2 = P_{\text{dy}}$ ) are the independent dynamical fields (dy) representing quarks. The dynamical part of the quark field at  $x^+ = 0$  is related to quark and antiquark creation and destruction operators by

$$\begin{aligned} P_{\text{dy}} \psi_{i,c}(0, x^-, \mathbf{x}_\perp) &= \frac{1}{(2\pi)^3} \int_0^\infty \frac{dk^+}{2k^+} \int d\mathbf{k}_\perp \\ &\times \sum_s \{ P_{\text{dy}} u(k, s) e^{-ik \cdot x} b(k^+, \mathbf{k}_\perp, c, s; i) \\ &+ P_{\text{dy}} v(k, s) e^{+ik \cdot x} d^\dagger(k^+, \mathbf{k}_\perp, c, s; i) \}. \end{aligned} \quad (11)$$

Here  $k \cdot x = k^+ x^- - \mathbf{k}_\perp \cdot \mathbf{x}_\perp$ , and  $d^\dagger$  is an antiquark creation operator, analogous to the quark creation operator  $b^\dagger$ . The field  $\psi$  carries a flavor index  $i$ . It also carries a color index  $c$ , which we normally suppress. The spinors  $u$  and  $v$  are the usual solutions of the free Dirac equation, normalized to  $\bar{u}(k, s) \gamma^+ u(k, s) = 2k^+$  and  $\bar{v}(k, s) \gamma^+ v(k, s) = 2k^+$ . Then one easily finds that  $P_{\text{dy}} u(k, s)$  and  $P_{\text{dy}} v(k, s)$  depend only on the  $+$  component of  $k$ .

When we combine Eqs. (10) and (11), we directly obtain

$$\begin{aligned} f_{i/p}^{(0)}(\xi) \langle P'|P\rangle &= \frac{P^+}{2\pi} \int dy^- e^{-i\xi P^+ y^-} \int dx^- d\mathbf{x}_\perp \\ &\times \langle P'| \bar{\psi}_i(0, x^- + y^-, \mathbf{x}_\perp) \gamma^+ \psi_i(0, x^-, \mathbf{x}_\perp) |P\rangle. \end{aligned} \quad (12)$$

We can eliminate the factor  $\langle P'|P\rangle$  by using translation invariance to write

$$\begin{aligned} & \langle P'| \bar{\psi}_i(0, x^- + y^-, \mathbf{x}_\perp) \gamma^+ \psi_i(0, x^-, \mathbf{x}_\perp) |P\rangle \\ &= e^{i(P'-P) \cdot x^- - (\mathbf{P}'_\perp - \mathbf{P}_\perp) \cdot \mathbf{x}_\perp} \\ &\times \langle P'| \bar{\psi}_i(0, y^-, \mathbf{0}) \gamma^+ \psi_i(0, 0, \mathbf{0}) |P\rangle. \end{aligned} \quad (13)$$

Then we can perform the  $x^-$  and  $\mathbf{x}_\perp$  integrations to give delta functions that set  $P'$  to  $P$ . We normalize our proton state vectors to

$$\langle P'|P\rangle = (2\pi)^3 2P^+ \delta(P'^+ - P^+) \delta(\mathbf{P}'_\perp - \mathbf{P}_\perp). \quad (14)$$

Then the delta functions from  $\langle P'|P\rangle$  in Eq. (12) cancel. We set  $P'$  to  $P$  to get



$$f_{i/p}^{(0)}(\xi) = \frac{1}{4\pi} \int dy^- e^{-i\xi P^+ y^-} \langle P | \bar{\psi}_i(0, y^-, \mathbf{0}) \gamma^+ \psi_i(0) | P \rangle. \quad (15)$$

We present this result in some detail to emphasize that the PDF for quarks is simply the proton matrix element of the number density operator for quarks as obtained in canonically quantized field theory.

### 3. Gauge invariance

Next, without changing  $f_{i/p}^{(0)}(\xi)$ , we can rewrite the definition in a way that makes it gauge invariant. The canonical field theory that our derivation has relied on makes use of the lightlike axial gauge  $A^+(x) = 0$  for the gluon field. In an arbitrary gauge, we merely insert a Wilson line factor

$$W(y^-, 0) = \mathcal{P} \exp \left( -ig \int_0^{y^-} d\bar{y}^- A^+(0, \bar{y}^-, \mathbf{0})_a t_a \right). \quad (16)$$

This is a matrix in the color indices carried by the quark fields;  $t_a$  is the SU(3) generator matrix in the  $\mathbf{3}$  representation. The  $\mathcal{P}$  indicates path ordering of the operators and matrices, with more positive  $y^-$  values to the left. The revised definition is

$$f_{i/p}^{(0)}(\xi) = \frac{1}{4\pi} \int dy^- e^{-i\xi P^+ y^-} \times \langle P | \bar{\psi}_i(0, y^-, \mathbf{0}) \gamma^+ W(y^-, 0) \psi_i(0) | P \rangle. \quad (17)$$

The factor  $W$  is just 1 if we use  $A^+(x) = 0$  gauge. If we change the gauge by a unitary transformation  $U(x)$ , we replace

$$\begin{aligned} \psi_i(0) &\rightarrow U(0) \psi_i(0), \\ \bar{\psi}_i(0, y^-, \mathbf{0}) &\rightarrow \bar{\psi}_i(0, y^-, \mathbf{0}) U(0, y^-, \mathbf{0})^{-1}, \\ W(y^-, 0) &\rightarrow U(0, y^-, \mathbf{0}) W(y^-, 0) U(0)^{-1}. \end{aligned} \quad (18)$$

Thus, when we include the operator  $W$ , the right-hand side of the equation is invariant under a change of gauge.

If we use a covariant (Bethe-Salpeter) wave function for the proton state, we can use Eq. (17) for perturbative calculations. The field  $\psi_i(0)$  absorbs a quark line from the wave function. Similarly,  $\bar{\psi}_i(0, y^-, \mathbf{0})$  creates a quark line that goes into the conjugate wave function. These quark lines can emit and absorb gluons. The factor  $W(y^-, 0)$  is conveniently written as  $W(y^-, \infty)$  times  $W(\infty, 0)$ . The operator  $W$  contains gluon fields that create and absorb gluons. In a simple intuitive picture, we do not just destroy a quark at position 0, leaving its color with nowhere to go. Instead we scatter it so that it moves to infinity along a fixed lightlike line in the minus direction, carrying its color with it. Then its color comes back to  $(0, y^-, \mathbf{0})$  to provide the color for the quark that we create.

### 4. Renormalization

The function  $f_{i/p}^{(0)}(\xi)$  has thus far been defined using “bare” fields, a bare coupling, bare parton masses, and a bare operator product of fields in a canonical formulation of the field theory. This will not do. Even the simplest one-loop calculation reveals that the bare  $\alpha_s$ , quark masses, and  $f_{i/p}^{(0)}(\xi)$  contain

ultraviolet (UV) divergences. Thus, we need to renormalize everything. The standard way to do this is to apply  $\overline{\text{MS}}$  renormalization with scale  $\mu^2$ . For this, we need to choose a number  $N_f$  of active flavors.<sup>4</sup> The  $\overline{\text{MS}}$ -renormalized entities acquire dependence on  $\mu^2$ , which can be chosen so as to improve perturbative convergence for the short-distance cross section  $\hat{\sigma}$  in Eq. (3). Physically, quark and gluon interactions at distance scales smaller than  $1/\mu^2$  are not resolved in these objects.

This gives us our final definition for the quark distribution (Collins and Soper, 1982)

$$f_{i/p}(\xi, \mu^2) = \frac{1}{4\pi} \int dy^- e^{-i\xi P^+ y^-} \times \langle P | \bar{\psi}_i(0, y^-, \mathbf{0}) \gamma^+ W(y^-, 0) \psi_i(0) | P \rangle, \quad (19)$$

where  $W(y^-, 0)$  is given by Eq. (16). We understand now that the formulas refer to fields and couplings and field products that are renormalized with the  $\overline{\text{MS}}$  prescription for all active quarks and gluons.

For antiquarks, the analogous definition is

$$f_{\bar{i}/p}(\xi, \mu^2) = \frac{1}{4\pi} \int dy^- e^{-i\xi P^+ y^-} \times \langle P | \text{Tr}[\gamma^+ \psi_i(0, y^-, \mathbf{0}) W(y^-, 0) \bar{\psi}_i(0)] | P \rangle, \quad (20)$$

where the color generator matrices in  $W(y^-, 0)$  are in the  $\bar{\mathbf{3}}$  representation of SU(3).

Note that no approximations are made in Eq. (19). In particular, we do not treat the quarks as massless. We cannot calculate  $f_{i/p}(\xi, \mu^2)$  at any finite order of perturbation theory, but we could, in principle, calculate it using lattice gauge theory. In such a calculation, we use our best estimates for the parameters in the QCD Lagrangian, including the strong coupling and the quark masses.

In fact, PDFs can be calculated using lattice gauge theory (Lin *et al.*, 2018), but the accuracy of such calculations is still limited. One can obtain much better accuracy by fitting the parton distributions to the data, as described in this review. However, the definition of the parton distributions is not affected by the approximations that we make in the fitting procedure. For instance, the calculated cross sections used in the fit can be leading order (LO), next-to-leading order (NLO), or next-to-next-to-leading order (NNLO). The resulting fits are often referred to as LO, NLO, and NNLO. However, it is the fits that carry these designations. The functions  $f_{i/p}(\xi, \mu^2)$  that we are trying to estimate are nonperturbative objects whose definitions are independent of the fitting method.

<sup>4</sup>For instance, if we are following the  $N_f = 5$  convention, then neither  $\alpha_s$  nor the PDFs include contributions from top quarks. Then top-quark virtual loops can still occur within the Feynman diagrams, but they are treated using the Collins-Wilczek-Zee prescription (Collins, Wilczek, and Zee, 1978), in which the UV divergencies that they introduce are subtracted at zero incoming momenta and do not affect scale dependence of  $\alpha_s$  or the PDFs. Depending on the circumstances, one uses different numbers of active flavors, as we discuss in Sec. II.C.

## 5. Gluons

We previously defined the PDFs for quarks and antiquarks by beginning with the number operator for quarks or antiquarks in unrenormalized canonical field theory using null-plane quantization in  $A^+ = 0$  gauge. The starting definition is then generalized to be gauge invariant and to use  $\overline{\text{MS}}$ -renormalized operators. One can follow the same sort of logic for the gluon field. We simply state the result (Collins and Soper, 1982)

$$f_{g/p}(\xi, \mu^2) = \frac{1}{2\pi\xi P^+} \int dy^- e^{-i\xi P^+ y^-} \times \langle P | G(0, y^-, \mathbf{0})^{+\nu} W(y^-, 0) G(0)_\nu^+ | P \rangle, \quad (21)$$

where  $G_{\mu\nu}$  is the gluon field operator,

$$G_{\mu\nu}^a = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a - gf^{abc} A_\mu^b A_\nu^c, \quad (22)$$

and the Wilson line operator  $W(y^-, 0)$  is given by Eq. (16), now using SU(3) generator matrices  $(t_a)_{bc} = -if_{abc}$  in the adjoint representation.

## 6. Evolution equation

We take a closer look at the renormalization of the PDFs. The  $\overline{\text{MS}}$  renormalization of the strong coupling  $\alpha_s \equiv g^2/(4\pi)$  and the fields  $\psi_i(x)$  and  $A^\mu(x)_a$  in  $n = 4 - 2\epsilon$  dimensions proceeds in the usual way by subtracting  $1/\epsilon$  poles and some finite terms from two-point subgraphs, three-point subgraphs, and four-gluon subgraphs with loops containing gluons and the  $N_f$  active quarks. Another sort of pole arises in matrix elements for PDFs from operator products like  $\bar{\psi}_i(0, y^-, \mathbf{0})\gamma^+\psi_i(0)$  in Eq. (19). Consider a graph in which a gluon is emitted from a propagator representing the quark that is destroyed by  $\psi_i(0)$ , then absorbed by a propagator representing the quark created by  $\bar{\psi}_i(0, y^-, \mathbf{0})$ . This gluon line creates a loop subgraph that is UV divergent in four dimensions.

We subtract the divergence using the  $\overline{\text{MS}}$  prescription, which creates dependence of  $f_{a/p}(\xi, \mu_F^2)$  on the *factorization scale*  $\mu_F^2$ , possibly different from the renormalization scale  $\mu_R^2$ . In much of our discussion, we assume that  $\mu_R^2$  and  $\mu_F^2$  are the same and denote both as  $\mu^2$ .

By examining the structure of the UV divergences, one finds that the functions  $f_{a/p}(\xi, \mu_F^2)$  obey the DGLAP evolution equations

$$\frac{d}{d\log\mu_F^2} f_{a/p}(\xi, \mu_F^2) = \sum_{\hat{a}} \int_{\xi}^1 \frac{dz}{z} P_{a\hat{a}}(z, \alpha_s(\mu_F^2)) f_{\hat{a}/p}(\xi/z, \mu_F^2). \quad (23)$$

The functions  $f_{a/p}(\xi, \mu_F^2)$  are nonperturbative, but since the dependence on  $\mu_F^2$  arises from the UV divergences of graphs for  $f_{a/p}(\xi, \mu_F^2)$ , the evolution kernels  $P_{ab}(z, \alpha_s(\mu_F^2))$  are perturbatively calculable as expansions in powers of  $\alpha_s(\mu_F^2)$ :

$$P_{a\hat{a}}(z, \alpha_s(\mu_F^2)) = \frac{\alpha_s(\mu_F^2)}{2\pi} P_{a\hat{a}}^{(1)}(z) + \left[ \frac{\alpha_s(\mu_F^2)}{2\pi} \right]^2 P_{a\hat{a}}^{(2)}(z) + \dots \quad (24)$$

The exact evolution kernels  $P_{ab}$  have been known up to three loops (NNLO) since 2004 (Moch, Vermaseren, and Vogt, 2004; Vogt, Moch, and Vermaseren, 2004), with active efforts now under way on computing the four-loop terms  $P_{ab}^{(4)}$ ; see Ueda (2018). It is significant that the functions  $P_{a\hat{a}}^{(n)}(z)$  do not depend on quark masses. In graphs for the PDFs, there are masses in quark propagators,  $(\not{k} + m)/(k^2 - m^2 + i\epsilon)$ . However, the ultraviolet poles of these graphs are determined by the behavior of the propagators for  $k \rightarrow \infty$ . In this limit, the masses do not contribute. This is an advantage of using the  $\overline{\text{MS}}$  scheme for renormalizing the PDFs.

## B. Infrared safety and factorization

PDFs are used to compute cross sections in collisions of a lepton with a hadron and collisions between two hadrons. We concentrate here on hadron-hadron collisions since these are currently the subject of investigation at the LHC. Lepton-hadron collisions, as in deeply inelastic scattering, are simpler.

In one sense, the use of PDFs to describe proton-proton collisions is simple. Suppose that we are interested in the cross section  $d^2\sigma/(dp_T dy)$ , to produce a jet with transverse momentum  $p_T$  and rapidity  $y$  plus anything else in the collision of a hadron of type  $A$  and a hadron of type  $B$ . Or suppose that we are interested in the cross section  $d\sigma/dy$ , to produce an on-shell Higgs boson with rapidity  $y$  plus anything else. We can consider many cases at once by saying that we are interested in a cross section  $\sigma[F]$  to measure a general observable  $F$  that is infrared safe in the sense that is explained in Sec. II.B.2. Then the PDFs relate  $\sigma[F]$  to a calculated cross section  $\hat{\sigma}_{ab}[F]$  for the collision of two partons. In its briefest form, the relation is

$$\sigma[F] \approx \sum_{a,b} \iint d\xi_a d\xi_b f_{a/A}(\xi_a, \mu^2) f_{b/B}(\xi_b, \mu^2) \hat{\sigma}_{ab}[F]. \quad (25)$$

Here we sum over the possible flavors  $a$  and  $b$  of partons that we might find in the respective hadrons. We integrate over the momentum fractions  $\xi_a$  and  $\xi_b$  of these partons. Then we multiply by  $\hat{\sigma}_{ab}[F]$ , which plays the role of a cross section for the collision of these partons to produce the final state that we are looking for.

We say that Eq. (25) expresses factorization.<sup>5</sup> Factorization seems simple, but it is not. First, it works only when the observable to be measured  $\sigma[F]$  has a certain property, “infrared safety.” Second, factorization is approximate, and we need to understand what is left out. Third, its validity is not self-evident, as one finds when trying to calculate  $\hat{\sigma}_{ab}[F]$  beyond the leading order and encountering infinities if the

<sup>5</sup>The word “factorization” is applied to many formulas in which a physical quantity is expressed as a convolution of a product of factors. Equation (25) is sometimes called inclusive collinear factorization.

calculation is not carefully formulated. Fourth, while  $\hat{\sigma}_{ab}[F]$  plays the role of a cross section to produce a certain final state in the collision of two partons, its calculation beyond the leading order in perturbative QCD actually involves subtractions, as discussed in Sec. II.B.2.

### 1. Kinematics

Consider a hard scattering process in the collisions of two high-energy hadrons  $A$  and  $B$ . The hadrons carry momenta  $P_A$  and  $P_B$ . The hadron energies are high enough that we can simplify the equations describing the collision kinematics by treating the colliding hadrons as massless. Then with a suitable choice of reference frame, the hadron momenta are

$$\begin{aligned} P_A &= (P_A^+, 0, \mathbf{0}), \\ P_B &= (0, P_B^-, \mathbf{0}). \end{aligned} \quad (26)$$

We then imagine a parton-level process in which a parton from hadron  $A$  with flavor  $a$  and momentum  $\xi_a P_A$  collides with a parton from hadron  $B$  with flavor  $b$  and momentum  $\xi_b P_B$ . This collision produces  $m$  partons with flavors  $f_i$  and momenta  $p_i$ . Each final-state parton has rapidity  $y_i$  and transverse momentum  $\mathbf{p}_{i,\perp}$  so that the components of its momentum are

$$p_i = \left( e^{y_i} \sqrt{(\mathbf{p}_{i,\perp}^2 + m_i^2)/2}, e^{-y_i} \sqrt{(\mathbf{p}_{i,\perp}^2 + m_i^2)/2}, \mathbf{p}_{i,\perp} \right). \quad (27)$$

Then momentum conservation gives us

$$\begin{aligned} \sum_{i=2}^m \mathbf{p}_{i,\perp} &= -\mathbf{p}_{1,\perp}, \\ \sum_{i=1}^m e^{y_i} \sqrt{(\mathbf{p}_{i,\perp}^2 + m_i^2)/2} &= \xi_a P_A^+, \\ \sum_{i=1}^m e^{-y_i} \sqrt{(\mathbf{p}_{i,\perp}^2 + m_i^2)/2} &= \xi_b P_B^-. \end{aligned} \quad (28)$$

### 2. Infrared safety

For the factorization to work, the observable should be infrared safe. The basic physical idea for this was introduced by [Sterman and Weinberg \(1977\)](#). We will follow the development of [Kunszt and Soper \(1992\)](#) and define what we mean by measuring the cross section for an observable  $F$  and what it means for  $F$  to be infrared safe. To keep the discussion simple, we temporarily assume that all of the partons involved are light quarks and gluons, which we consider to be massless, and that the observable does not distinguish the flavors of the partons. For this simple case, we express the parton-level cross section for an observable  $F$  using the definition

$$\begin{aligned} \hat{\sigma}_{ab}[F] &= \frac{1}{2!} \int dy_1 dy_2 d\mathbf{p}_{2,\perp} \frac{d\hat{\sigma}_2}{dy_1 dy_2 d\mathbf{p}_{2,\perp}} F_2(p_1, p_2) \\ &+ \frac{1}{3!} \int dy_1 dy_2 dy_3 d\mathbf{p}_{2,\perp} d\mathbf{p}_{3,\perp} \\ &\times \frac{d\hat{\sigma}_3}{dy_1 dy_2 dy_3 d\mathbf{p}_{2,\perp} d\mathbf{p}_{3,\perp}} F_3(p_1, p_2, p_3) + \dots \end{aligned} \quad (29)$$

Here we start with the cross section to produce  $m$  partons with momenta  $\{p_1, \dots, p_m\}$ . We multiply the cross section by a function  $F_m(p_1, \dots, p_m)$  that specifies the measurement that we want to make on the final-state partons. These functions are taken to be symmetric under interchange of their arguments. Accordingly, we divide by the number  $m!$  of permutations of the parton labels. We integrate over the momenta of the final-state partons. The transverse momentum of parton 1 and the needed momentum fractions for the incoming partons are determined by Eq. (28). Finally, we sum over the number  $m$  of final-state partons.

An example may be useful. If we want to evaluate the cross section for the sum of the absolute values of the transverse momenta of the partons to be bigger than 100 GeV, then we choose a step function  $F_m = \theta(\sum_{i=1}^m |\mathbf{p}_{i,\perp}| > 100 \text{ GeV})$ . More generally, the  $F_m$  for jet cross sections are made of step functions and delta functions constructed according to the jet algorithm used.

Infrared safety is a property of the functions  $F_m$  that relates each function  $F_{m+1}(p_1, \dots, p_m, p_{m+1})$  to the function  $F_m(p_1, \dots, p_m)$  with one fewer parton. There are two requirements needed for  $F$  to be infrared safe.

First, consider the following limit in which partons  $m+1$  and  $m$  become collinear:

$$\begin{aligned} p_{m+1} &\rightarrow z \tilde{p}_m, \\ p_m &\rightarrow (1-z) \tilde{p}_m. \end{aligned} \quad (30)$$

Here  $\tilde{p}_m$  is a lightlike momentum and  $0 \leq z \leq 1$ . Therefore,  $p_m + p_{m+1} \rightarrow \tilde{p}_m$ . We can concentrate only on partons with labels  $m+1$  and  $m$  because the functions  $F$  are assumed to be symmetric under interchange of the parton labels. For  $F$  to be infrared safe, we demand that

$$F_{m+1}(p_1, \dots, p_{m-1}, p_m, p_{m+1}) \rightarrow F_m(p_1, \dots, p_{m-1}, \tilde{p}_m) \quad (31)$$

in the collinear limit [Eq. (30)].

Second, also consider the following limit in which parton  $m+1$  becomes collinear to one of the beams:

$$p_{m+1} \rightarrow \lambda P_A \quad (32)$$

or

$$p_{m+1} \rightarrow \lambda P_B. \quad (33)$$

Here  $\lambda \geq 0$ . When  $\lambda = 0$ , parton  $m+1$  is simply becoming infinitely soft. For  $F$  to be infrared safe, we demand that

$$F_{m+1}(p_1, \dots, p_m, p_{m+1}) \rightarrow F_m(p_1, \dots, p_m) \quad (34)$$

in either limit (32) or limit (33).

When the partonic masses are negligible, infrared safety means that the result of the measurement is not sensitive to whether or not one parton splits into two almost collinear partons, and it is not sensitive to any partons that have small momenta transverse to the beam directions.

Sometimes an observable  $F$  with this property is referred to as infrared and collinear safe instead of just infrared (IR) safe. The meaning is the same.

Cross sections to produce jets are infrared safe as long as we use a suitable algorithm, encoded in the functions  $F_m$ , to define jets. However, the cross section to produce a jet containing a specified number of charged particles would not be infrared safe: a collinear splitting of a gluon to a quark plus an antiquark increases the number of charged particles by 2.

In Eq. (29), we have cross sections to produce specified partons. This is the formula that we use to construct  $\hat{\sigma}_{ab}[F]$  on the right-hand side of Eq. (25). On the left-hand side of Eq. (25), we measure the cross section  $\sigma[F]$  experimentally by using the same functions  $F_m$  applied to hadrons instead of partons. Relating a hadron cross section to a parton cross section is evidently an approximation, so there is an error in Eq. (25) no matter how many orders of perturbation theory we use in the calculation. The physical idea is that the infrared-safe measurement involves a large scale  $Q$ , which is incorporated into the functions  $F_m$ . Processes that involve scales  $M$  with  $M \ll Q$  do not substantially affect the measurement. In particular, combining partons to form hadrons involves scales  $M \sim 0.3$  GeV. Thus, turning partons into hadrons does not substantially affect a measurement with a much larger scale  $Q$ .

While we can best understand the idea of IR safety by sticking to massless partons and flavor-independent measurement functions, it is important to keep in mind that a quark does have a mass, however small or large; QCD factorization need not assume that the partons are massless. For quark masses  $m_i$  that are small compared to the scale  $Q$  of the measurement that we have in mind, we amend the definition of infrared safety of the functions  $F_m$  to include taking the limit  $m_i \rightarrow 0$ . Sometimes a quark mass is comparable to the scale  $Q$ . For instance, this typically applies to top quarks. It also applies to, say, charm quarks when we take  $Q$  to be only a couple of GeV. In such cases, we need not consider limits  $m_i \rightarrow 0$  for “heavy” quarks and can operate with parton distribution functions only for the “light” quarks; see Sec. II.C.

### 3. Factorization

With the needed preparation accomplished, we can now state how the PDFs are used to calculate the cross section for whatever observable  $F$  we want, as long as  $F$  is infrared safe. For this condition to apply, the observable  $F$  must be sufficiently inclusive. The formula we use was stated in Eq. (25), and we restate it here in a slightly more detailed form:

$$\sigma[F] = \sum_{a,b} \int d\xi_a \int d\xi_b f_{a/A}(\xi_a, \mu_F^2) f_{b/B}(\xi_b, \mu_F^2) \times \hat{\sigma}_{a,b,\xi_a,\xi_b,\mu_F^2}[F] + \mathcal{O}(M/Q). \quad (35)$$

Our convention in Eq. (35) is to use the name  $\mu_F$ , the “factorization scale,” for the scale parameter in the parton distribution functions. The physical cross section cannot depend on  $\mu_F$ , so  $\hat{\sigma}$  must also depend on  $\mu_F$ . Note that in Eq. (35), we have not yet expanded in powers of  $\alpha_s$ .

The intuitive basis for Eq. (35) is simple. The factor  $f_{a/A}(\xi_a, \mu_F^2) d\xi_a$  plays the role of the probability to find a parton of flavor  $a$  in a hadron of flavor  $A$ . For the other hadron, the corresponding probability is  $f_{b/B}(\xi_b, \mu_F^2) d\xi_b$ . Then  $\hat{\sigma}[F]$  plays the role of a cross section to obtain the observable  $F$  from the scattering of these partons, as given in Eq. (29). Naturally, this parton-level cross section depends on the parton variables  $a, b, \xi_a, \xi_b$ , as indicated by the subscript notation. Here the differential cross sections to produce  $m$  final-state partons contain delta functions that relate the momentum fractions  $\xi_a$  and  $\xi_b$  to the final-state parton momenta, according to Eq. (28).

A similar formula applies for lepton-hadron scattering, a process that provides important constraints on the PDF parametrizations; see Sec. II.D. Then there is only one hadron in the initial state, so the following formula is simpler:

$$\sigma[F] = \sum_a \int d\xi f_{a/A}(\xi, \mu_F^2) \hat{\sigma}_{a,\xi,\mu_F^2}[F] + \mathcal{O}(M/Q). \quad (36)$$

The cross section  $\hat{\sigma}[F]$  in Eq. (35) [or Eq. (36)] has a perturbative expansion in powers of  $\alpha_s(\mu_R^2)$ , where the renormalization scale  $\mu_R^2$  can be chosen independently from the factorization scale  $\mu_F^2$ . That is,

$$\hat{\sigma}_{a,b,\xi_a,\xi_b,\mu_F^2}[F] = \left[ \frac{\alpha_s(\mu_R^2)}{2\pi} \right]^B \hat{\sigma}_{a,b,\xi_a,\xi_b,\mu_F^2,\mu_R^2}^{(B)}[F] + \left[ \frac{\alpha_s(\mu_R^2)}{2\pi} \right]^{B+1} \hat{\sigma}_{a,b,\xi_a,\xi_b,\mu_F^2,\mu_R^2}^{(B+1)}[F] + \dots \quad (37)$$

Here  $B$  is the integer that tells us how many powers of  $\alpha_s$  appear in the Born-level cross section: e.g., 0 for  $Z$  boson production, and 2 for two jet production. Perturbative calculations can be at lowest order (LO), corresponding to one term in the expansion, next-to-lowest order (NLO) with two terms, sometimes NNLO, and, in general,  $N^k$ LO.

When  $\hat{\sigma}$  is expanded in powers of  $\alpha_s$ , as in Eq. (37), our convention is to use the name  $\mu_R$ , the “renormalization scale,” for the scale parameter in  $\alpha_s$ . The physical cross section cannot depend on  $\mu_R$ , so the functions  $\hat{\sigma}^{B+k}$  in Eq. (37) must also depend on  $\mu_R$ . (Other conventions for the precise meaning of  $\mu_F$  and  $\mu_R$  are possible.)

One useful property of Eqs. (36) and (37) is that the dependence of the calculated cross section on  $\mu_F^2$  and  $\mu_R^2$  diminishes as we go to higher orders. Indeed, the cross section in nature,  $\sigma[F]$ , does not depend on  $\mu_F^2$  or  $\mu_R^2$ . Thus, if we calculate to order  $\alpha_s^{B+k}$ , the derivative of the calculated cross section with respect to  $\mu_F^2$  and  $\mu_R^2$  is of order  $\alpha_s^{B+k+1}$ . Because of this property, one often uses the effect of varying  $\mu_F$  or  $\mu_R$  by a fixed factor (e.g., 2 or 1/2) to provide an estimate of the error caused by calculating only to a finite perturbative order.

Note that, for terms of the order of  $\alpha_s^{B+k}$  in  $\hat{\sigma}$  to match the terms of the order of  $\alpha_s^k$  in the evolution equation for the parton distributions, we need to include at least the terms up to  $\alpha_s^k P_{a\hat{a}}^{(k)}(z)$  in Eqs. (23) and (24) giving the evolution of the PDFs. Since the lowest-order term in the evolution kernel is  $P_{a\hat{a}}^{(1)}(z)$ , including terms up to  $P_{a\hat{a}}^{(k)}(z)$  is referred to as  $N^{k-1}$ LO

evolution, while we say that including terms up to  $\alpha_s^{B+k}$  in the partonic cross section gives an  $N^k$ LO calculation. Thus, for example, if we have a NNLO cross-section calculation, we do not obtain the proper matching for the  $\mu_F$  scale dependence unless we have at least NLO evolution for the PDFs. Normally, one uses even higher-order evolution for the PDFs because this evolution determines the accuracy with which we know the PDFs at a high scale, given experimental inputs at much lower scales.

The error terms  $\mathcal{O}(M/Q)$  in Eqs. (35) and (36) arise from power-suppressed contributions that are beyond the accuracy of the factorized representation (Collins, Soper, and Serman, 1983, 1985, 1988; Bodwin, 1985; Collins, 1998, 2013). No matter how many terms are included in  $\hat{\sigma}$ , there are contributions that are left out. These terms are suppressed by a power of  $M \sim 1$  GeV divided by a large scale parameter  $Q$  that characterizes the hard scattering process to be measured. Here we choose 1 GeV as a nominal scale for hadronic bound state physics. This value is somewhat smaller than the scale  $\mu$  at which  $\alpha_s(\mu^2) = 1/2$ , somewhat larger than the proton mass and about 4 times larger than the inverse of the radius of a proton.

When data are included in a PDF fit but, for these data,  $M/Q$  is not small enough to be completely negligible, it may be useful to include a nonperturbative model for the power-suppressed corrections. The model can have one or more parameters that can be fit to the data. For instance, one can add a contribution to the cross section that is proportional to  $\kappa^2/Q^2$ , where  $\kappa$  is a parameter to be fit. This possibility is especially relevant for deeply inelastic scattering (see Sec. II.D), for which one can make use of a theoretical expansion known as the operator product expansion (Wilson, 1969) to suggest the form of the first power-suppressed contribution; see such fits, for example, in Alekhin, Blumlein, and Moch, 2012 and Alekhin *et al.* (2017).

The power-suppressed contributions in Eq. (35) arise from the approximations needed to obtain the result. For instance, if a loop momentum  $l$  flows through the wave function of quarks in a proton, we have to neglect  $l$  compared to the hard momenta, say, the transverse momentum of an observed jet. We can illustrate the issue of power corrections and, at the same time, see something about the interplay of ultraviolet and infrared singularities by looking at a simple model for an integral that might occur in a calculation.

Consider a model integral for a correction to the Born cross section for a process involving only one initial-state hadron. In this model, the Born cross section is

$$I_{\text{Born}} = \int_0^\infty dl^2 |\psi(l^2)|^2 \times 1, \quad (38)$$

where  $l^2$  represents the transverse momentum of a quark that is restricted to be of the order of 1 GeV<sup>2</sup> by a hadronic wave function  $|\psi(l^2)|^2$ . The Born-level PDF is  $\int_0^\infty dl^2 |\psi(l^2)|^2$ . The Born hard scattering function is simply 1. At order  $\alpha_s$ , we should have an  $\alpha_s$  correction to the Born PDF times the Born hard scattering function, plus the Born PDF times an  $\alpha_s$  correction to the hard scattering function.

In our model, the order  $\alpha_s$  correction to the cross section is

$$I = \int_0^\infty dl^2 |\psi(l^2)|^2 \alpha_s \int_0^\infty dk^2 \left[ \frac{\mu_F^2}{k^2} \right]^\epsilon \frac{1}{k^2 + l^2} \frac{Q^2}{k^2 + Q^2}. \quad (39)$$

Here  $Q^2$  is the ‘‘hard scale,’’ with  $l^2 \ll Q^2$ . We supply a factor  $[\mu_F^2/k^2]^\epsilon$  that mimics dimensional regularization in  $n = 4 - 2\epsilon$  dimensions. This integral is, however, convergent in the IR and in the UV, so we could simply set  $\epsilon \rightarrow 0$  after its computation.

For  $k^2 \gg l^2$ , we could approximate  $1/(k^2 + l^2) \rightarrow 1/k^2$ . This would define a part of the integral  $I$  that represents the probability distribution  $|\psi(l^2)|^2$  times an order  $\alpha_s^1$  contribution to the hard scattering function,

$$I_{\text{UV}} = \int_0^\infty dl^2 |\psi(l^2)|^2 \alpha_s \left( \int_0^\infty dk^2 \left[ \frac{\mu_F^2}{k^2} \right]^\epsilon \frac{1}{k^2} \frac{Q^2}{k^2 + Q^2} + \frac{1}{\epsilon} \right). \quad (40)$$

The integral is IR divergent. We have subtracted its IR pole, which is proportional to  $-1/\epsilon$ .

For  $k^2 \ll Q^2$ , we could approximate  $Q^2/(k^2 + Q^2) \rightarrow 1$ . The integral then defines an  $\alpha_s$  correction to the PDF times the Born hard scattering function,

$$I_{\text{IR}} = \int_0^\infty dl^2 |\psi(l^2)|^2 \alpha_s \left( \int_0^\infty dk^2 \left[ \frac{\mu_F^2}{k^2} \right]^\epsilon \frac{1}{k^2 + l^2} - \frac{1}{\epsilon} \right) \times 1. \quad (41)$$

The PDF integral is UV divergent, so we have ‘‘renormalized’’ it by subtracting its UV pole, which is proportional to  $+1/\epsilon$ .

We can also use

$$I_0 = \int_0^\infty dl^2 |\psi(l^2)|^2 \alpha_s \int_0^\infty dk^2 \left[ \frac{\mu_F^2}{k^2} \right]^\epsilon \frac{1}{k^2} = 0. \quad (42)$$

The  $k$  integral equals zero because it equals  $1/\epsilon - 1/\epsilon$ .

A simple calculation using the integrands of our integrals gives

$$I = I_{\text{UV}} + I_{\text{IR}} - I_0 - \int_0^\infty dl^2 |\psi(l^2)|^2 \times \frac{l^2}{Q^2} \alpha_s \int_0^\infty dk^2 \left[ \frac{\mu_F^2}{k^2} \right]^\epsilon \frac{1}{k^2 + l^2} \frac{Q^2}{k^2 + Q^2}. \quad (43)$$

Thus, our integral can be decomposed into the UV integral (including an IR subtraction), the IR integral (including its UV renormalization subtraction), and a remainder that is suppressed by a factor of  $l^2/Q^2$ , where  $l^2$  is of the order of 1 GeV<sup>2</sup>. This is the power-suppressed correction.

Notice that dimensional regularization with factorization scale  $\mu_F^2$  followed by subtracting  $1/\epsilon$  poles serves two functions. The integral  $I_{\text{UV}}$  in Eq. (40) is a model for how one calculates the hard scattering function at NLO. This integral has an IR divergence, which is removed by subtracting its pole. This subtraction matches the subtraction necessary to renormalize UV divergence in the correction to the model parton distribution function in Eq. (41).

This model calculation is somewhat misleading because it suggests that PDFs have a useful perturbative expansion. They do not. To compute a hard cross section in real QCD, one can replace the distribution of partons in a proton by the distribution of partons in a parton, which is singular but perturbatively calculable; see Collins (2013) for details.

Not much is known about the general form of the power corrections for hadron-hadron collisions.<sup>6</sup> It is important that they are there, but if  $Q$  is of the order of hundreds of GeV, then the power corrections are completely negligible. However, if  $Q$  is of the order of 5 GeV, then we ought not to claim 1% accuracy in the calculation of  $\sigma[F]$ , no matter how many orders of perturbation theory we use.

Equation (35), representing inclusive collinear factorization, is the basis of every prediction for hard processes at hadron colliders like the LHC, including both standard model processes and processes that might produce new heavy particles. As far as we know, it is a theorem for infrared-safe QCD observables dependent on energy-momentum variables that are of the same order of magnitude. There are other formulas in QCD that go under the name of factorization and typically apply either to amplitudes rather than cross sections or to observables dependent on several momentum variables of disparate orders of magnitude. These include  $k_T$  factorization and soft-collinear-effective-theory factorization. These other forms of factorization may fail, thus requiring a more complex analysis, or are at least more subject to doubt than Eq. (35); see, for example, Collins and Qiu, 2007; Catani, Florian, and Rodrigo, 2012; Forshaw, Seymour, and Siodmok, 2012; Rothstein and Stewart, 2016; Schwartz, Yan, and Zhu, 2018).

Early attempts to establish inclusive collinear factorization (Amati, Petronzio, and Veneziano, 1978; Ellis *et al.*, 1979) were instructive but incomplete. Later proofs of Eq. (35) (Collins, Soper, and Serman, 1983, 1985, 1988; Bodwin, 1985; Collins, 1998) are far from simple. They could perhaps benefit from more scrutiny than they have received. One issue is that the published proofs have considered only the Drell-Yan process, not more complex processes like jet production. A more serious issue is that there is no known general method that can deal with the boundaries between integration regions in the Feynman diagrams. On the other hand, any breakdown in the inclusive collinear factorization in Eq. (35) could lead to infinities in calculations of  $\hat{\sigma}$ , and no problems have been observed to date even in N<sup>3</sup>LO calculations (Anastasiou *et al.*, 2015).

### C. Treatment of heavy quarks

To accurately describe data at energies from one to thousands of GeV, modern global PDF fits not only change the number  $N_f$  of active flavors depending on the scales  $\mu_{R,F}^2$  but also retain relevant quark mass dependence in the hard scattering cross sections. The most comprehensive approach

to do this is to work in one of the GM-VFN factorization schemes (Aivazis *et al.*, 1994; Buza *et al.*, 1998; Thorne and Roberts, 1998; Chuvakin, Smith, and van Neerven, 2000; Kramer and Spiesberger, 2004; Kniehl *et al.*, 2005a, 2005b; Thorne, 2006; Forte *et al.*, 2010. The massive fixed-flavor number schemes Gluck, Hoffmann, and Reya, 1982; Gluck, Godbole, and Reya, 1988) are also applicable under the right circumstances and may result in simpler predictions. Such computations are a complex subject that we cannot cover in any depth here. We illustrate some of the key ideas by mostly following Krämer, Olness, and Soper (2000).

Consider the perturbative calculation of an infrared-safe cross section  $\sigma$  with scale  $Q^2$  when  $Q^2 \gg m_i^2$ , where  $i$  denotes any of the  $u, d, s, c,$  and  $b$  quarks. We can greatly simplify this calculation by neglecting masses of the five quarks. But what if  $Q$  is high enough and we expect that top quarks contribute either in the final state or in the virtual loop corrections? We rarely can set  $m_t = 0$ , as  $m_t \approx 174$  GeV is so large that we seldom have  $Q^2 \gg m_t^2$ , even at the LHC.

There is a simple answer: in the range of energies comparable to  $m_t$ , we can use the  $\overline{\text{MS}}$  scheme with five active quark flavors  $u, d, s, c,$  and  $b$ . Top quarks are included in the relevant Feynman graphs, but in accord with the Collins-Wilczek-Zee (CWZ) prescription (Collins, Wilczek, and Zee, 1978) we use the zero-momentum subtraction, instead of  $\overline{\text{MS}}$  subtraction, for the UV renormalization of loop subgraphs with top-quark lines. Then, terms involving  $m_t$  appear only in the hard cross section  $\hat{\sigma}$  of Eq. (35).

One consequence of this is that the evolution equation for  $\alpha_s(\mu_R^2)$  uses the five-flavor beta function. Radiative contributions to the top-quark mass also need to be renormalized, as discussed later. The five-flavor scheme introduces nonzero PDFs  $f_{i/p}(\xi, \mu_F^2)$  for  $i \in \{g, u, \bar{u}, d, \bar{d}, \dots, b, \bar{b}\}$ . There are no parton distributions for  $i = t$  or  $i = \bar{t}$  in this scheme. The PDFs  $f_{i/p}(\xi, \mu_F^2)$  are as previously defined and evolve using five-flavor DGLAP kernels. Another (not obvious) result is that top-quark contributions are negligible in the limit  $Q^2 \ll m_t^2$ : that is, top quarks decouple when the momentum scale  $Q$  of the problem is much smaller than the top-quark mass.

In a general case, we can distinguish among several versions of an  $N_f$ -flavor scheme. In the zero-mass (ZM) scheme, only the Feynman graphs with massless active quarks  $q_i$  with  $i \leq N_f$  are included in  $\hat{\sigma}$  in the factorized hadronic cross section (35). In the fixed-flavor number (FFN) scheme, the massless active quark contributions for  $i \leq N_f$  are included as in the ZM scheme, and the Feynman graphs with massive inactive (anti)quarks  $q_i$  with  $i > N_f$  are included only in the hard cross section  $\hat{\sigma}$ . Finally, the most complete GM scheme retains non-negligible quark mass terms from both active and inactive quarks in all parts of the hadronic cross section.

When discussing the region  $Q^2 \gg m_b^2$ , we can use the five-flavor FFN scheme and set masses of  $u, d, s, c,$  and  $b$  quarks to zero. Consider now the region  $Q^2 \sim m_b^2$  and  $Q^2 \gg m_c^2$ . Here we can use the four-flavor FFN scheme. There are parton distributions for flavors  $u, d, s,$  and  $c$ , and these quarks are considered massless, while there are no parton distributions for  $b$  and  $\bar{b}$ . In fact, the four-flavor scheme is also fine for the

<sup>6</sup>See, however, Qiu and Serman (1991a, 1991b).

subregion of  $Q^2 \gg m_b^2$  in which  $\alpha_s \log(Q^2/m_b^2) \ll 1$ , where we do not need to sum logs of  $Q^2/m_b^2$ .<sup>7</sup>

We now have two possible FFN schemes for calculating a cross section at  $m_b^2/Q^2 \ll 1$ , but  $\alpha_s \log(Q^2/m_b^2) \ll 1$ : the five-flavor scheme with  $\alpha_s^{(5)}(\mu^2)$  and  $f_{i/p}^{(5)}(\xi, \mu^2)$ , and the four-flavor scheme with  $\alpha_s^{(4)}(\mu^2)$  and  $f_{i/p}^{(4)}(\xi, \mu^2)$ . (Here we set  $\mu_R^2 = \mu_F^2 = \mu^2$ .) The physical predictions must be the same, order by order in  $\alpha_s$ , in either scheme. This condition gives us matching relations between  $\alpha_s$  and the PDFs in the four-flavor and five-flavor schemes. At lowest order in  $\alpha_s$ , these relations are simple. We should not use  $\alpha_s^{(5)}(\mu^2)$  or  $f_{i/p}^{(5)}(\xi, \mu^2)$  for calculating physical cross sections unless  $\mu^2 \gg m_b^2$ , but if we simply use their analytic forms for  $\mu^2 = m_b^2$ , we have

$$\begin{aligned} \alpha_s^{(5)}(m_b^2) &= \alpha_s^{(4)}(m_b^2), \\ f_{i/p}^{(5)}(\xi, m_b^2) &= f_{i/p}^{(4)}(\xi, m_b^2) \quad \text{for } i \notin \{b, \bar{b}\}, \\ f_{b/p}^{(5)}(\xi, m_b^2) &= f_{b/p}^{(5)}(\xi, m_b^2) = 0. \end{aligned} \quad (44)$$

At higher orders of  $\alpha_s$ ,  $f_{b/p}^{(5)}(\xi, m_b^2) \neq 0$ , and the matching conditions are different and depend on whether  $m_b$  is an  $\overline{\text{MS}}$  or pole mass (Buza *et al.*, 1996; Bierenbaum, Blumlein, and Klein, 2007; Ablinger *et al.*, 2017). Then, to obtain the  $f_{i/p}^{(5)}(\xi, \mu^2)$  for  $\mu^2 > m_b^2$ , we solve the  $N_f = 5$  evolution equation with a boundary condition [Eq. (44)] at  $\mu^2 = m_b^2$ . One can also use a different scale  $\mu_b^2 \neq m_b^2$ , as well as alternative prescriptions, for the matching (Bertone *et al.*, 2017, 2018; Kusina *et al.*, 2013).

We can derive analogous matching relations between the three-flavor and four-flavor schemes at  $\mu = m_c \approx 1.3$  GeV. The full range  $\mu^2 \geq m_c^2$  is then described by a sequence of the schemes with  $N_f = 3, 4$ , and 5 that together compose a VFN scheme.

The scheme described thus far is conceptually simple but involves awkward switches between different values of  $N_f$  at still unspecified energy values. For instance, suppose we wish to switch between the  $N_f = 3$  FFN calculation and the  $N_f = 4$  calculation at a switching value  $\mu = Q \equiv \mu_4$  somewhere above  $m_c$ . At the corresponding value of  $Q = \mu_4$ , the calculated cross section is discontinuous, which is detrimental in a PDF fit. If  $\mu_4^2$  is too close to  $m_c^2$ , the  $N_f = 4$  FFN cross section misses important terms proportional to  $m_c^2$ . If  $\mu_4^2$  is too far above  $m_c^2$ , the missing higher-order collinear logarithms  $\alpha_s \log(Q^2/m_c^2)$  make the  $N_f = 3$  FFN cross section unreliable in the upper range of  $Q$ .

To avoid this, many global fits use some version of a GM-VFN scheme; see the previous references. These schemes allow for  $N_f$  switching exactly at the corresponding quark mass,  $\mu_i = m_i$ , and they achieve a smooth interpolation

<sup>7</sup>One can also use a three-flavor FFN scheme. For instance, this scheme was used for fitting  $c$ -quark and  $b$ -quark production in deeply inelastic scattering by Alekhin *et al.* (2017). It introduces PDFs for  $u$ ,  $d$ , and  $s$  quarks, which are treated as massless. The  $c$  and  $b$  quarks appear, with their masses, in  $\hat{\sigma}$ .

between the switching points. In such a scheme, one retains numerically non-negligible masses for any quark type in  $\hat{\sigma}$ . For instance, the ZM-VFN and GM-VFN schemes differ only in the treatment of the terms of the order of  $\mathcal{O}(m_i^2/Q^2)$  in the short-distance cross sections. They have the same mass dependence of PDFs. We are simply including the essential  $\mathcal{O}(m_i^2/Q^2)$  terms in  $\hat{\sigma}$ .

The logic that we just outlined is closely followed by the simplified Aivazis-Collins-Olness-Tung (S-ACOT) scheme (Aivazis *et al.*, 1994; Krämer, Olness, and Soper, 2000). It was proved to all  $\alpha_s$  orders by Collins (1998) and applied in DIS up to NNLO (Guzzi *et al.*, 2012) for use in fits by the CTEQ-TEA Collaboration. In the ACOT family of schemes, the flavor number in  $\alpha_s$ , masses, and PDFs is varied according to the CWZ prescription. Other GM-VFN schemes are perturbatively equivalent to the (S-)ACOT scheme. Comparisons between the GM-VFN approaches were made by Alekhin *et al.* (2010), Binoth *et al.* (2010), Guzzi *et al.* (2012), and Gao, Harland-Lang, and Rojo (2018).

We have discussed variable-flavor-number schemes with up to five active flavors. At a 100 TeV collider, introduction of a top-quark PDF may also be warranted; see Sec. III.D in Mangano *et al.* (2017). Groups doing global fitting often also present PDFs determined in fixed-flavor number schemes.

## 1. Heavy-quark masses

All nonzero quark masses in the calculation require renormalization. Either the heavy-quark  $\overline{\text{MS}}$  masses or the pole masses provided by Tanabashi *et al.* (2018) are used as input parameters when fitting the PDFs. Their values can even be extracted from PDF fits (Gao, Guzzi, and Nadolsky, 2013; Ball *et al.*, 2016; Alekhin *et al.*, 2017; Gizhko *et al.*, 2017). The  $\overline{\text{MS}}$  mass for a charm quark is better defined in PQCD and more precisely constrained by the world data. On the other hand, some PQCD calculations in the global fits use the pole mass as the input. Perturbative relations to convert the  $\overline{\text{MS}}$  mass into the pole mass, or back, are known to a high order in PQCD (Chetyrkin, Kuhn, and Steinhauser, 2000).

## 2. Fitted charm quarks

The PDF for charm quarks is of special interest. Consider first the most standard treatment. Suppose, hypothetically, that we have data for a cross section at a scale  $Q$  around  $m_c$ , and that in this cross section the hard interaction produces a charm quark and an antiquark in the final state. The three-flavor scheme is useful for predicting this cross section, containing nonperturbative parametrizations of  $f_{i/p}^{(3)}(\xi, \mu_0^2)$  for  $i \in \{g, u, \bar{u}, d, \bar{d}, s, \bar{s}\}$ . There is no PDF for charm quarks. If we now change to a four-flavor scheme, we have PDFs  $f_{i/p}^{(4)}(\xi, \mu^2)$  for  $i \in \{g, u, \bar{u}, d, \bar{d}, s, \bar{s}, c, \bar{c}\}$ . These would be obtained by matching to the three-flavor PDFs at scale  $\mu^2 = m_c^2$ . For the charm quark, the matching gives  $f_{c/p}^{(4)}(\xi, m_c^2) = 0$  at leading order. Matching gives us what may be called perturbative charm because the charm PDF for  $\mu^2 > m_c^2$  is produced solely by the perturbative matching and DGLAP evolution.

Note that the leading-order  $3 \rightarrow 4$  matching condition  $f_{c/p}^{(4)}(\xi, m_c^2) = 0$  can be traced to the assumption that the power-suppressed corrections of the order of  $M^2/m_c^2$  to a hypothetical cross section at  $Q^2 \sim m_c^2$  would be negligible. If we think that this is perhaps not the case, we could allow  $f_{c/p}^{(4)}(\xi, m_c^2)$  to be nonzero. Then solving the DGLAP evolution equation with the boundary condition at  $\mu^2 = m_c^2$  would result in a possibly nonzero value of  $f_{c/p}^{(4)}(\xi, m_c^2)$ , which in turn influences the prediction of experimental results at scales  $Q^2$  larger than  $m_c^2$ , even  $Q^2 \gg m_c^2$ . If we put these experimental results into the PDF-fitting program, we can fit  $f_{c/p}^{(4)}(\xi, m_c^2)$ . This gives us fitted charm.

The logic of fitted charm just outlined implies that if we do consider experiments at  $Q^2 \sim m_c^2$ , we should allow for fitted nonperturbative charm quark contributions to the predictions for these experiments. If we use a four-flavor scheme with  $m_c \neq 0$  as described later, these contributions can come from  $f_{c/p}^{(4)}(\xi, \mu^2)$  for  $\mu^2 \sim m_c^2$ .

Representative Feynman diagrams that contribute to the fitted component of the charm PDF were given by Hou *et al.* (2018) in their Fig. 3. These diagrams introduce contributions of the order of  $\mathcal{O}(M^2/m_c^2)$  to the charm PDF. The available data can accommodate or even mildly prefer contributions  $f_{c/p}^{(4)}(\xi, m_c^2)$  carrying up to about 1% of the net proton's momentum (Jimenez-Delgado *et al.*, 2015; Ball *et al.*, 2016; Hou *et al.*, 2018).

#### D. Deeply inelastic scattering

We conclude the theory overview with a discussion of DIS  $\ell(k) + A(p_A) \rightarrow \ell'(k') + X$ , an important class of processes for the determination of PDFs (Devenish and Cooper-Sarkar, 2004). Here  $\ell$  and  $\ell'$  can be electrons, neutrinos, or muons with specified momenta  $k$  and  $k'$ . Hadron  $A$  is a proton, nucleus, or pion with momentum  $p_A$ , and  $X$  stands for unobserved particles. The interaction between the leptons and the hadron proceeds with an exchange of a virtual  $\gamma^*$ ,  $Z$ , or  $W$  boson with momentum  $q = k - k'$ .

Not only were the measurements in DIS historically influential in the development of QCD, while diverse DIS data from HERA and fixed targets still serve as the backbone for global fits, but projections (Abdul Khalek *et al.*, 2018; Hobbs *et al.*, 2019) also show that DIS data will continue to provide essential constraints on the PDFs in the high-luminosity LHC era.

It is conventional to define three Lorentz-invariant variables,

$$Q^2 = -q^2, \quad x_{bj} = \frac{Q^2}{2P_A \cdot q}, \quad y = \frac{P_A \cdot q}{P_A \cdot k}. \quad (45)$$

“Deeply inelastic” means that  $Q^2 \gg 1 \text{ GeV}^2$  and that  $x_{bj}$  is not too small or too close to 1. The use of the variable  $x_{bj}$  was first suggested by Bjorken (1969), who proposed that the cross section would have simple properties in the deeply inelastic limit.

If only the final-state lepton is observed, one often writes the spin-independent cross section as a linear combination of three *structure functions*  $F_1(x_{bj}, Q^2)$ ,  $F_2(x_{bj}, Q^2)$ , and  $F_3(x_{bj}, Q^2)$ . To determine the  $F_i$ 's, one needs data from two c.m. energies  $\sqrt{s}$ . Otherwise, an approximate assumption is needed to extract the  $F_i$ . The structure function  $F_3$  is nonzero only if the current violates parity.

The structure functions can be written in terms of PDFs in the form

$$F_i(x_{bj}, Q^2) = \sum_a \int d\xi f_{a/A}(\xi, \mu^2) \hat{F}_{i,a}\left(\frac{x_{bj}}{\xi}, \frac{\mu^2}{Q^2}\right) + \mathcal{O}(M/Q), \quad (46)$$

where  $M \sim 1 \text{ GeV}$ . Usually, one chooses  $\mu^2 = Q^2$ . This is really just another form of Eq. (36). It offers one nice result. In  $eA$  scattering with a virtual photon exchange from the electron, at lowest order in  $\alpha_s$ ,  $\hat{F}_2$  contains a delta function that sets  $\xi \rightarrow x_{bj}$ :

$$F_2(x_{bj}, Q^2) \approx \sum_i Q_i^2 x_{bj} f_i(x_{bj}, Q^2), \quad (47)$$

where  $Q_i$  is the electric charge (in units of  $e$ ) of partons of type  $i$ :  $Q_{u,c,t} = 2/3$ , and  $Q_{d,s,b} = -1/3$ . Because of the charge factors  $Q_i^2$ , neutral-current DIS via photon exchange is about 4 times more sensitive to up-type (anti)quark PDFs than to down-type ones. It is more difficult for DIS to constrain the  $d$ -quark and especially  $s$ -quark PDFs, so the uncertainties on these flavors tend to be higher than for  $u$  and  $c$  PDFs, as we see in Fig. 1. We caution, however, that the simple result in Eq. (47) does not hold beyond the lowest order, and that the structure functions are not to be confused with PDFs.

Note that we have used  $\xi$  for the momentum fraction argument of PDFs and  $x_{bj}$  for the kinematic variable defined in Eq. (45). In the particle physics literature, one often sees the notation  $x$  used for the PDF momentum fraction argument  $\xi$ . In fact, we sometimes use  $x$  with this meaning in this review. We generally use  $x_{bj}$  and not simply  $x$  for the kinematic variable defined in Eq. (45).

### III. STATISTICAL INFERENCE IN FITTING THE PARTON DISTRIBUTIONS

In this section, we derive the key statistical results relevant for the extraction of the PDFs from the experimental data. We use the simplest possible framework, in which experimental errors can be approximated as having a Gaussian distribution, and the theoretical predictions are approximated as linear functions of the parameters used to describe the parton distribution functions. This framework is sometimes called the Hessian method (Pumplin *et al.*, 2001) since a certain matrix called the Hessian matrix plays a prominent role. The Hessian approach is motivated by the observation that many essential features of the PDF fits are captured by assuming an approximately Gaussian behavior of the underlying probability distribution of the experimental data. The PDF functional forms can be determined (inferred) from the experimental data by applying Bayes's theorem (Alekhin,



1999), as is summarized in Sec. III.A. The Hessian approximation provides a simplified solution to the problem of Bayesian inference when the PDFs are well constrained, and the deviations from the most likely solution for PDFs are relatively small. The only nonstandard feature that we add is the inclusion of a set of parameters  $R_k$  that represent the possibility that the theory, with an ideal choice of parameters for the PDFs, is not perfect and may not fit the data exactly.

### A. Bayes's theorem

Fitting parton distributions to data involves accounting for the statistical and systematic errors in the data. Thus, we need a statistical analysis. For this, we use a Bayesian framework in this review. The alternative is a frequentist framework, but we find that the Bayesian framework is simple to understand and makes us more aware of assumptions that otherwise remain obscure.

We begin with Bayes's theorem. At its base, this is a simple matter of counting. Consider a population in which each individual can have one or both of two characteristics  $T_1$  and  $D$ . For a concrete example, the population might consist of people in California.  $T_1$  might be "has a certain genetic marker," while  $D$  might be "tests positive for this genetic marker." Denote by  $P(T_1)$  the probability that an individual has the characteristic  $T_1$ . That is,  $P(T_1)$  is the number of individuals with the characteristic  $T_1$  divided by the total number of individuals. Similarly, let  $P(D)$  be the probability that an individual has the characteristic  $D$ . Denote by  $P(T_1|D)$  the conditional probability that an individual who is known to have the characteristic  $D$  has the characteristic  $T_1$ . That is,  $P(T_1|D)$  is the number of individuals with both characteristics  $T_1$  and  $D$  divided by the number of individuals with the characteristic  $D$ . Similarly, let  $P(D|T_1)$  be the conditional probability that an individual who is known to have the characteristic  $T_1$  has the characteristic  $D$ . With these definitions, we have Bayes's theorem

$$P(T_1|D)P(D) = P(D|T_1)P(T_1). \quad (48)$$

By evaluating the likelihood  $P(D|T_1)$  according to an explicit prescription presented in Sec. III.B and by having estimates for  $P(D)$  and  $P(T_1)$ , we could infer the "posterior probability"  $P(T_1|D)$  by rearranging the factors in Eq. (48) as follows:

$$P(T_1|D) = \frac{P(D|T_1)P(T_1)}{P(D)}. \quad (49)$$

Now consider another characteristic  $T_2$ , for instance, "does not have the genetic marker." Then Bayes's theorem gives us

$$\frac{P(T_1|D)}{P(T_2|D)} = \frac{P(D|T_1)P(T_1)}{P(D|T_2)P(T_2)}. \quad (50)$$

Note that  $P(D)$  cancels in this equation. If we know the quantities on the right-hand side of Eq. (50), this tells us the relative probabilities of finding characteristics  $T_1$  and  $T_2$  among individuals who are known to have the characteristic  $D$ .

Equation (50) is directly applicable and useful in cases in which there is a large number of individuals, and we know the probabilities on the right-hand side of the equation. For instance, a physician of a patient who tests positive for a rare health condition will want to use Eq. (50) to help prescribe a specific treatment to address the condition.

We do, however, use Eq. (50) in a more subtle case. Suppose that there is only one individual. Suppose further that you have a subjective belief, based on your prior experience, that the probability that this individual has property  $T_i$  is  $P(T_i)$  for  $i \in \{1, 2\}$ . This "prior probability" could be formed based on your past observations.

Now suppose that the individual is observed to have property  $D$ . Assume that you know how to compute the probability  $P(D|T_i)$  for an individual to have property  $D$  if the individual has property  $T_i$ . You can turn this knowledge around with the help of Bayes's theorem [Eq. (49)] to calculate the probability  $P(T_i|D)$  for having property  $T_i$  on the condition that the individual was observed to have property  $D$ . You can also use Eq. (50) to compute the "posterior probability ratio"  $P(T_1|D)/P(T_2|D)$  by multiplying the prior probability ratio  $P(T_1)/P(T_2)$ , reflecting your knowledge before the measurement, by the ratio  $P(D|T_1)/P(D|T_2)$  of the likelihoods calculated for each property  $T_i$ . Then you declare an updated prior probability ratio

$$\left(\frac{P(T_1)}{P(T_2)}\right)_{\text{new}} \equiv \frac{P(T_1|D)}{P(T_2|D)} = \frac{P(D|T_1)}{P(D|T_2)} \left(\frac{P(T_1)}{P(T_2)}\right)_{\text{old}}. \quad (51)$$

In our physics application,  $T_1$  and  $T_2$  are possible theoretical models describing a physical system, or perhaps one model of the system with two choices of parameters. Even when we cannot interpret  $P(T_i)$  by counting the number of instances of the system in which  $T_i$  holds, we often have some idea about which theory is more likely, and we express this belief as the prior probabilities  $P(T_i)$ . The prior probabilities are often based on previous experiments but are partly subjective. Thus, your prior probabilities may not be the same as mine. Now we make a measurement and observe that the system has a property  $D$ . With these new data in mind, we update the probability ratio for the two theories according to Eq. (51).

Your new estimate is still partly subjective and may not agree with mine. However, as data accumulate, it frequently happens that the subjective nature of our probabilities ceases to really matter. Suppose, for example, that your prior probability ratio is  $P(T_1)/P(T_2) = 10$  and mine is  $P(T_1)/P(T_2) = 0.1$ . Then we do not agree which theory is more likely. However, suppose that, after a lot of data become available, the likelihood ratio for the data is  $P(D|T_1)/P(D|T_2) = 10^6$ . Then your posterior probability ratio is  $P(T_1|D)/P(T_2|D) = 10^7$ , while mine is  $P(T_1|D)/P(T_2|D) = 10^5$ . Both ratios are large and strongly disfavor theory  $T_2$  in comparison to theory  $T_1$ . At this point, I agree that you are right and we stop discussing the matter.

Notice that our subjective probability estimates change by the same factor, equal to the likelihood ratio  $P(D|T_1)/P(D|T_2)$ . We agree on this factor because we agree on the data  $D$  as well as on the calculation. For the agreement

to be possible, the theories must be natural, in the sense that the theory predictions  $T$  and, by extension, the probabilities depending on them are smooth functions of theoretical parameters. The technical reason is that estimation of correlated uncertainties requires inversion of a large Hessian matrix  $H_{\alpha\beta}$  constructed from derivatives  $\partial T_k/\partial a_\alpha$  of model predictions  $T_k$  with respect to parameters  $a_\alpha$ . This inversion is numerically stable only if  $\partial T_k/\partial a_\alpha$  are well behaved. Naturalness is thus required for reliable estimates of the probabilities and derived quantities, including the model-discriminating ratios  $P(D|T_1)/P(D|T_2)$  and uncertainties on the theoretical predictions.

## B. Gaussian model for the distribution of data

We wish to fit parameters  $a_\alpha$ ,  $\alpha \in \{1, \dots, N_P\}$ , to the data. The data are given by values  $D_k$ ,  $k \in \{1, \dots, N_D\}$ . Each  $D_k$  represents the number of counts divided by an integrated luminosity in a certain bin of measured momenta in a certain experiment that is included in the fit. In the Gaussian approximation, we suppose that after accounting for experimental errors the data have the form

$$D_k = \langle D_k \rangle + \sigma_k \Delta_k + \sigma_k \sum_J \beta_{kJ} \bar{\lambda}_J. \quad (52)$$

Here  $\langle D_k \rangle$  is the value that the datum  $D_k$  would have if there were no experimental uncertainties from either counting statistics or systematic effects such as detector calibration. The value  $\sigma_k$  is the statistical uncertainty (1 standard deviation quoted in the data tables with the  $\pm$  sign) associated with  $D_k$ . The variable  $\Delta_k$  represents fluctuations in  $D_k$  from counting statistics or other sources. In accord with our Gaussian approximation, the  $\Delta_k$  are normalized to be independent Gaussian random variables with mean 0 and variance 1. We use the notation  $\mathcal{N}(0, 1)$  for this distribution. In a typical particle experiment, there can be experimental systematic uncertainties, for example, associated with some imprecision in measurements of luminosity or particle energies. In general, the systematic uncertainties can be decomposed into two types: random fluctuations that are fully uncorrelated point by point and contribute to  $\sigma_k \Delta_k$  in Eq. (52), and fully correlated variations between the experimental data points, contributing to the last term in Eq. (52). Then we would interpret  $\sigma_k$  as the full uncorrelated uncertainty, composed from the uncorrelated statistical and systematic uncertainties added in quadrature.

Let the number of sources of correlated systematic uncertainties be  $N_\lambda$ . We represent the correlated systematic uncertainties with a correlation matrix  $\sigma_k \beta_{kJ}$ . This is a matrix with indices  $k$  and  $J$ . There is no sum over  $k$ . Then in Eq. (52) there is a sum over an index  $J$  that labels the sources of correlated systematic uncertainties. The contribution to  $D_k$  from a systematic source  $J$  is written as  $\sigma_k \beta_{kJ} \bar{\lambda}_J$ , where  $\bar{\lambda}_J$  is an independent random variable with an  $\mathcal{N}(0, 1)$  distribution.<sup>8</sup>

<sup>8</sup>In general, both statistical and systematic uncertainties can be asymmetric and can be quoted as such by the experimental groups. In practice, neglecting the asymmetry has been often acceptable in the PDF analyses.

An example may be helpful. Each  $D_k$  is obtained by dividing a number of counts by a measured value  $\mathcal{L}_0$  of the integrated luminosity  $\mathcal{L}$  for the corresponding experiment. There is some uncertainty in the luminosity measurement, which we express by writing  $\mathcal{L} = \mathcal{L}_0 + \mathcal{L}_0 \sigma_{\mathcal{L}} \bar{\lambda}_{J_{\mathcal{L}}}$ , where  $\sigma_{\mathcal{L}}$  is an estimated fractional uncertainty in the luminosity,  $J_{\mathcal{L}}$  is the index we choose for this source of uncertainty, and  $\bar{\lambda}_{J_{\mathcal{L}}}$  is the corresponding Gaussian random variable. Then in Eq. (52),

$$\sigma_k \beta_{kJ_{\mathcal{L}}} = \langle D_k \rangle \sigma_{\mathcal{L}}. \quad (53)$$

We refer to the random variables  $\bar{\lambda}_J$  as correlated systematic error variables. They are analogous to the variables  $\Delta_k$ , whose fluctuations embody the uncorrelated errors. In one way of analyzing the data, we do not try to determine the values of the  $\bar{\lambda}_J$ . In another analysis, introduced in Sec. III.D, we introduce nuisance parameters  $\lambda_J$  that are intended to estimate the values of the  $\bar{\lambda}_J$ .

The distribution of data  $D$  is given by noting that if  $f$  is a function of the  $\Delta_k$  and  $\bar{\lambda}_J$ , its expectation value is

$$\begin{aligned} \langle f \rangle &= (2\pi)^{-(N_D+N_\lambda)/2} \int d^{N_D} \Delta \int d^{N_\lambda} \bar{\lambda} f(\Delta, \bar{\lambda}) \\ &\times \exp\left(-\frac{1}{2} \sum_k \Delta_k^2 - \frac{1}{2} \sum_J \bar{\lambda}_J^2\right). \end{aligned} \quad (54)$$

Using Eq. (52) in Eq. (54) and observing that  $\langle \Delta_k \rangle = 0$  and  $\langle \bar{\lambda}_J \rangle = 0$ , we see that the expectation value of  $D_k$  is the value  $\langle D_k \rangle$  that appears in Eq. (52). Then, using Eq. (52) in Eq. (54) again and noting that  $\langle \Delta_i \Delta_j \rangle = \delta_{ij}$ ,  $\langle \bar{\lambda}_J \bar{\lambda}_L \rangle = \delta_{JL}$ , and  $\langle \Delta_i \bar{\lambda}_J \rangle = 0$ , we find that

$$\langle (D_i - \langle D_i \rangle)(D_j - \langle D_j \rangle) \rangle = C_{ij}^{-1}, \quad (55)$$

where

$$C_{ij}^{-1} = \sigma_i \sigma_j \left\{ \delta_{ij} + \sum_J \beta_{iJ} \beta_{jJ} \right\}. \quad (56)$$

The matrix  $C^{-1}$  is the inverse of a matrix  $C$  that is called the covariance matrix. When experimental systematic errors  $\beta_{iJ}$  are present,  $C$  is generally not a diagonal matrix.<sup>9</sup> We can use this simple result to show that if  $f$  is a function of the data  $D_k$ , then its expectation value is

$$\begin{aligned} \langle f \rangle &= \frac{\sqrt{\det C}}{(2\pi)^{N_D/2}} \int d^{N_D} D f(D) \\ &\times \exp\left(-\frac{1}{2} \sum_{ij} (D_i - \langle D_i \rangle)(D_j - \langle D_j \rangle) C_{ij}\right). \end{aligned} \quad (57)$$

This gives the probability density for the data  $D_k$ , with the variables  $\bar{\lambda}_J$  integrated out.

<sup>9</sup>We note that if we start with the matrix  $C^{-1}$  instead of Eq. (52), we can diagonalize  $C^{-1}$  to obtain a nonunique representation in the form of Eq. (56), which is consistent with Eq. (52).

The relation between Eqs. (55) and (57) is an important general result that we can use whenever variables  $y_i$  are distributed in a generalized Gaussian fashion, that is, with a probability density proportional to  $\exp(-\sum y_i y_j M_{ij})$  for a matrix  $M$ . To prove it, we use Eq. (57) to compute  $\langle (D_i - \langle D_i \rangle)(D_j - \langle D_j \rangle) \rangle$ . We change variables in Eq. (57) to  $x_i = (\sqrt{C})_{ij}(D_j - \langle D_j \rangle)$ .<sup>10</sup> We use  $d^{N_D} D = (1/\sqrt{\det C}) d^{N_D} x$ . Then the exponent is  $-1/2 \sum_i x_i^2$ . We then obtain  $\langle x_k x_l \rangle = \delta_{kl}$ , which gives us the result in Eq. (55).

### C. Determining theory parameters using $\chi^2$

We now introduce the theoretical predictions  $T_k(a)$  for the data  $D_k$ . The prediction depends on some parameters  $a$ , most notably the parameters of the PDFs. We define  $\langle D_k \rangle$  as the value of  $D_k$  if the experimental errors are negligible. For a given choice of the parameters  $a$ , the theoretical prediction may not be correct, but if the prediction is perfect then  $T_k(a) = \langle D_k \rangle$ .

If we substitute  $\langle D_k \rangle \rightarrow T_k(a)$  into Eq. (57), we see that the probability of obtaining the experimental results  $D$  if the theory represented by  $T(a)$  is correct is

$$P(D|T(a)) = d\mu(D) \exp[-\frac{1}{2} \chi^2(D, a)], \quad (58)$$

where

$$\chi^2(D, a) \equiv \sum_{ij} [D_i - T_i(a)][D_j - T_j(a)] C_{ij} \quad (59)$$

and the data space measure is

$$d\mu(D) \equiv (2\pi)^{-N_D/2} \sqrt{\det C} d^{N_D} D. \quad (60)$$

Consider two choices  $a_1$  and  $a_2$  for the parameters. Suppose that before seeing the experimental results  $D$  we judge the probability that theory  $T(a_1)$  is correct as  $P(T(a_1))$ , and we judge the probability that theory  $T(a_2)$  is correct as  $P(T(a_2))$ . Perhaps these prior probabilities are based on previous experiments, or perhaps they are based on some sort of dynamical model of parton behavior. Whatever our prior belief was, it should be modified after we know the experimental results. Let the new probabilities based on the experimental results  $D$  be  $P(T(a_1)|D)$  and  $P(T(a_2)|D)$ , respectively. Then Bayes's theorem [Eq. (50)] gives us

$$\frac{P(T(a_1)|D)}{P(T(a_2)|D)} = \frac{P(D|T(a_1)) P(T(a_1))}{P(D|T(a_2)) P(T(a_2))}.$$

The information from experiment is contained in the likelihood ratio

$$\frac{P(D|T(a_1))}{P(D|T(a_2))} = \exp\left(-\frac{\chi^2(D, a_1) - \chi^2(D, a_2)}{2}\right). \quad (61)$$

Thus,  $\chi^2(D, a)$  is the function that we need for parameter estimation: differences in  $\chi^2$  give us the likelihood ratio that

<sup>10</sup>We can define the matrix  $\sqrt{C}$  because  $C$  is a real symmetric matrix with all positive eigenvalues.

tells us how to adjust our judgments of which parameter choices are favored using Bayes's theorem [Eq. (50)]. The maximum of the likelihood  $P(D|T(a))$  with respect to  $a$  is achieved at the global minimum of  $\chi^2(D, a)$ .

We can use  $\chi^2(D, a)$  as a measure of goodness of fit: the theory matches the experimental results well when the differences  $D_i - T_i(a)$  are small, meaning that  $\chi^2(D, a)$  is small. However, there are reasons to be critical of  $\chi^2(D, a)$  as the only measure of goodness of fit when the number  $N_D$  of data is large. We return to this question in Sec. IV.

### D. Another definition of $\chi^2$

Given a sample of data  $D_k$ , we can measure how well the theory matches the data by the parameter  $\chi^2(D, a)$  defined in Eq. (59). We can rewrite Eq. (59) as

$$\begin{aligned} \chi^2(D, a) &= \sum_{ij} (D_i - \langle D_i \rangle)(D_j - \langle D_j \rangle) C_{ij} \\ &+ 2 \sum_{ij} (D_i - \langle D_i \rangle)[\langle D_j \rangle - T_j(a)] C_{ij} \\ &+ \sum_{ij} [\langle D_i \rangle - T_i(a)][\langle D_j \rangle - T_j(a)] C_{ij}, \end{aligned} \quad (62)$$

with the expectation value

$$\langle \chi^2(D, a) \rangle = N_D + \chi^2(\langle D \rangle, a), \quad (63)$$

which we derive using Eqs. (59) and (55). This is minimized when  $T_i(a) = \langle D_i \rangle$ .

The function  $\chi^2(D, a)$  is derived from the differential probability  $P(D|T(a))$  to obtain the data  $D$  if the theory  $T(a)$  is correct; cf. Eq. (58). We can define another version of  $\chi^2$ , based on the differential probability  $P(D|T(a), \lambda)$  to obtain the data  $D$  if the theory  $T(a)$  is correct and the random systematic error variables  $\bar{\lambda}$  take the values  $\bar{\lambda}_J = \lambda_J$ . Using Eqs. (54) and (52) with  $\langle D_k \rangle$  replaced by  $T_k(a)$ , this probability is

$$\begin{aligned} P(D|T(a), \lambda) &= (2\pi)^{-(N_D+N_\lambda)/2} \int d^{N_D} \Delta \int d^{N_\lambda} \bar{\lambda} \\ &\times \exp\left(-\frac{1}{2} \sum_k \Delta_k^2 - \frac{1}{2} \sum_J \bar{\lambda}_J^2\right) \\ &\times \prod_k \delta\left(D_k - \left[T_k(a) + \sigma_k \Delta_k + \sigma_k \sum_J \beta_{kJ} \bar{\lambda}_J\right]\right) \\ &\times \prod_J \delta(\bar{\lambda}_J - \lambda_J). \end{aligned} \quad (64)$$

Performing the integrations, we have

$$P(D|T(a), \lambda) = (2\pi)^{-(N_D+N_\lambda)/2} \left[ \prod_k \frac{1}{\sigma_k} \right] \exp\left[-\frac{1}{2} \chi^2(D, a, \lambda)\right], \quad (65)$$

where

$$\chi^2(D, a, \lambda) = \sum_k \left[ \frac{D_k - T_k(a)}{\sigma_k} - \sum_I \beta_{kI} \lambda_I \right]^2 + \sum_J \lambda_J^2. \quad (66)$$

Then, as in Eq. (61), the information from experiment needed to apply Bayes's theorem to the determination of  $a$  and  $\lambda$  is contained in the likelihood ratio

$$\frac{P(D|T(a_1), \lambda_1)}{P(D|T(a_2), \lambda_2)} = \exp \left( -\frac{\chi^2(D, a_1, \lambda_1) - \chi^2(D, a_2, \lambda_2)}{2} \right). \quad (67)$$

Note that the parameters  $\lambda_J$  are not necessarily equal to the true systematic error variables  $\bar{\lambda}_J$ . Rather, the  $\lambda_J$  are parameters that one can fit to the data  $D_k$ . The best-fit values of  $\lambda_J$  then approximate the true  $\bar{\lambda}_J$ . The  $\lambda_J$  are called nuisance parameters. We maximize the likelihood of obtaining the observed data with parton parameters  $a$  and nuisance parameters  $\lambda$  by minimizing  $\chi^2(D, a, \lambda)$  with respect to  $a$  and  $\lambda$ .

The function  $\chi^2(D, a, \lambda)$  is useful if we want to use data to estimate not only the true values of the parton parameters,  $\bar{a}$ , but also the systematic error parameters  $\bar{\lambda}$ . Since we are normally not so interested in the  $\bar{\lambda}$  values, the function  $\chi^2(D, a, \lambda)$  may seem less important than the function  $\chi^2(D, a)$ . However, the function  $\chi^2(D, a, \lambda)$  has the advantage that it does not involve the covariance matrix  $C_{ij}$ .

It is significant that if we fit values of  $\lambda$  by minimizing  $\chi^2(D, a, \lambda)$ , we can obtain  $\chi^2(D, a)$ . With the manipulations outlined in the [Appendix](#), we can write  $\chi^2(D, a, \lambda)$  in the following instructive form:

$$\chi^2(D, a, \lambda) = \sum_{ij} [D_i - T_i(a)][D_j - T_j(a)] C_{ij} + \sum_{IJ} \lambda'_I \lambda'_J B_{IJ}, \quad (68)$$

where  $C$  is the covariance matrix defined in Eq. (56),  $B$  is a matrix with elements

$$B_{IJ} \equiv \delta_{IJ} + \sum_k \beta_{kI} \beta_{kJ}, \quad (69)$$

and  $\lambda'$  is a shifted version of  $\lambda$

$$\lambda'_I = \lambda_I - \sum_J \sum_k \frac{D_k - T_k(a)}{\sigma_k} \beta_{kJ} B_{JI}^{-1}. \quad (70)$$

The minimum of  $\chi^2(D, a, \lambda)$  with respect to  $\lambda$  occurs when  $\lambda' = 0$ , corresponding to  $\lambda = \lambda^{\text{fit}}$ . Thus,

$$\min_{\lambda} \chi^2(D, a, \lambda) \equiv \chi^2(D, a, \lambda^{\text{fit}}) = \chi^2(D, a). \quad (71)$$

The PDF-fitting groups use either form of  $\chi^2$ ; see, for example, the review of various conventions for  $\chi^2$  in [Appendix A of Ball \*et al.\* \(2013b\)](#).

In  $\chi^2(D, a)$ , the experimental systematic errors are encoded in the matrix  $C$ . This form is used, e.g., by the NNPDF analyses. In  $\chi^2(D, a, \lambda)$ , we have the systematic errors expressed explicitly using the parameters  $\lambda$ . This is the

convention adapted for the CTEQ analyses, starting with CTEQ6 ([Pumplin \*et al.\*, 2002](#)). There is then an extra term  $\sum_J \lambda_J^2$  in  $\chi^2(D, a, \lambda)$ . We are instructed to fit the parameters  $\lambda$  to the data by minimizing  $\chi^2(D, a, \lambda)$ . Later, we see how to fit the theory parameters  $a$  by minimizing  $\chi^2(D, a)$  with respect to the parameters  $a$ . This is then equivalent to minimizing  $\chi^2(D, a, \lambda)$  with respect to  $a$  and  $\lambda$ .

If we use  $\chi^2(D, a)$ , then we do not need to be concerned with the systematic error parameters  $\bar{\lambda}$ . With  $\chi^2(D, a)$ , we have a matrix  $C_{ij}$ . The fact that this matrix is not diagonal indicates that the errors are correlated. The presence of  $C_{ij}$  makes the formulas a little complicated, but there are no real conceptual complications:  $C_{ij}$  acts as a metric tensor on the space of the data, so one could think of  $u_i C_{ij} v_j$  as simply  $u \cdot v$ .

The minimum of  $\chi^2(D, a, \lambda)$  occurs at values  $\lambda_J^{\text{fit}}$  of the nuisance parameters. What is the relation between the  $\lambda_J^{\text{fit}}$  and the systematic error variables  $\bar{\lambda}_J$ ? The correlated error variables  $\bar{\lambda}_J$  influence the data  $D_k$ , but the uncorrelated error variables  $\Delta_k$  also influence the  $D_k$  data, and furthermore we do not know the exact parton parameters  $a_\alpha$ , so one cannot expect to recover the  $\bar{\lambda}_J$  exactly from the data. However, we see later that the  $\lambda_J^{\text{fit}}$  approximate the  $\bar{\lambda}_J$  ones when there are many data  $D_k$  and the parameters  $\beta_{kJ}$  that give the influence of the  $\bar{\lambda}_J$  on the data are not too small. Specifically, we will see that the  $\lambda_J^{\text{fit}}$  approximate well the  $\bar{\lambda}_J$  ones when the matrix elements  $B_{IJ}$  are large. This happens when the sum over the data index  $k$  in Eq. (69) includes many terms and the parameters  $\beta_{kJ}$  are not too small.

The analysis is simple. We begin with  $\chi^2(D, a, \lambda)$  in Eq. (66) and substitute in

$$D_k - T_k(a) = \sigma_k \Delta_k + \sigma_k \sum_J \beta_{kJ} \bar{\lambda}_J - [T_k(a) - \langle D_k \rangle] \quad (72)$$

from Eq. (52). This gives

$$\chi^2(D, a, \lambda) = \sum_k \left[ \Delta_k + \frac{T_k(a) - \langle D_k \rangle}{\sigma_k} - \sum_J \beta_{kJ} (\lambda_J - \bar{\lambda}_J) \right]^2 + \sum_J \lambda_J^2. \quad (73)$$

The partial derivatives vanish at the best-fit  $\lambda_I = \lambda_I^{\text{fit}}$ ,

$$0 = -\frac{1}{2} \frac{\partial \chi^2}{\partial \lambda_I} \Big|_{\lambda = \lambda^{\text{fit}}} = \sum_k \left[ \Delta_k + \frac{T_k(a) - \langle D_k \rangle}{\sigma_k} \right] \beta_{kI} - \bar{\lambda}_I - \sum_J B_{IJ} (\lambda_J^{\text{fit}} - \bar{\lambda}_J), \quad (74)$$

so

$$\lambda_I^{\text{fit}} - \bar{\lambda}_I = \sum_J B_{IJ}^{-1} \left\{ \sum_k \beta_{kJ} \left[ \Delta_k + \frac{T_k(a) - \langle D_k \rangle}{\sigma_k} \right] - \bar{\lambda}_J \right\}. \quad (75)$$

On the right-hand side of Eq. (75), the  $\bar{\lambda}_J$  are of order 1, the quantities  $T_k(a) - \langle D_k \rangle$  should be small if we use values of  $a$  fit to the data, and, since the  $\Delta_k$  are independent random variables with  $\mathcal{N}(0, 1)$  distributions,  $\sum_k \beta_{kJ} \Delta_k$  should have

fluctuations of the order of a typical  $\beta_{kJ}$  coefficient times the square root of the number of contributing indices  $k$ . Thus, the quantity in brackets is not large. However, the matrix elements  $B_{IJ}^{-1}$  are small. Thus, we expect the quantity  $\lambda_I^{\text{fit}} - \bar{\lambda}_I$  to be small.

If we use  $\chi^2(D, a, \lambda)$ , then the treatment of the systematic error parameters  $\lambda$  is similar to the treatment of the theory parameters  $a$ . We obtain values  $\lambda^{\text{fit}} \approx \bar{\lambda}$ . The values  $\bar{\lambda}$  are, by assumption, distributed according to the  $\mathcal{N}(0, 1)$  distribution. Thus, the values  $\lambda^{\text{fit}}$  should be approximately distributed with this distribution. In Sec. IV.E, we use this to test the assumptions that we have used.

We can also consider the limit in which  $\sum_k \beta_{kI} \beta_{kJ}$  is small rather than large for some value of  $I$  and all values of  $J$ . Then we have  $B_{IJ} \approx \delta_{IJ}$ . This gives us

$$\lambda_I^{\text{fit}} \approx \sum_k \beta_{kI} \left[ \Delta_k + \frac{T_k(a) - \langle D_k \rangle}{\sigma_k} \right]. \quad (76)$$

Then  $\lambda_I^{\text{fit}}$  is not approximately  $\bar{\lambda}_I$  but is instead quite small. The test in Sec. IV.E suggests that this happens for some of the nuisance parameters.

### E. Dependence on the theory parameters

In Sec. III.C, we introduced theory predictions  $T_k(a)$  for the data  $D_k$ . The  $T_k(a)$  depend on a number  $N_p$  of parameters  $a_\alpha$ . Now we suppose that in the neighborhood of the global  $\chi^2$  minimum the functions  $T_k(a)$  are approximately linear in the parameters. To keep the notation as simple as possible, we define the origin of the parameter space so that the neighborhood of the global minimum is a region near  $a = 0$ . That is, if we are fitting a function  $x^{A_1}(1-x)^{A_2}$  for parameters  $\{A_1, A_2\}$  and preliminary fits give  $A_1 \approx -1.3$  and  $A_2 \approx 4.5$ , we define the new parameters by  $A_1 = -1.3 + a_1$  and  $A_2 = 4.5 + a_2$ . Then we are interested in small values of  $\{a_1, a_2\}$ . We assume for the purpose of examining the fitting procedure that the following linear approximation is adequate:

$$T_k(a) = T_k(0) + T_{k\alpha} a_\alpha. \quad (77)$$

Here and to follow, we use the Einstein summation convention for parameter indices  $\alpha, \beta$ , etc.

Finding the best-fit PDFs does not typically rely on the approximation (77) of linear dependence of  $T_k(a)$  on the PDF parameters  $a$ . However, the PDF error analysis using the Hessian method does rely on this approximation. We adopt it in this review. Using Eq. (77), we find that  $\chi^2$  is a quadratic function of the parameters  $a$ . In fact, within the range of  $a$  important for the fit, this quadratic dependence is a reasonable approximation, as one may note, for example, in Fig. 11 in Sec. IV.I. One should be careful to check whether this approximation is adequate in an actual fit; see Sec. IV.

When the statistical and systematic experimental errors that contribute to the data in Eq. (52) vanish, the data  $D_k$  equal their expectation values  $\langle D_k \rangle$ . We suppose that there are ideal values  $\bar{a}$  of the parton parameters, related to the expectation values  $\langle D_k \rangle$  of the data by

$$T_k(\bar{a}) = \langle D_k \rangle + R_k. \quad (78)$$

We make the definition of the ideal parameters  $\bar{a}$  more precise in Eq. (97). In Eq. (78), we include constants  $R_k$  that represent imperfections in the theory, such that even when we use parameters  $\bar{a}$ , the theory does not match  $\langle D_k \rangle$  exactly. Of course, one commonly assumes that the  $R_k$  are zero, but in this review we want to at least consider the possibility that something goes wrong. For example, the imperfections represented by  $R_k$  could arise because we have omitted higher-order contributions, there is beyond-the-standard-model physics in the data but not in the theory, or the parametrization that we use for the PDFs cannot match the true PDFs exactly.

Another way to include imperfections in the theory is to incorporate theory errors into the analysis. In Eq. (52), we can set the expectation value  $\langle D_k \rangle$  of the data in bin  $k$  to the cross section in that bin calculated exactly in the standard model. We call this exact cross section  $\bar{T}_k(\bar{a})$  so that Eq. (52) becomes

$$D_k = \bar{T}_k(\bar{a}) + \sigma_k \Delta_k + \sigma_k \sum_{J \in \mathcal{E}} \beta_{kJ} \bar{\lambda}_J, \quad (79)$$

where  $\mathcal{E}$  is the set of experimental systematic errors. Now we do not have the exact prediction  $\bar{T}_k(\bar{a})$  available. All we have is the cross section  $T_k(\bar{a})$  calculated, say, at NNLO (but with true PDF and related parameters  $\bar{a}$ ). In the linear approximation that we use, we can parametrize our ignorance in the form

$$\bar{T}_k(\bar{a}) = T_k(\bar{a}) + \sigma_k \sum_{J \in \mathcal{T}} \beta_{kJ} \bar{\lambda}_J, \quad (80)$$

where  $\mathcal{T}$  is a set of sources of theory errors and the parameters  $\bar{\lambda}_J$  are random variables that are chosen from a distribution such as  $\mathcal{N}(0, 1)$ . Then the term  $\sigma_k \sum_{J \in \mathcal{T}} \beta_{kJ} \bar{\lambda}_J$  is our estimate of theory contributions to the true cross section  $\bar{T}_k(\bar{a})$  that are not included in our NNLO calculation  $T_k(\bar{a})$ . The parameters  $\beta_{kJ}$  represent our estimate of the dependence of  $\bar{T}_k(\bar{a})$  on the  $J$ th source of theoretical uncertainty. Of course, in reality only one value  $\lambda_J^{\text{true}}$  of each  $\bar{\lambda}_J$  is realized in an exact calculation. The  $R_k$  in Eq. (78) are then

$$R_k = -\sigma_k \sum_{J \in \mathcal{T}} \beta_{kJ} \lambda_J^{\text{true}}. \quad (81)$$

When we combine Eqs. (79) and (80), we obtain

$$D_k = T_k(\bar{a}) + \sigma_k \Delta_k + \sigma_k \sum_{J \in \mathcal{E} \cup \mathcal{T}} \beta_{kJ} \bar{\lambda}_J, \quad (82)$$

where both experimental systematic errors and our estimated theory errors are now included. The representations of theory errors were discussed by Olness and Soper (2010), Cacciari and Houdeau (2011), Gao (2011), Forte, Isgrò, and Vita (2014), Abdul Khalek *et al.* (2019a), and Harland-Lang and Thorne (2019). Until recently, PDF fits typically omitted theory errors. In this review, we do not include theory errors and instead represent imperfections in the theory by the  $R_k$  in Eq. (78).

There is a certain freedom in the definition of  $\bar{a}$  and  $R_k$  in Eq. (78). We can use this freedom to simplify the later

analysis. Suppose that we say that Eq. (78) applies for ideal parameters  $\bar{a}^{(0)}$  and imperfection parameters  $R_k^{(0)}$ :

$$T_k(\bar{a}^{(0)}) = \langle D_k \rangle + R_k^{(0)}. \quad (83)$$

Let  $\bar{a}^{(0)} = \bar{a} + \delta a$ , where the  $\delta a_\alpha$  are small parameters that we are free to choose. Then

$$T_k(\bar{a}) + T_{k\alpha}\delta a_\alpha = \langle D_k \rangle + R_k^{(0)}. \quad (84)$$

This gives us Eq. (78) with  $R_k = R_k^{(0)} - T_{k\alpha}\delta a_\alpha$ . What should the  $\delta a_\alpha$  be? In the following analysis, the vector  $\sum_{kj} R_k C_{kj} T_{j\beta}$  plays an important role. This vector equals

$$\sum_{kj} R_k C_{kj} T_{j\beta} = \sum_{kj} R_k^{(0)} C_{kj} T_{j\beta} - H_{\alpha\beta} \delta a_\alpha, \quad (85)$$

where

$$H_{\alpha\beta} \equiv \sum_{kj} T_{k\alpha} C_{kj} T_{j\beta} \quad (86)$$

is the Hessian matrix, which will play a major role in the subsequent analysis. We choose

$$\delta a_\alpha = H_{\alpha\beta}^{-1} \sum_{kj} R_k^{(0)} C_{kj} T_{j\beta} \quad (87)$$

so that

$$\sum_{kj} R_k C_{kj} T_{j\beta} = 0. \quad (88)$$

This is a useful property, as we see later.

We define  $\chi^2(D, a)$  by Eq. (59) so that

$$\begin{aligned} \chi^2(D, a) &= \sum_{ij} [D_i - T_i(0) - T_{i\alpha} a_\alpha] \\ &\quad \times [D_j - T_j(0) - T_{j\beta} a_\beta] C_{ij}. \end{aligned} \quad (89)$$

The minimum of  $\chi^2(D, a)$  is at parameters such that

$$0 = -\frac{1}{2} \frac{\partial \chi^2}{\partial a_\beta} = \sum_{ij} [D_i - T_i(0) - T_{i\alpha} a_\alpha] C_{ij} T_{j\beta}. \quad (90)$$

This is

$$H_{\beta\alpha} a_\alpha = \mathcal{D}_\beta, \quad (91)$$

where

$$\mathcal{D}_\beta \equiv \sum_{ij} [D_i - T_i(0)] C_{ij} T_{j\beta}. \quad (92)$$

Thus, the fit parameters are

$$a_\alpha^{\text{fit}} = H_{\alpha\beta}^{-1} \mathcal{D}_\beta. \quad (93)$$

What is the expectation value of  $a_\alpha^{\text{fit}}$ ? To answer this question, we need a result for  $\bar{a}_\alpha$ . Using Eq. (78), we have

$$T_{i\alpha} \bar{a}_\alpha = \langle D_i \rangle - T_i(0) + R_i. \quad (94)$$

Thus,

$$\sum_{ij} C_{ji} T_{j\beta} T_{i\alpha} \bar{a}_\alpha = \sum_{ij} C_{ji} T_{j\beta} (\langle D_i \rangle - T_i(0) + R_i). \quad (95)$$

Using the definition (86) of the Hessian matrix and the property (88) of  $R_k$ , this is

$$H_{\beta\alpha} \bar{a}_\alpha = \sum_{ij} [\langle D_i \rangle - T_i(0)] C_{ij} T_{j\beta}. \quad (96)$$

Thus,

$$\bar{a}_\alpha = H_{\alpha\beta}^{-1} \sum_{ij} [\langle D_i \rangle - T_i(0)] C_{ij} T_{j\beta}. \quad (97)$$

Comparing this to Eq. (93) gives us

$$a_\alpha^{\text{fit}} - \bar{a}_\alpha = H_{\alpha\beta}^{-1} \sum_{ij} (D_i - \langle D_i \rangle) C_{ij} T_{j\beta}. \quad (98)$$

The result is that the expectation value of  $a_{\text{fit}}$  is  $\bar{a}$ :

$$\langle a_\alpha^{\text{fit}} - \bar{a}_\alpha \rangle = 0. \quad (99)$$

This also gives the correlations of the parameters  $a_\alpha$  with the data  $D_i$  and with each other. From Eqs. (98) and (55), we have

$$\langle (D_k - \langle D_k \rangle) (a_\alpha^{\text{fit}} - \bar{a}_\alpha) \rangle = \sum_{ij} H_{\alpha\beta}^{-1} C_{ki}^{-1} C_{ij} T_{j\beta}. \quad (100)$$

Thus,

$$\langle (D_k - \langle D_k \rangle) (a_\alpha^{\text{fit}} - \bar{a}_\alpha) \rangle = H_{\alpha\beta}^{-1} T_{k\beta}. \quad (101)$$

For  $\langle (a_\alpha^{\text{fit}} - \bar{a}_\alpha) (a_\beta^{\text{fit}} - \bar{a}_\beta) \rangle$ , Eq. (98) gives

$$\begin{aligned} &\langle (a_\alpha^{\text{fit}} - \bar{a}_\alpha) (a_\beta^{\text{fit}} - \bar{a}_\beta) \rangle \\ &= H_{\alpha\sigma}^{-1} H_{\beta\tau}^{-1} \sum_{ijkl} C_{ik} T_{k\sigma} C_{jl} T_{l\tau} \langle (D_i - \langle D_i \rangle) (D_j - \langle D_j \rangle) \rangle \\ &= H_{\alpha\sigma}^{-1} H_{\beta\tau}^{-1} \sum_{ijkl} C_{ik} T_{k\sigma} C_{jl} T_{l\tau} C_{ij}^{-1} \\ &= H_{\alpha\sigma}^{-1} H_{\beta\tau}^{-1} \sum_{kl} C_{lk} T_{k\sigma} T_{l\tau} \\ &= H_{\alpha\sigma}^{-1} H_{\beta\tau}^{-1} H_{\sigma\tau}. \end{aligned} \quad (102)$$

Thus,

$$\langle (a_\alpha^{\text{fit}} - \bar{a}_\alpha) (a_\beta^{\text{fit}} - \bar{a}_\beta) \rangle = H_{\alpha\beta}^{-1}. \quad (103)$$

Compare this to Eq. (55),  $\langle (D_i - \langle D_i \rangle) (D_j - \langle D_j \rangle) \rangle = C_{ij}^{-1}$ . Thus, we see the importance of the Hessian matrix: its inverse is the correlation matrix for the fitted parton parameters.

## F. Distribution of the parameters

Since the data are Gaussian distributed and the fit parameters are linearly related to the data, the fit parameters are Gaussian distributed. The expectation values (103) give us the distribution of the best-fit parameters, analogously to what we find in Eq. (57): given a function  $f(a_{\text{fit}} - \bar{a})$ , we have

$$\langle f \rangle = \sqrt{\det H} (2\pi)^{-N_p/2} \int d^{N_p} (a_{\text{fit}} - \bar{a}) f(a_{\text{fit}} - \bar{a}) \times \exp \left[ -\frac{1}{2} H_{\alpha\beta} (a_{\alpha}^{\text{fit}} - \bar{a}_{\alpha}) (a_{\beta}^{\text{fit}} - \bar{a}_{\beta}) \right]. \quad (104)$$

Let us be clear about where this comes from. We consider an ensemble of repetitions of the experiments. As the data fluctuate, the values of  $a_{\text{fit}}$  fluctuate according to the distribution (104). However, we can turn this around. Given the data  $D_k$ , we find the corresponding best-fit parameters  $a_{\text{fit}}$ . We do not know  $\bar{a}$ . But if we repeat the experiments many times, the values of the difference  $(a_{\text{fit}} - \bar{a})$  fluctuate around zero according to Eq. (104). Then Eq. (104) gives us a measure of the error in estimating  $\bar{a}$  by  $a_{\text{fit}}$ .

Equation (104) applies in the  $N_p$ -dimensional space of fit parameters. It is instructive to consider how this works in a particular coordinate system. We let  $\{e^{(1)}, e^{(2)}, \dots, e^{(N_p)}\}$  be a set of basis vectors for the parameter space. We take these basis vectors to be orthogonal and normalized using  $H$  as the metric tensor<sup>11</sup>

$$e_{\alpha}^{(n)} H_{\alpha\beta} e_{\beta}^{(m)} = \delta_{mn}. \quad (105)$$

The corresponding completeness relation is

$$\sum_n e_{\alpha}^{(n)} e_{\beta}^{(n)} = H_{\alpha\beta}^{-1}. \quad (106)$$

[To prove this, we define  $\sum_n e_{\alpha}^{(n)} e_{\beta}^{(n)} = A_{\alpha\beta}$  and use Eq. (105) to show that  $A_{\alpha\beta} H_{\beta\gamma} e_{\gamma}^{(n)} = e_{\alpha}^{(n)}$ .]

It is often useful to choose the basis vectors to be the eigenvectors of  $H$ :  $H_{\alpha\beta} e_{\beta}^{(n)} = h_n e_{\alpha}^{(n)}$ . However, any choice of basis vectors obeying Eq. (105) will do. For instance,  $e^{(1)}$  could be a vector normalized to  $e_{\alpha}^{(1)} H_{\alpha\beta} e_{\beta}^{(1)} = 1$ , pointing in a direction that is of particular interest. Then the other  $e^{(n)}$  could be chosen to satisfy Eq. (105). We use this construction in Sec. III.G.

Using the basis vectors  $e^{(n)}$ , we can expand a general vector of parameters  $a$  about  $a_{\text{fit}}$  in the form

$$a_{\alpha}(t) = a_{\alpha}^{\text{fit}} + \sum_n t_n e_{\alpha}^{(n)}. \quad (107)$$

Here the argument  $t$  in  $a(t)$  denotes  $\{t_1, \dots, t_{N_p}\}$ . How does  $\chi^2$  depend on the parameters  $t_n$ ? To find out, we evaluate

<sup>11</sup>We do not distinguish between upper and lower indices  $\alpha, \beta, \dots$ . If we did, parton parameters would have upper indices  $a^{\alpha}$  and the metric tensor  $H$  would have lower indices  $H_{\alpha\beta}$ . Then  $e^{(n)}$  and  $H^{-1}$  would have upper indices  $e_{\alpha}^{(n)}$  and  $(H^{-1})^{\alpha\beta}$ .

$\chi^2(D, a(t))$ . Using Eqs. (86), (88), (89), (94), and (98), we obtain for a general choice of  $a$

$$\chi^2(D, a) = \sum_{ij} [D_i - \langle D_i \rangle - R_i] [D_j - \langle D_j \rangle - R_j] C_{ij} - 2H_{\alpha\beta} (a_{\alpha}^{\text{fit}} - \bar{a}_{\alpha}) (a_{\beta} - \bar{a}_{\beta}) + H_{\alpha\beta} (a_{\alpha} - \bar{a}_{\alpha}) (a_{\beta} - \bar{a}_{\beta}). \quad (108)$$

Here  $a_{\text{fit}}$  is the parameter choice that we get from fitting the data  $D_k$ ,  $\bar{a}$  is what we get by averaging  $a_{\text{fit}}$  over an imagined ensemble of experiments, and  $a$  represents the parameters that we are free to vary. Then if we substitute  $a(t)$  given in Eq. (107) for  $a$ , we get

$$\chi^2(D, a(t)) = \sum_{ij} [D_i - \langle D_i \rangle - R_i] [D_j - \langle D_j \rangle - R_j] C_{ij} - H_{\alpha\beta} (a_{\alpha}^{\text{fit}} - \bar{a}_{\alpha}) (a_{\beta}^{\text{fit}} - \bar{a}_{\beta}) + \sum_n t_n^2. \quad (109)$$

That is, varying  $a$  from  $a_{\text{fit}}$  in any direction  $e_{\alpha}^{(n)}$  by  $t_n = 1$  increases  $\chi^2$  by 1.

The distribution (104) of differences of  $a_{\text{fit}}$  from  $\bar{a}$  can be rewritten in the  $e^{(n)}$  basis. We define

$$a_{\alpha}^{\text{fit}} - \bar{a}_{\alpha} = -\sum_n t_n e_{\alpha}^{(n)}. \quad (110)$$

Then in Eq. (104), we can regard  $f$  as a function of the eigenvector coordinates  $t_n$  instead of the original parameters  $a_{\alpha}^{\text{fit}} - \bar{a}_{\alpha}$ , and we can change integration variables to the  $t_n$ , giving us

$$\langle f \rangle = \int_{-\infty}^{\infty} \frac{dt_1}{\sqrt{2\pi}} \dots \int_{-\infty}^{\infty} \frac{dt_{N_p}}{\sqrt{2\pi}} f(t_1, \dots, t_{N_p}) \exp \left( -\frac{1}{2} \sum_{n=1}^{N_p} t_n^2 \right). \quad (111)$$

Each coordinate  $t_n$  follows an  $\mathcal{N}(0, 1)$  distribution.

In particular, if we are interested only in the component of  $\bar{a} - a_{\text{fit}}$  in the direction  $e^{(1)}$ , we can let the function  $f$  in Eq. (111) depend only on  $t_1$ . Then

$$\langle f \rangle = \int_{-\infty}^{\infty} \frac{dt_1}{\sqrt{2\pi}} f(t_1) \exp \left( -\frac{1}{2} t_1^2 \right). \quad (112)$$

Comparing this to Eq. (109), we see that a “ $2\sigma$ ” value  $t_1$ , that is,  $t_1 = 2$ , increases  $\chi^2$  by 4 from its best-fit value.

It is of interest to understand the distribution  $\rho(R^2, N_p)$  of  $R^2 = \sum_n t_n^2$ ,

$$\rho(R^2, N_p) = \int_{-\infty}^{\infty} \frac{dt_1}{\sqrt{2\pi}} \dots \int_{-\infty}^{\infty} \frac{dt_{N_p}}{\sqrt{2\pi}} \delta \left( \sum_{n=1}^{N_p} t_n^2 - R^2 \right) \times \exp \left( -\frac{1}{2} \sum_{n=1}^{N_p} t_n^2 \right). \quad (113)$$

This is the  $\chi^2$  distribution with  $N_p$  degrees of freedom. In fitting parton distributions,  $N_p$  is quite large, say,  $N_p = 25$ .

The mean value of any of the  $t_n^2$  is  $t_n^2 = 1$ , but the mean value of  $R^2$  is much larger, as we see in Sec. III.H:  $\langle R^2 \rangle = N_P$ . Using the  $t_n$  as coordinates, the hypersurface  $\sum_{n=1}^{N_P} t_n^2 = R^2$  is a sphere. In terms of the original parton parameters  $\bar{a}_\alpha - a_\alpha^{\text{fit}}$ , it is an ellipsoid. With the values of  $\bar{a}_\alpha - a_\alpha^{\text{fit}}$  distributed according to Eq. (104), for the ellipsoid to include 95% of the points we should choose  $R \approx 6.1$  for  $N_P = 25$ . In contrast, if we look at just one  $t_n$ , then for the interval  $t_n^2 < R_n^2$  to include 95% of the points  $t_n$ , we should choose  $R_n \approx 2$ . But with  $N_P = 25$ , the fraction  $P$  of points inside an ellipsoid with  $R = 2$  is  $P \approx 5 \times 10^{-7}$ . This discussion illustrates that when we discuss the uncertainties in the determination of the parton parameters we need to carefully distinguish whether we are discussing an uncertainty interval in one dimension or in 25 dimensions. As another consequence of the large- $N_P$  geometry, when the PDF probability distribution is sampled by randomly varying the PDF parameters, the overwhelming majority of such Monte Carlo parameter replicas are likely to be bad fits with  $P \sim 0$  (Hou *et al.*, 2017). Thus, though the estimates of the first and second moments of the  $N_P$ -dimensional probability distribution from the Monte Carlo sample converge to their true values with about 100–1000 replicas, the Monte Carlo method tends to be highly inefficient for exploring the neighborhood  $\sum_n t_n^2 \leq R^2$ , with  $R$  of order unity. In contrast, the analytic minimization of  $\chi^2$  by the gradient descent, as implemented in CTEQ and MMHT fits, directly finds the neighborhood  $\sum_n t_n^2 \leq R^2$  around the global minimum and renders the probability distribution within this neighborhood. The analytic minimization and Monte Carlo sampling approaches thus offer complementary strengths when examining the multidimensional probability distribution of PDF parameters and associated PDF uncertainties.

### G. Calculation of a cross section

We can now ask another question. Suppose that  $\sigma$  is a cross section that is determined by the PDFs. Then  $\sigma$  is a function  $\sigma(a)$  of the parton parameters  $a$ . We need data with no errors to determine the ideal parameters  $\bar{a}$ , so we never know  $\sigma(\bar{a})$  exactly. However, we can fit the parameters and estimate  $\sigma(\bar{a})$  by  $\sigma(a_{\text{fit}})$ .

What is the expected error resulting from using the fit parameters? To analyze this, we begin by defining the shift in the cross section

$$\Delta\sigma = \sigma(\bar{a}) - \sigma(a_{\text{fit}}). \quad (114)$$

Then with a linear approximation we write

$$\Delta\sigma = (\bar{a}_\alpha - a_\alpha^{\text{fit}})\sigma'_\alpha, \quad (115)$$

with  $\sigma'_\alpha \equiv \partial\sigma(a)/\partial a_\alpha$ .

Now we define a special vector in the space of parameters according to

$$e(\sigma)_\alpha = \frac{H_{\alpha\beta}^{-1}\sigma'_\beta}{\sqrt{\sigma'_\gamma H_{\gamma\delta}^{-1}\sigma'_\delta}}. \quad (116)$$

This vector is normalized to

$$e(\sigma)_\alpha H_{\alpha\beta} e(\sigma)_\beta = 1. \quad (117)$$

It is useful to define more vectors,  $\{e(\sigma)^{(2)}, \dots, e(\sigma)^{(N_P)}\}$  such that these vectors, together with  $e(\sigma) \equiv e(\sigma)^{(1)}$ , form a basis for the parameter space and such that the basis vectors are orthogonal and normalized using the metric tensor  $H_{\alpha\beta}$  as in Eq. (105).

With the aid of these basis vectors, we can write a general parameter vector as

$$a_\alpha = a_\alpha^{\text{fit}} + t_1 e(\sigma)_\alpha + \sum_{n \geq 2} t_n e(\sigma)_\alpha^{(n)}. \quad (118)$$

The corresponding change in  $\chi^2$ , using Eq. (109), is

$$\begin{aligned} \chi^2 \left( D, a_{\text{fit}} + t_1 e(\sigma) + \sum_{n \geq 2} t_n e(\sigma)^{(n)} \right) \\ = \chi^2(D, a_{\text{fit}}) + t_1^2 + \sum_{n \geq 2} t_n^2. \end{aligned} \quad (119)$$

We set  $a \rightarrow \bar{a}$  in the definition Eq. (118) so that

$$\bar{a}_\alpha - a_\alpha^{\text{fit}} = t_1 e(\sigma)_\alpha + \sum_{n \geq 2} t_n e(\sigma)_\alpha^{(n)}. \quad (120)$$

Then, according to Eq. (111), as  $a_{\text{fit}}$  varies in an ensemble experiment set, the expansion parameters  $\{t_1, \dots, t_{N_P}\}$  fluctuate following  $\mathcal{N}(0, 1)$  distributions.

We can use this result to analyze the fluctuations in the following cross section from Eq. (115):

$$\Delta\sigma = t_1 e(\sigma)_\alpha \sigma'_\alpha + \sum_{n \geq 2} t_n e(\sigma)_\alpha^{(n)} \sigma'_\alpha. \quad (121)$$

Using Eq. (116), this becomes

$$\Delta\sigma = \sqrt{\sigma'_\gamma H_{\gamma\delta}^{-1} \sigma'_\delta} \left\{ t_1 e(\sigma)_\alpha H_{\alpha\beta} e(\sigma)_\beta + \sum_{n \geq 2} t_n e(\sigma)_\alpha^{(n)} H_{\alpha\beta} e(\sigma)_\beta^{(n)} \right\}. \quad (122)$$

Because of the orthonormality condition (105), only the first term survives and we obtain

$$\Delta\sigma = \sqrt{\sigma'_\gamma H_{\gamma\delta}^{-1} \sigma'_\delta} t_1. \quad (123)$$

Thus, the fluctuations in the cross section are given entirely by the fluctuations of the parameters along the special direction  $e(\sigma)^{(1)}$ . There is a coefficient  $[\sigma'_\gamma H_{\gamma\delta}^{-1} \sigma'_\delta]^{1/2}$  that is larger when the cross section is a fast varying function of the parton parameters. The remaining factor  $t_1$  fluctuates following an  $\mathcal{N}(0, 1)$  distribution as the data fluctuate. This means that if we want  $\Delta\sigma$  to represent, say, a 2 standard deviation error on  $\sigma$ , we set  $t_1 = 2$ . Furthermore,  $t_1$  has the property that when the parameters vary from the best-fit parameters according to  $a = a_{\text{fit}} + t_1 e(\sigma)^{(1)}$ , the  $\chi^2$  increases by  $t_1^2$ .



### 1. Formulas to compute PDF uncertainties

There is a standard practical method for calculating  $\Delta\sigma$ . We choose basis vectors that are not specially adapted to the cross section  $\sigma(a)$ . We choose the basis vectors  $e^{(n)}$ ,  $n = 1, \dots, N_p$  to obey the orthonormality condition (105). Typically, the basis vectors  $e^{(n)}$  are chosen as eigenvectors of the Hessian matrix  $H$ . Commonly, there are  $2N_p$  error fits,  $a_{\text{fit}} \pm \tilde{t}e^{(n)}$ , that come with a set of published PDFs. Here  $\tilde{t}$  is defined by the published analysis. If we assume that linear approximations are adequate, then we need only  $N_p$  error fits,  $a_{\text{fit}} + \tilde{t}e^{(n)}$ , with positive  $\tilde{t}$ . Error fits with two signs (Nadolsky and Sullivan, 2001) allow for a more complete treatment than we give here, which allows for nonlinear contributions.

One can use the basis vectors  $e^{(n)}$  to evaluate the uncertainty in  $\sigma(a_{\text{fit}})$ . For each direction  $n$ , define

$$\Delta\sigma^{(n)} = \sigma(a_{\text{fit}} + te^{(n)}) - \sigma(a_{\text{fit}}). \quad (124)$$

Recall from Eq. (109) that if we set  $a = a_{\text{fit}} + te^{(n)}$ , then  $\chi^2(D, a)$  increases by  $t^2$  compared to  $\chi^2(D, a_{\text{fit}})$ . As long as we use a linear approximation, we have

$$\Delta\sigma^{(n)} = t\sigma'_\alpha e_\alpha^{(n)}. \quad (125)$$

Now sum the squares of  $\Delta\sigma^{(n)}$ :

$$\sum_n (\Delta\sigma^{(n)})^2 = t^2 \sum_n \sigma'_\alpha e_\alpha^{(n)} e_\beta^{(n)} \sigma'_\beta. \quad (126)$$

Using the completeness relation (106) for the basis vectors  $e^{(n)}$ , this is

$$\sum_n (\Delta\sigma^{(n)})^2 = t^2 \sigma'_\alpha H_{\alpha\beta}^{-1} \sigma'_\beta. \quad (127)$$

According to Eq. (123), we can estimate the error in  $\sigma$  by

$$(\Delta\sigma)^2 = t^2 \sigma'_\alpha H_{\alpha\beta}^{-1} \sigma'_\beta, \quad (128)$$

where, for example, we choose  $t = 2$  if we want a  $2\sigma$  error estimate. Thus, we can obtain  $\Delta\sigma$  by using variations in the eigenvector directions  $e^{(n)}$ :

$$(\Delta\sigma)^2 = \sum_n (\Delta\sigma^{(n)})^2. \quad (129)$$

That is, we need to calculate  $N_p$  error contributions  $\Delta\sigma^{(n)}$  by using the parton error sets according to Eq. (124). Then adding the errors  $\Delta\sigma^{(n)}$  in quadrature gives the total error  $\Delta\sigma$ .

There is a second standard practical method for calculating  $\Delta\sigma$ . This method derives from the publications of the NNPDF Collaboration (Ball *et al.*, 2010, 2013a, 2015, 2017). With the NNPDF approach, there is effectively a large number of parameters, and the distribution of the parameters is not strictly Gaussian. The distribution of results is represented by giving a large sample of parton distribution sets. Within the linear approximations that we use in this review, one would generate a large sample of parton distributions based on

parameters  $a_\alpha = a_\alpha^{\text{fit}} + \sum_n t_n e_\alpha^{(n)}$  as in Eq. (107), with Gaussian random variables  $t_n$ . Given this sample, one calculates  $\sigma(a)$  for each example and thus obtains a corresponding sample of  $\sigma(a)$  values, from which one obtains statistical properties of the sample such as  $\langle\sigma\rangle$  and  $(\Delta\sigma)^2$ .

We have spoken of  $\sigma(a)$  as a cross section. More broadly,  $\sigma(a)$  in this section could be any physical quantity that depends on the parton parameters  $a$ . In particular,  $\sigma(a)$  could be a parton distribution function  $f_{a/A}(x, \mu^2)$  for a parton flavor  $a$ , evaluated at a momentum fraction  $x$  and a scale  $\mu$ . Then the previously presented calculation gives us an error estimate  $\Delta f_{a/A}(x, \mu^2)$ .

### H. Expectation value and variance of $\chi^2$

In this section, we investigate the value of  $\chi^2$  obtained in the fit. Start with Eq. (108) for  $\chi^2(D, a)$ . We fit the parameters  $a$  to minimize  $\chi^2(D, a)$  for given data  $D$ . Using the fit parameters gives

$$\chi^2(D, a^{\text{fit}}) = \sum_{ij} [D_i - \langle D_i \rangle - R_i][D_j - \langle D_j \rangle - R_j] C_{ij} - H_{\alpha\beta} (a_\alpha^{\text{fit}} - \bar{a}_\alpha)(a_\beta^{\text{fit}} - \bar{a}_\beta). \quad (130)$$

Using Eq. (98) for  $a_\alpha^{\text{fit}} - \bar{a}_\alpha$ , this is

$$\chi^2 = \sum_{ij} [D_i - \langle D_i \rangle - R_i][D_j - \langle D_j \rangle - R_j] C_{ij} - \sum_{ij} [D_i - \langle D_i \rangle][D_j - \langle D_j \rangle] M_{ij}, \quad (131)$$

where the matrix  $M$  is

$$M = CTH^{-1}T^TC. \quad (132)$$

Note that here we eliminate the parameters  $a_\alpha$  entirely.

We can use this to evaluate the expectation value  $\langle\chi^2\rangle$  of  $\chi^2$  and its variance

$$\langle(\chi^2 - \langle\chi^2\rangle)^2\rangle = \langle(\chi^2)^2\rangle - \langle\chi^2\rangle^2. \quad (133)$$

We use Eq. (57) for the probability distribution of the data  $D$ . This gives

$$\langle[D_i - \langle D_i \rangle][D_j - \langle D_j \rangle]\rangle = C_{ij}^{-1} \quad (134)$$

and

$$\begin{aligned} \langle[D_i - \langle D_i \rangle][D_j - \langle D_j \rangle][D_k - \langle D_k \rangle][D_l - \langle D_l \rangle]\rangle \\ = C_{ij}^{-1} C_{kl}^{-1} + C_{ik}^{-1} C_{jl}^{-1} + C_{il}^{-1} C_{jk}^{-1}. \end{aligned} \quad (135)$$

To derive this, one can change the variables in Eq. (57) to  $x_i = \sum_j (\sqrt{C})_{ij} (D_j - \langle D_j \rangle)$ . Then the symmetries of the integrand imply that  $\langle x_i x_j \rangle \propto \delta_{ij}$  and  $\langle x_i x_j x_k x_l \rangle \propto \delta_{ij} \delta_{kl} + \delta_{ik} \delta_{jl} + \delta_{il} \delta_{jk}$ . The coefficients of proportionality are simple to evaluate, giving Eqs. (134) and (135).

Now a certain amount of algebra with the matrices leads to results containing the factors

$$\sum_{i=1}^{N_D} \delta_{ii} = N_D, \quad \sum_{\alpha=1}^{N_P} \delta_{\alpha\alpha} = N_P. \quad (136)$$

The results are

$$\langle \chi^2 \rangle = N_D - N_P + \sum_{ij} R_i C_{ij} R_j \quad (137)$$

and

$$\langle (\chi^2 - \langle \chi^2 \rangle)^2 \rangle = 2(N_D - N_P) + 4 \sum_{ij} R_i C_{ij} R_j. \quad (138)$$

These results are often used to provide an indication of whether a good fit has been found. The parameters  $R_i$  that we have introduced represent an imperfection in the theory: if the parton distributions do not have enough available parameters, we expect  $R_i \neq 0$ . If there are enough parameters and if the rest of the theoretical model is correct, then the  $R_i$  parameters should vanish. In that case,  $\langle \chi^2 \rangle$  should be close to  $N_D - N_P$ . For example, if  $N_D = 3000$  and  $N_P = 25$ , then  $\chi^2$  should be around 2975. The square root of the variance of  $\chi^2$  in this case is  $\sqrt{5950} \approx 77$ . Thus, we expect to find  $\chi^2 \approx 2975 \pm 77$ . The fit is bad if  $\chi^2$  is too high or too low.

#### IV. TESTS OF PERFORMANCE OF THE FIT

In Sec. III.C, we examined the function  $\chi^2(D, a)$  defined in Eq. (59) that is minimized to determine PDF parameters  $a$  from the experimental data  $D$ , giving values  $a_{\text{fit}}$ . However, this fitting procedure produces correct results only if the data  $D$  are reliable within their errors as given by the experiments and if the adopted theory is actually a good description of nature for some parameter combination  $a_{\text{fit}}$ . If  $\chi^2(D, a_{\text{fit}})$  does not lie within certain limits, one can conclude that something is wrong with the fit. However, it has been known since the inception of the global QCD analysis in the late 1980s that the value of the global  $\chi^2(D, a_{\text{fit}})$  is an essential, but far from sufficient, measure of the goodness of fit; see, e.g., [Morfin and Tung \(1991\)](#).

In this section, we argue that the PDF fit should pass a number of tests to fulfill what one might call a strong set of goodness-of-fit criteria. Several of these tests involve looking at quantities derived from the fitting procedure that should follow a predicted distribution if the statistical assumptions on which the fit is based are valid. One can then test whether the quantities are in fact distributed as predicted. We include the distribution of the nuisance parameters, the distribution of the residuals for the fitted data, and the distribution of  $\chi^2$  values for many subsets of the data. Another test looks at whether individual subsets of the data are statistically consistent with the global  $a_{\text{fit}}$  in individual directions in the space of parameters. In Sec. IV.D, we discuss the closure test, which is a powerful test of the fitting methodology and is used, in particular, by NNPDF ([Ball et al., 2015](#)).

These tests, taken together, are more constraining and difficult to satisfy than the standard weak goodness-of-fit

criterion based on the value of the global  $\chi^2(D, a)$ . The point of the strong set of goodness-of-fit criteria is to find places where the statistical assumptions used in the fitting procedure break down. We will find, in examples, that some of the assumptions used in the fit do break down. When this happens, the PDF error estimates that emerge from the fit cannot be realistic. One possibility is that any observed problems are the result of understated estimates of the experimental systematic errors. We propose ways to deal with this in Secs. IV.I and IV.J. The method usually used to adjust the final PDF error estimates is to apply a tolerance criterion. This is a complicated subject, which we address in Sec. IV.L.

One of the tests, which we investigate in Sec. IV.H, looks at  $\chi^2$  for subsets of the data. Data subsets are often examined visually to rule out systematic discrepancies by comparing data and theory predictions in the figures. This is a reasonable, but slow and imprecise, test. It can be made more precise using the procedure in Secs. IV.G and IV.H.

A standard goodness-of-fit criterion based on the value of the overall  $\chi^2(D, a_{\text{fit}})$  is called the hypothesis-testing criterion ([Collins and Pumplin, 2001](#)). We see in Eq. (137) that if the theory is perfect so that  $R_j = 0$ , the expectation value of this quantity is  $\langle \chi^2(D, a_{\text{fit}}) \rangle = N_D - N_P$ . However, Eq. (138) shows that  $\chi^2(D, a_{\text{fit}})$  is expected to fluctuate by about  $\sqrt{2(N_D - N_P)}$ . Thus, we surely have a bad fit at the  $2\sigma$  level if  $\chi^2(D, a_{\text{fit}}) - (N_D - N_P)$  is bigger than  $2 \times \sqrt{2(N_D - N_P)}$ , that is, about 150 for  $N_D - N_P \approx 3000$ .

However, if we fit the parameters  $a$  without changing the functional form, a small difference in  $\chi^2$  values of the order of  $2^2 = 4$  is already significant, while  $2 \times \sqrt{2(N_D - N_P)} \approx 150$  is far too large. In this restricted situation, the parameter-fitting criterion, which assigns the 95% (or 68%) probability level to the increase  $\Delta\chi^2 = 4$  (or 1), adequately estimates the uncertainty on parameters of this fixed model, as long as the statistical assumptions on which the fit is based are all valid. Multiple functional forms can give good fits to the data. To the parameter-fitting uncertainty, one must add the uncertainty due to the choice of the functional form of PDFs, as discussed in Sec. IV.C.

As previously outlined, we will look at quantities derived from the fitting procedure that should follow a predicted distribution. Call the quantities  $q_j$ . When testing for a possible systematic deviation from the predicted distribution for the  $q_j$ , we find it useful to transform the observed quantity to a form  $x_j = x(q_j)$  such that the expected distribution of  $x_j$  is  $\mathcal{N}(0, 1)$ , the standard normal (Gaussian) distribution with the mean of 0 and variance of 1. For a quantitative estimate of the probability that the observed distribution of the  $x_j$  was sampled from  $\mathcal{N}(0, 1)$ , we can apply the standard Anderson-Darling test ([Anderson and Darling, 1952](#)). The test yields a “distance”  $A_{\text{obs}}$  of the observed cumulative probability distribution of  $x_j$  values from that for  $\mathcal{N}(0, 1)$ . Then it calculates the probability  $P_{A-D}$  that the same number of randomly drawn  $x_j$  values from  $\mathcal{N}(0, 1)$  will have a distance  $A$  with  $A > A_{\text{obs}}$ . With this test, (a)  $P_{A-D}$  always lies between 0 and 1, (b)  $P_{A-D}$  is close to 1 (or 0) if the histogram matches the  $\mathcal{N}(0, 1)$  distribution closely (or poorly), and (c) if we repeat the sampling procedure many times with data actually drawn from the

$\mathcal{N}(0, 1)$  distribution, the values of  $P_{A-D}$  are uniformly distributed between 0 and 1.

We begin with a preliminary question: do we have enough fitting parameters to obtain a good fit to the data?

### A. Testing with resampled data

The PDFs  $f_{a/p}(\xi, \mu^2)$  must use a sufficiently flexible functional form to reproduce only regular, but no random, features of the hadronic data. However, the functional form for PDFs is known only semiquantitatively based on considerations like the positivity of cross sections, asymptotic limits at small and large  $x$ , and nucleon sum rules. One resorts to a phenomenological form  $f_{a/p}(\xi, \mu_0^2)$  for the PDFs at the initial scale  $\mu_0^2$  and must decide how many parameters  $a_\alpha$  to use. If the number  $N_p$  of parameters is too small, the theory may not be perfect. If there are too many, no global minimum of  $\chi^2$  may exist, or we may overfit the data. In this section, we explore the dependence on the number of free parameters for a given parametrization form. One can also vary the PDF functional forms, as discussed in Sec. IV.C.

To estimate the optimal number of parameters, we return to  $\chi^2$  using Eq. (98) in Eq. (108):

$$\begin{aligned} \chi^2(D, a) &= \sum_{ij} [D_i - \langle D_i \rangle - R_i][D_j - \langle D_j \rangle - R_j] C_{ij} \\ &\quad - 2 \sum_{ij} [D_i - \langle D_i \rangle] C_{ij} T_{j\beta} (a_\beta - \bar{a}_\beta) \\ &\quad + H_{a\beta} (a_\alpha - \bar{a}_\alpha) (a_\beta - \bar{a}_\beta). \end{aligned} \quad (139)$$

Suppose that we obtain a set of parameters  $(a_\alpha)_1$  by minimizing  $\chi^2(D_1, a_1)$  for the fitted data sample  $D_1$  with  $N_D$  data points. Then we use the same parameters to calculate  $\chi^2(D_2, a_1)$  for a control data sample  $D_2$  that is obtained by repeating the experiment with different random fluctuations. ( $N_D$ ,  $R_i$ , and  $C_{ij}$  are the same for  $D_1$  and  $D_2$ .) What do we get? From Eqs. (55), (99), and (103), the  $\chi^2$  expectation for sample  $D_2$ , but using the parameters  $a_1$  fitted to  $D_1$ , is given by

$$\begin{aligned} \langle \chi^2(D_2, a_1) \rangle &= \sum_{ij} \sum_{ij} C_{ij}^{-1} C_{ij} + \sum_{ij} R_i R_j C_{ij} + H_{a\beta} H_{a\beta}^{-1} \\ &= N_D + N_p + \sum_{ij} R_i R_j C_{ij}. \end{aligned} \quad (140)$$

This is bigger by  $+2N_p$  than

$$\langle \chi^2(D_1, a_1) \rangle = N_D - N_p + \sum_{ij} R_i R_j C_{ij}, \quad (141)$$

obtained by using the data sample  $D_1$  to evaluate the fit  $a_1$  obtained from this data. If there are more than just a few parameters, this is a big change.

Although having “too many” parameters makes  $\langle \chi^2(D_2, a_1) \rangle$  larger, we could imagine that with “too few” parameters we get a bad fit because the fit cannot get close to the true PDFs. In our treatment here, with too few parameters  $\sum_{ij} R_i R_j C_{ij}$  is large.

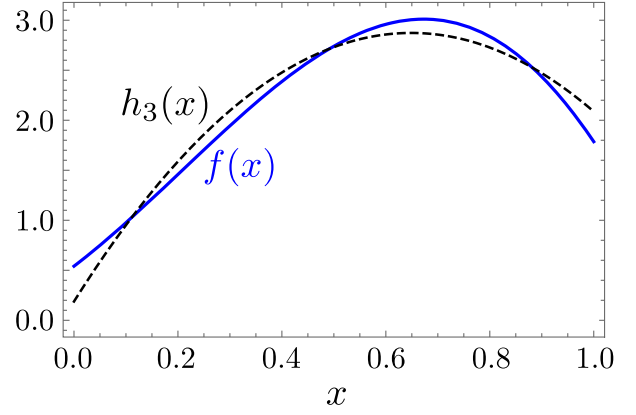


FIG. 2. The function  $f(x)$  in Eq. (142) and a three-parameter polynomial fit  $h_3(x)$  [Eq. (143)] to this function. Here there are not enough parameters to get a good fit.

We can illustrate this with a toy example, showing what may happen in parametrization studies with real parton distributions, such as the one given by Martin *et al.* (2013). Suppose that we fit a test function  $h_{N_p}(x)$  with  $N_p$  parameters to pseudodata generated by random fluctuations around the function

$$f(x) = 3(x + 0.2)^{1.2}(1.2 - x)^{1.2}(1 + 2.3x). \quad (142)$$

For the fit, we use a polynomial

$$h_{N_p}(x) = \sum_{i=1}^{N_p} a_i x^{i-1}. \quad (143)$$

If we use  $h_3(x)$  with just three parameters, we do not get a good fit, as we see in Fig. 2. Here for typical choices  $x_j$  of  $x$ , the measures  $R_j = f(x_j) - h_3(x_j)$  of how well the “theory” matches the exact function are typically of the order of 0.1. However, if we use four parameters to form  $h_4(x)$ , we can get a good fit with  $R_j \sim 0.01$ . If we increase  $N_p$  beyond 4, the measures  $R_j$  are even smaller. Then increasing  $N_p$  when fitting data, with its fluctuations, does not help produce a better fit to the true function  $f(x)$ . However, increasing  $N_p$  does make  $\chi^2$  smaller because we start better fitting the random fluctuations.

To illustrate what happens, we generate “toy data”  $y_i = f(x_i) + 0.2f(x_i)r_i$  at coordinate values  $x_i = 0.01i - 0.05$  for  $i = 1, \dots, N_D$ , where  $N_D = 100$ . These are shown as scattered points in Fig. 3(a). The  $r_i$ ’s are random numbers sampled from  $\mathcal{N}(0, 1)$ . Then fitting the data using  $h_4(x)$  produces the black dashed curve in Fig. 3(a). This is already a fairly good fit to  $f(x)$ , with  $R_j \sim 0.1$ , but we can allow ourselves even more parameters, for example, by fitting  $h_{13}(x)$  with 13 parameters. The 13-parameter fit is shown as the solid blue curve in Fig. 3(a). This produces a smaller value of  $\chi^2$ , but not a better fit to  $f(x)$ .

It appears from Fig. 3(a) that what we are doing is fitting the fluctuations in the data. To test this, we can generate a second set of data  $D_2$  using the same  $f(x)$  and a different set of  $r_i$ . We measure  $\chi^2(D_2, a_1)$  of the original fit  $h_{13}(x)$  to the new data

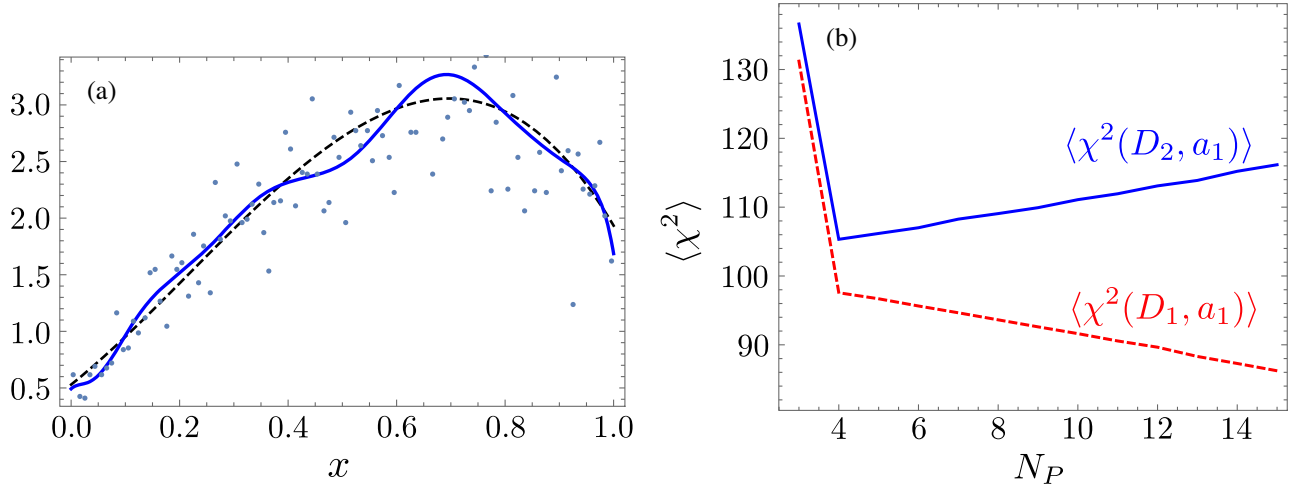


FIG. 3. (a) Data  $\{y_i, x_i\}$  generated from  $f(x)$  [Eq. (142)], shown with a four-parameter (dashed curve) and 13-parameter (solid curve) polynomial fits to the data. (b) Average over a large number of trials of  $\chi^2(D_1, a_1)$  and  $\chi^2(D_2, a_1)$  as a function of the number of parameters  $N_P$ . In each trial, a polynomial with  $N_P$  parameters is fit to data  $D_1$  generated from  $f(x)$ . Then  $\chi^2(D_2, a_1)$  is measured for that polynomial compared to an independent data sample  $D_2$  generated from  $f(x)$ .

sample  $D_2$ . In Fig. 3(b), for each  $N_P$ , we repeat this procedure many times and show  $\chi^2(D_2, a_1)$  averaged over many such trials as a function of  $N_P$ . We see that  $N_P = 3$  is not enough: there is a substantial decrease in  $\langle \chi^2(D_2, a_1) \rangle$  if we increase  $N_P$  to 4. However, beyond  $N_P = 4$ ,  $\langle \chi^2(D_2, a_1) \rangle$  increases with  $N_P$ , in agreement with Eq. (140). The rise of  $\chi^2(D_2, a_1)$  for  $N_P > 4$  is suggestive of overfitting the randomly fluctuating data: while increasing  $N_P$  improves  $\chi^2(D_1, a_1)$  for the fitted sample, when  $N_P$  is too large, it increases  $\chi^2(D_2, a_1)$  for the control sample, indicating that the fit adapts to random fluctuations in  $D_1$ .

Of course, the optimal number of fitted parameters depends on the size of the fluctuations in the data. Having 20% fluctuations in the data is not representative of what one finds in data used in PDF global fits. We can repeat this same exercise using 2% fluctuations. Then the minimum of  $\chi^2(D_2, a_1)$  [as in Fig. 3(b)] occurs at  $N_P = 5$  instead of  $N_P = 4$ .

## B. Dependence on the number of PDF parameters

A strategy comparing  $\chi^2$  values of the fitted and control samples is routinely employed to prevent overfitting of the data in the approach utilizing neural-network (NN) parton distribution functions (Ball *et al.*, 2010, 2013a, 2015, 2017). The PDF of each flavor is given by a NN of a certain configuration, which behaves essentially as a flexible function, with its optimal number of parameters selected so as to satisfy a twofold condition that the resulting PDFs render acceptable fits to the fitted and control samples in each run, or replica, of the global analysis. Both the fitted sample  $D_1$  and the control sample  $D_2$  are obtained by randomly fluctuating the central values of the data according to the Gaussian distributions provided by the standard deviations of the data. The fit consists in training the NN to maximize agreement with the fitted sample. When training the NN on sample  $D_1$ ,  $\chi^2(D_1, a_1)$  is improved to an arbitrary accuracy by training the

neural network long enough. For the control sample  $D_2$ , the  $\chi^2(D_2, a_1)$  initially decreases and then grows after some number of training cycles. The training is stopped when  $\chi^2(D_2, a_1)$  starts growing. The NN obtained at this point most optimally approximates both  $D_1$  and  $D_2$  samples without overfitting  $D_1$ .

In the traditional approach used by groups other than NNPDF, the PDFs are parametrized by a set of fixed functional forms; if the number of free parameters  $N_P$  is too small or too large, the fit is too poor or unstable. For example, the recent MMHT fits approximate each PDF by a Chebyshev polynomial of degree 6 (Martin *et al.*, 2013; Thorne *et al.*, 2019). CT18 error PDFs (Hou *et al.*, 2019) are parametrized with functional forms containing Bernstein polynomials and up to five free parameters.

We can illustrate this behavior with an example using real data. The CT14HERA2 parton distributions (Hou, Sayipjamal *et al.*, 2017) are parametrized using a generic form

$$f_a(x, Q_0) = A_0 x^{A_1} (1-x)^{A_2} P(x; A_3, A_4, \dots). \quad (144)$$

The  $x^{A_1}$  and  $(1-x)^{A_2}$  prefactors capture the typical behavior in the  $x \rightarrow 0$  and  $x \rightarrow 1$  limits, respectively. The function  $P(x; A_3, A_4, \dots)$  is constructed as a linear combination of Bernstein basis polynomials

$$P(x; A_3, A_4, \dots) = \sum_{k=1,2,\dots} A_{k+2} b_{n,k}(x), \quad (145)$$

with  $b_{n,m}(x) = \binom{n}{m} x^m (1-x)^{n-m}$ . Up to four Bernstein polynomials per flavor are introduced in the CT14HERA2 parametrization, with a total of  $N_P = 26$  parameters. However, we can try to choose  $N_P \leq 26$  or  $N_P > 26$ . The number  $N_P$  can be easily varied by adding or removing Bernstein polynomials with nonzero coefficients in the functions  $P(x; A_3, A_4, \dots)$  in  $f_a(x, Q_0)$ .

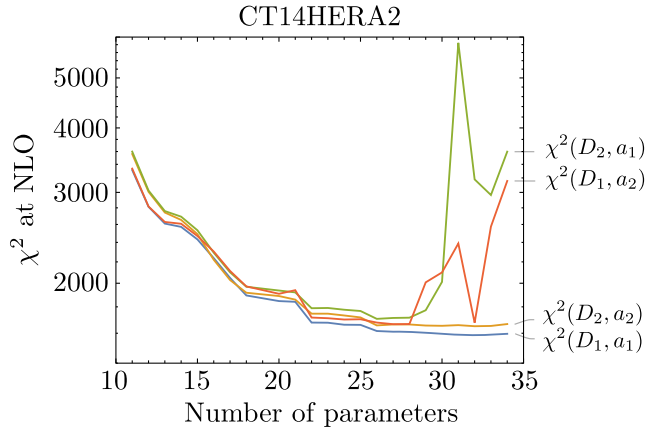


FIG. 4.  $\chi^2$  from the fitted and control samples of data from the CT14HERA2 NLO resampling exercise.

We divide the CT14HERA2 dataset into two equal parts, assigning each datum to the half set  $D_1$  or half set  $D_2$  at random. Then we fit the  $N_p$  parameters to dataset  $D_1$ , giving parameters  $a_1$ . We measure  $\chi^2(D_1, a_1)$  for this fitted data using the data  $D_1$  to which it was fitted. We also measure the  $\chi^2(D_2, a_1)$  for the fitted parameters  $a_1$  using the second half set  $D_2$ . Alternatively, we fit the  $N_p$  parameters to dataset  $D_2$ , giving parameters  $a_2$ . We measure  $\chi^2(D_2, a_2)$  for this fitted data using the data  $D_2$  to which it was fitted. Then we also measure the  $\chi^2(D_1, a_2)$  for this fit using the other dataset  $D_1$ .

We show the results in Fig. 4. As we increase the number of parameters,  $\chi^2(D_1, a_1)$  and  $\chi^2(D_2, a_2)$  decrease. For  $N_p < 28$ ,  $\chi^2(D_2, a_1)$  and  $\chi^2(D_1, a_2)$  also decrease, although there is not much decrease beyond  $N_p = 26$ . For  $N_p > 28$ , the behavior is different. For the fitted samples, the  $\chi^2$  values continue to decrease, although the numerical minimization by the MINUIT program (James and Roos, 1975) becomes less stable. On the other hand,  $\chi^2(D_2, a_1)$  and  $\chi^2(D_1, a_2)$  exhibit large fluctuations for  $N_p > 28$  and, additionally, their values increase dramatically for the largest  $N_p$  in the figure.

We could also examine  $\chi^2(D, a)$  for the entire dataset  $D = D_1 + D_2$ . In this case,  $\chi^2(D, a)$  also initially decreases as  $N_p$  increases, but for large enough  $N_p$  the fits become unstable: first,  $\chi^2$  is no longer a well-behaved function of PDF parameters; for even higher  $N_p$ , a unique best-fit PDF set itself may not be found. This behavior is an indication of overfitting. The number  $N_p$  beyond which overfitting happens depends on the functional forms of PDFs and other factors. With the resampling exercise, we detect traces of overfitting in a half of the full dataset. The presented exercise and other parametrization studies indicate that with the parametrization forms adopted in the latest CT PDFs about 25–30 free parameters is optimal for describing the fitted dataset.

A way to understand the results in Fig. 4 is to note that the analysis in Sec. IV.A of  $\langle \chi^2 \rangle$  was based on the assumptions (from Sec. III.E) that one needs to consider only a small region of the parameter space and that, within this region, the theory predictions  $T_k(a)$  are approximately linear functions of the parameters  $a$ . The analysis then predicts a slow change of  $\chi^2(D_2, a_1)$  [and  $\chi^2(D_1, a_2)$ ] according to Eq. (140) as we

increase the number of parameters  $N_p$ . This is not what we see in Fig. 4.

To help interpret this, suppose that we use the data  $D_1$  and determine the fit parameters  $a_1$  from  $D_1$ . We see in Fig. 4 a large increase of  $\chi^2(D_2, a_1)$  and large fluctuations for  $N_p > 28$ . What apparently happens is that when there are too many parameters the data do not strongly constrain the components of  $a$  in one or more “flat” directions in the parameter space. The value of  $\chi^2(D_1, a)$  changes little as  $a$  varies in these directions. One needs to move through a wide range of  $a$  in a flat direction before  $\chi^2(D_1, a)$  increases significantly. In this wide range, the  $T_k(a)$  are not approximately linear functions of  $a$ , and  $\chi^2(D_1, a)$  is not approximately quadratic. Then  $\chi^2(D_1, a)$  may have multiple minima at locations that correspond to unreasonable behavior of the PDFs. The best-fit parameters  $a_1$  can be at an unphysical location that strongly depends on the fluctuations in the data  $D_1$ . We can test how reasonable the fit  $a_1$  is by testing it against the second dataset  $D_2$ . With too many parameters, we can get the large values of  $\chi^2(D_2, a_1)$  that we see in Fig. 4. We conclude that if one uses a family of functions for the PDFs that have a variable number of parameters, then a simple test like that shown in Fig. 4, ideally repeated for several random partitions of the data in two halves, can estimate how many parameters one should allow.

### C. Dependence on the PDF functional form

There is another sort of test available. The PDFs are unknown functions, but one represents them using fixed functional forms with a finite number of unknown parameters.<sup>12</sup> PDF parametrization studies thus constitute an essential step in global analyses. In Sec. IV.B, we examined the dependence of the fit results on the number of parameters within one general functional form. One can also change the general functional form. Thus, one should examine the tests listed previously and later in this section for a large class of PDF functional forms.

Universal approximation theorems (Cybenko, 1989; Hornik, Stinchcombe, and White, 1990; Hornik, 1991) demonstrate that feed-forward neural networks with one or multiple layers can approximate any continuous function and its derivatives to arbitrary accuracy. This fundamental result motivates the Neural Network PDF approach that parametrizes parton distribution functions by neural networks trained on the fitted data.

Alternatively, modern parametrizations based on Chebyshev or Bernstein polynomials also allow one to parametrize a variety of functional behaviors (Pumplin, 2010; Glazov, Moch, and Radescu, 2011; Martin *et al.*, 2013; Gao and Nadolsky, 2014; Hou *et al.*, 2019). Suppose that two choices for functional forms give good fits in the sense of reasonably meeting the array of goodness-of-fit criteria discussed in this section, but that they also give two sets of PDFs that differ outside of their parameter-fitting error estimates. This suggests that there is a theoretical systematic error associated with the choice of PDF functional form that needs

<sup>12</sup>The neural net approach avoids this limitation.

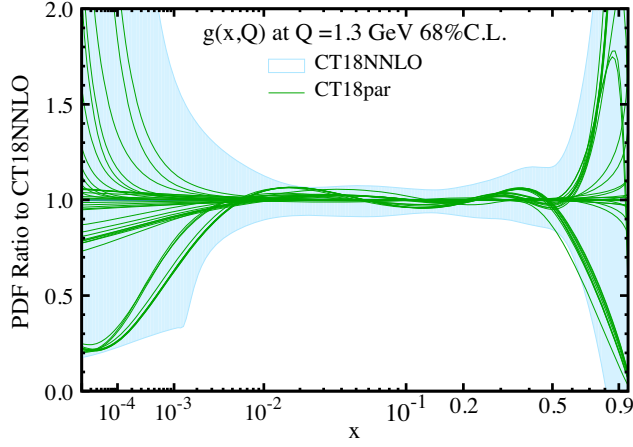


FIG. 5. Green solid lines: gluon PDFs from candidate fits of the CT18 NNLO analysis, obtained using alternative parametrization forms and plotted as ratios to the default CT18 NNLO  $g(x, Q)$ . Light blue band: 68% C.L. uncertainty band of the published CT18 NNLO PDFs. From Hou *et al.*, 2019.

to be added to the analysis. In a global analysis such as CT18, multiple parametrization forms are tried to estimate this source of PDF uncertainty. For example, the green solid lines in Fig. 5 illustrate predictions for the best-fit gluon PDFs obtained with about 100 candidate CT18 fits based on different parametrization forms. Notice that the solid lines show large deviations from each other for large  $x$  and small  $x$ , regions in which the gluon distribution is not well constrained by the data. The PDF uncertainty of the published CT18 NNLO set, shown as a blue band, is computed using a prescription that is broad enough to cover variations due to the choice of parametrization form; for details, see Hou *et al.* (2019).

#### D. Closure test

Given a PDF functional form depending on parameters  $a$ , PDF fitting of the data determines best-fit parameters  $a_{\text{fit}}$ . The fitting procedure is based on Eq. (58), which asserts that if the theory represented by parameters  $\bar{a}$  is correct, then in an ensemble of trials the data  $D_j$  is distributed with a probability  $P(D|T(\bar{a}))$  that involves the theory predictions  $T_i(\bar{a})$  and the covariance matrix and the covariance matrix  $C_{ij}$ . Then the parameters  $a^{\text{fit}}$  obtained from these data are distributed according to Eq. (104). We can test this. We can generate a set of pseudodata  $D'_k$  using Eq. (52) with the truth values  $\langle D'_k \rangle$  set equal to theory predictions  $T_i(a)$  for some realistic parameters  $a$ , and with  $\sigma_k$  and  $\beta_{k,j}$  obtained from the experimental groups. Then we can run the fitting procedure with the generated data  $D'_k$ , giving a new fit  $a'_{\text{fit}}$ . We should find that the new fit  $a'_{\text{fit}}$  agrees with the original parameters  $a$  within the errors generated by the fit. If they do not agree, we should understand why. This “closure test” is not commonly carried out by fitting groups who use the Hessian method described in this review, but it is a feature of fits by the NNPDF group (Ball *et al.*, 2015).

#### E. Test of the nuisance parameters

We now describe a goodness-of-fit test based on the distribution of the nuisance parameters. For this test, it is

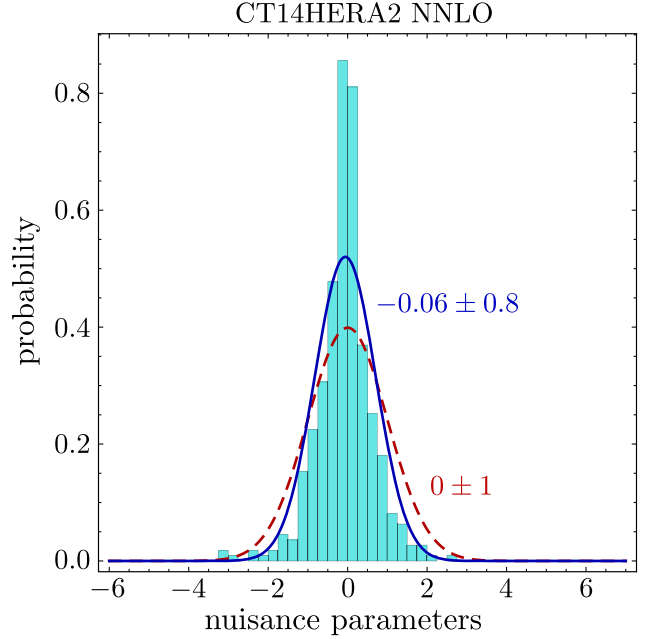


FIG. 6. Distribution of nuisance parameters  $\lambda_j^{\text{fit}}$  for the CT14 HERA2 NNLO parton distribution fit.

useful to use the form of  $\chi^2$  in which nuisance parameters  $\lambda_j$  appear explicitly,  $\chi^2(D, a, \lambda)$  as given in Sec. III.D and Eq. (66). Then we can fit values  $\lambda_{\text{fit}}$  of  $\lambda$  by minimizing  $\chi^2(D, a, \lambda)$ . The minimum value is  $\chi^2(D, a)$ , expressed in terms of the covariance matrix  $C_{ij}$  in Eq. (59). We have  $\chi^2(D, a_{\text{fit}}, \lambda_{\text{fit}}) = \chi^2(D, a_{\text{fit}})$ . We also argue after Eq. (75) that  $\lambda_j^{\text{fit}} \approx \bar{\lambda}_j$  in an accurate fit with enough data. Since the  $\bar{\lambda}_j$  are independent random variables distributed according to  $\mathcal{N}(0, 1)$ , the  $\lambda_j^{\text{fit}}$  are expected to follow the  $\mathcal{N}(0, 1)$  distribution, too. We can test these assumptions by making a histogram of  $\lambda_j^{\text{fit}}$ .

We show in Fig. 6 a histogram of the best-fit nuisance parameters  $\lambda_j^{\text{fit}}$  for the CT14 HERA2 NNLO fit along with a dashed, red curve showing the expected Gaussian distribution. The correlated systematic errors are implemented in the CT14HERA2 fit using the matrices  $\beta_{k,j}$  provided with the experimental datasets. The observed distribution is substantially narrower than the expected distribution. The mean for the observed distribution is  $-0.06$ , which is close to 0, but its standard deviation is 0.8, which is noticeably smaller than 1. We also show a blue solid curve giving a Gaussian distribution with this mean and standard deviation. This curve also does not match the observed distribution well.

For a quantitative estimate of the probability that the observed distribution was sampled from  $\mathcal{N}(0, 1)$ , we can apply the standard Anderson-Darling test (Anderson and Darling, 1952) described at the beginning of Sec. IV. For the histogram in Fig. 6, we find that  $P_{\text{A-D}} \sim 10^{-6}$ . This indicates that it is unlikely that the  $\lambda_j^{\text{fit}}$  values were generated from the expected  $\mathcal{N}(0, 1)$  distribution, as was self-evident from the figure.

There are a few more  $|\lambda_j^{\text{fit}}|$  that are larger than 2 than would be expected. However, the main feature that we see in Fig. 6 is that too many  $|\lambda_j^{\text{fit}}|$  are small, indicating possibly that the

corresponding  $\beta_{kJ}$  values are overestimated. This could indicate that the estimates are conservative in the sense that if one suspects that the  $\beta_{kJ}$  values for a source  $J$  of systematic error should be smaller but one cannot prove it with solid evidence, then the conservative approach is to leave these  $\beta_{kJ}$  values unchanged.

There is another reason that the values of some of the  $\lambda_J^{\text{fit}}$  might be smaller than expected. The expectation from Eq. (75) that  $\lambda_J^{\text{fit}} \approx \bar{\lambda}_J$  was based on the assumption that for each dataset  $E$  there are many pieces of data and few nuisance parameters. What happens if this is not the case, as suggested near the end of Sec. III.D? Suppose that we add a spurious nuisance parameter  $\lambda_L$  with  $\beta_{kL} = 0$  for all data indices  $k$ . Then, in Eq. (66) for  $\chi^2(D, a, \lambda)$ , the spurious  $\lambda_L$  does not contribute to the first term, which involves data, and contributes only to the “penalty” term  $\sum_J \lambda_J^2$ . Then minimizing  $\chi^2(D, a, \lambda)$  with respect to the  $\lambda_J$  gives  $\lambda_L^{\text{fit}} = 0$ . If a certain dataset  $E$  has many systematic errors and thus many nuisance parameters, it may be that some of their linear combinations are spurious in this sense. Then the same linear combinations of the fitted values  $\lambda_J^{\text{fit}}$  vanish so that the fitted values do not spread out over the entire  $\lambda_J$  space but only a subspace. In this case, the absolute values of these  $\lambda_J^{\text{fit}}$  will be smaller than would be given by an  $\mathcal{N}(0, 1)$  distribution. In case there are many nuisance parameters, some of which have little effect, one can approximate the matrix  $C^{-1}$  in Eq. (56) so as to leave effectively fewer, but better behaving, nuisance parameters, as described in Appendix B of Hou *et al.* (2019). We suggest that, when performing a fit, it is useful to check the  $\lambda_J^{\text{fit}}$  distribution. Values of  $\lambda_J^{\text{fit}}$  that are smaller than expected from an  $\mathcal{N}(0, 1)$  distribution suggest problems with the systematic errors used in the fit but are perhaps not of great concern. An excessive number of large  $\lambda_J^{\text{fit}}$  values could indicate a more serious difficulty that needs to be resolved.

#### F. Test of data residuals

Equation (66) gives  $\chi^2(D, a_{\text{fit}}, \lambda_{\text{fit}})$  as

$$\chi^2(D, a_{\text{fit}}, \lambda_{\text{fit}}) = \sum_k [r_k(a_{\text{fit}}, \lambda_{\text{fit}})]^2 + \sum_J [\lambda_J^{\text{fit}}]^2, \quad (146)$$

where

$$r_k(a_{\text{fit}}, \lambda_{\text{fit}}) = \frac{D_k - T_k(a_{\text{fit}})}{\sigma_k} - \sum_I \beta_{kI} \lambda_I^{\text{fit}} \quad (147)$$

is called the residual for datum  $k$  obtained in the fit. Using Eqs. (52), (77), and (78) with  $R_k = 0$ , this is

$$r_k(a_{\text{fit}}, \lambda_{\text{fit}}) = \Delta_k - \frac{T_{k\alpha}}{\sigma_k} (a_{\alpha}^{\text{fit}} - \bar{a}_{\alpha}) - \sum_I \beta_{kI} (\lambda_I^{\text{fit}} - \bar{\lambda}_I). \quad (148)$$

The  $\Delta_k$  introduced in Eq. (52) follow independent  $\mathcal{N}(0, 1)$  distributions. The values  $(a_{\alpha}^{\text{fit}} - \bar{a}_{\alpha})$  and  $(\lambda_I^{\text{fit}} - \bar{\lambda}_I)$  have expectation values of zero, but they have fluctuations that arise from the fluctuations in the data. If there are enough data, we expect the fluctuations of  $(a_{\alpha}^{\text{fit}} - \bar{a}_{\alpha})$  and  $(\bar{\lambda}_I - \lambda_I^{\text{fit}})$  to be small. Then the residuals  $r_k$  should also be approximately distributed as  $\mathcal{N}(0, 1)$ .

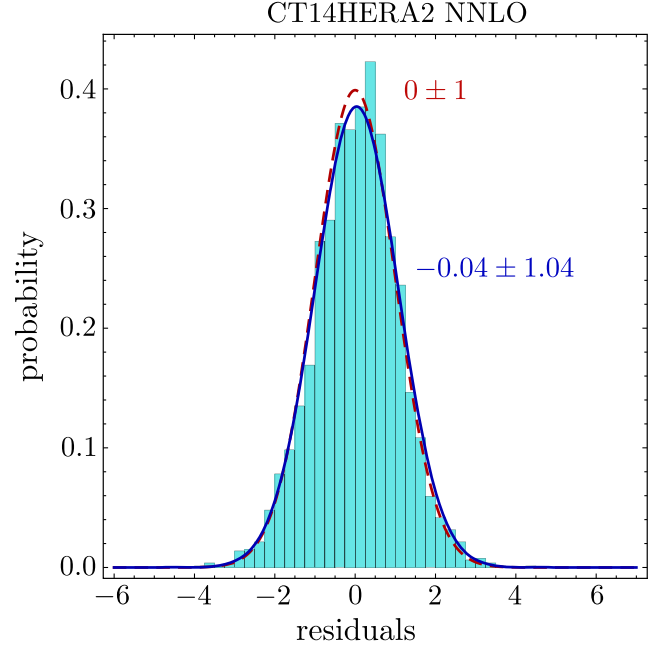


FIG. 7. Distribution of residuals for the CT14 HERA2 NNLO parton distribution fit.

We can test these assumptions by making a histogram of the values  $r_k$  obtained. We show in Fig. 7 a histogram of the residuals  $r_k$  for the CT14 HERA2 NNLO fit along with a curve showing the expected Gaussian distribution. Comparing these using the Anderson-Darling test gives  $P_{\text{A-D}} = 5.7 \times 10^{-3}$ . Thus, we can conclude with some confidence that the observed distribution of residuals was not drawn exactly from  $\mathcal{N}(0, 1)$ . However, we judge the difference between the two distributions not to be physically significant. After all, we expect the observed distribution to be only approximately an  $\mathcal{N}(0, 1)$  distribution. The mean for the observed distribution is 0.04, which is close to 0, and its standard deviation is 1.04, which is close to 1. We also show a blue solid curve giving a Gaussian distribution with this mean and standard deviation.

The distribution of residuals is another indicator that should be checked when performing a fit. In this case, no large discrepancies are observed.

#### G. Value of $\chi^2$ from an individual experiment

In this section, we describe a goodness-of-fit test based on a decomposition of the data into subsets. We sometimes refer to a subset of the data as an “experiment,” although we could divide the data into subsets in different ways.

Label the subset of the data that we wish to consider by an index  $E$ . Let  $D(E)$  refer to the data in subset  $E$ , that is, all data points  $D_i$  for  $i \in E$ . Recall Eq. (58), giving the probability of finding data  $D$  if the theory  $T(a)$  is correct. The analog of this that gives the probability of finding data  $D(E)$  if the theory  $T(a)$  is correct is

$$P(D(E)|T(a)) = d\mu(D) \exp[-\frac{1}{2}\chi^2(D(E), a)], \quad (149)$$

where

$$\chi^2(D(E), a) = \sum_{i,j \in E} [D_i - T_i(a)][D_j - T_j(a)] C_{ij}, \quad (150)$$

$d\mu(D(E))$  is the measure

$$d\mu(D(E)) = (2\pi)^{-N_E/2} \sqrt{\det C_E} d^{N_E} D, \quad (151)$$

$C_E$  is the matrix  $C_{ij}$  for  $i, j \in E$ , and  $N_E$  is the number of data in the subset  $E$ .

If we have a good fit, then the probability  $P(D(E)|T(a))$  should not be too small for each subset  $E$  of the data. That is, for each subset  $E$ ,  $\chi^2(D(E), a)$  should not be too large.

As mentioned, this is a much stronger criterion than the hypothesis-testing criterion. An individual experiment  $E$  may be badly fit in a large global fit [have an unacceptably high  $\chi^2(D(E), a)$ ], even while the total  $\chi^2(D, a)$  may look reasonable. In this section we ask, how large is too large for  $\chi^2(D(E), a)$ ? What should the distribution of this quantity be?

Consider the structure of the covariance matrix  $C_{ij}$  defined by Eq. (56). Suppose first that each experiment has independent systematic errors that are not shared among the experiments so that each source  $J$  of systematic error is associated with just one experiment  $E_J$ . Then  $\beta_{iJ} = 0$  unless  $i \in E_J$ . The covariance matrix is then block diagonal,  $C_{ij} = C_{ij}^{-1} = 0$ , unless  $i \in E$  and  $j \in E$  for the same experiment label  $E$ . Then the total  $\chi^2(D, a)$  is a sum of contributions  $\chi^2(D(E), a)$  from the separate experiments

$$\chi^2(D, a) = \sum_E \chi^2(D(E), a). \quad (152)$$

It is, however, not necessary that errors for experiment  $E$  are uncorrelated with the errors from other experiments. If for some of the datasets  $E$  the covariance matrix has elements  $C_{ij}$  that are nonzero for  $i \in E$  and  $j \notin E$ , then Eq. (152) will fail. However, if the theory  $T(a)$  is correct, the probability of finding data  $D(E)$  is still given by Eq. (149).

We return to Eq. (150). There are  $N_E$  data in the set  $E$ . Then, with the parameters  $a$  fixed to ideal values ( $a = \bar{a}$ , as in Sec. III.E) and flawless theory ( $R_k = 0$ ), we find that  $T_k(\bar{a}) = \langle D_k \rangle$ , and the distribution of  $\chi^2(D(E), \bar{a})$  is the standard  $\chi^2$  distribution with  $N_E$  degrees of freedom.

If, however, we use the best-fit parameters  $a_{\text{fit}}$  that are constrained by experiment  $E$  as well as the rest of the experiments, we find that  $\chi^2(D(E), a_{\text{fit}})$  is approximately equal to the standard  $\chi^2$  distribution with  $N_E$  degrees of freedom up to a subleading term that can be determined as follows.

From Eqs. (77), (78), and (59), we find that

$$T_k(a) = \langle D_k \rangle + R_k + T_{k\alpha}(a_\alpha - \bar{a}_\alpha) \quad (153)$$

and

$$\begin{aligned} \chi^2(D(E), a) = & \sum_{i,j \in E} [D_i - \langle D_i \rangle - T_{i\alpha}(a_\alpha - \bar{a}_\alpha) - R_i] \\ & \times [D_j - \langle D_j \rangle - T_{j\beta}(a_\beta - \bar{a}_\beta) - R_j] C_{ij}. \end{aligned} \quad (154)$$

What is the expectation value of this at  $a = a_{\text{fit}}$ ? We can refer to Eqs. (99), (103), (55) and (101) for  $a = a_{\text{fit}}$  to arrive at

$$\langle \chi^2(D(E), a_{\text{fit}}) \rangle = N_E - N_P(E) + \sum_{i,j \in E} R_i R_j C_{ij}, \quad (155)$$

where

$$N_P(E) = \sum_{i,j \in E} T_{i\alpha} T_{j\beta} C_{ij} H_{\alpha\beta}^{-1}. \quad (156)$$

The first term in Eq. (155) is simply the number of data  $N_E$  in the dataset. The third term is small if the theory is good ( $R_i \approx 0$ ). The second term  $-N_P(E)$  requires some analysis. We interpret  $N_P(E)$  as the effective number of parameters constrained by dataset  $E$ . To support this interpretation, we note from Eqs. (86) and (137) that if  $N_P(E)$  is summed over all datasets  $E$ , it gives the number  $N_P$  of parameters. Now the number of parameters  $N_P$  is much smaller than the total number of data  $N_D$ . This makes it plausible that  $N_P(E)$  is small compared to the number  $N_E$  of data in dataset  $E$   $N_E \gg N_P(E)$ . Then the first term in Eq. (155) dominates.

## H. Test of $\chi^2$ from individual experiments

What can we do with  $\chi^2(D(E), a_{\text{fit}})$ ? Its value is given by Eq. (154) with  $a = a_{\text{fit}}$ . When the best-fit parameters are close to the true ones ( $a_{\text{fit}} \approx \bar{a}$ ) and theory is nearly perfect ( $R_k \approx 0$ ), the  $\chi^2(D(E), a_{\text{fit}})$  distribution reduces to the form

$$\chi^2(D(E), a_{\text{fit}}) = \sum_{i,j \in E} (D_i - \langle D_i \rangle)(D_j - \langle D_j \rangle) C_{ij} + \dots$$

that, as we already know, obeys the  $\chi^2$  distribution with  $N_E$  degrees of freedom. Now  $a_{\text{fit}}$  is not exactly  $\bar{a}$  but is influenced by the data in dataset  $E$ . Thus, one might think that the number of degrees of freedom is  $N_E - N_P(E)$ , where  $N_P(E)$  [Eq. (156)] is the effective number of fit parameters associated with dataset  $E$ . However, the reasoning in Sec. IV.G suggests that  $N_P(E) \ll N_E$  so that the number of degrees of freedom is approximately simply the number  $N_E$  of data in dataset  $E$ . This expectation can be tested by calculating  $N_P(E)$  in the case in which  $\chi^2(D(E), a_{\text{fit}})$  appears to be anomalously small.

We now check to see whether the distributions of the observed  $\chi^2(D(E), a_{\text{fit}})$  values from the experiments  $E$  in actual PDF fits are close to the ideal distributions.

When  $N_E$  is large, say,  $N_E \gtrsim 30$ , the  $\chi^2$  distribution with  $N_E$  degrees of freedom approaches the Gaussian distribution with mean  $\langle \chi^2(D(E), a_{\text{fit}}) \rangle \approx N_E$  and standard deviation  $\langle [\chi^2(D(E), a_{\text{fit}}) - N_E]^2 \rangle \approx 2N_E$ , as we see in Sec. III.H. For  $N_E \lesssim 30$ , the non-Gaussian features are pronounced and  $\chi^2$  distributions with different  $N_E$  are not easily compared. Conveniently for our purpose, the variable

$$S_E \equiv \sqrt{2\chi^2(D(E), a_{\text{fit}}) - \sqrt{2N_E - 1}} \quad (157)$$

fluctuates with a distribution that is accurately<sup>13</sup> an  $\mathcal{N}(0, 1)$  distribution (Fisher, 1925; Lai, Guzzi *et al.*, 2010), namely,

<sup>13</sup>Other definitions (Lewis, 1988) of  $S_E$  are more accurate but with the simple form of Eq. (157), the distribution function  $\rho(S_E)$  matches the Gaussian distribution  $(2\pi)^{-1/2} \exp(-S_E^2/2)$  to within 0.04 for  $N_E = 5$  and to within 0.01 for  $N_E = 50$ .



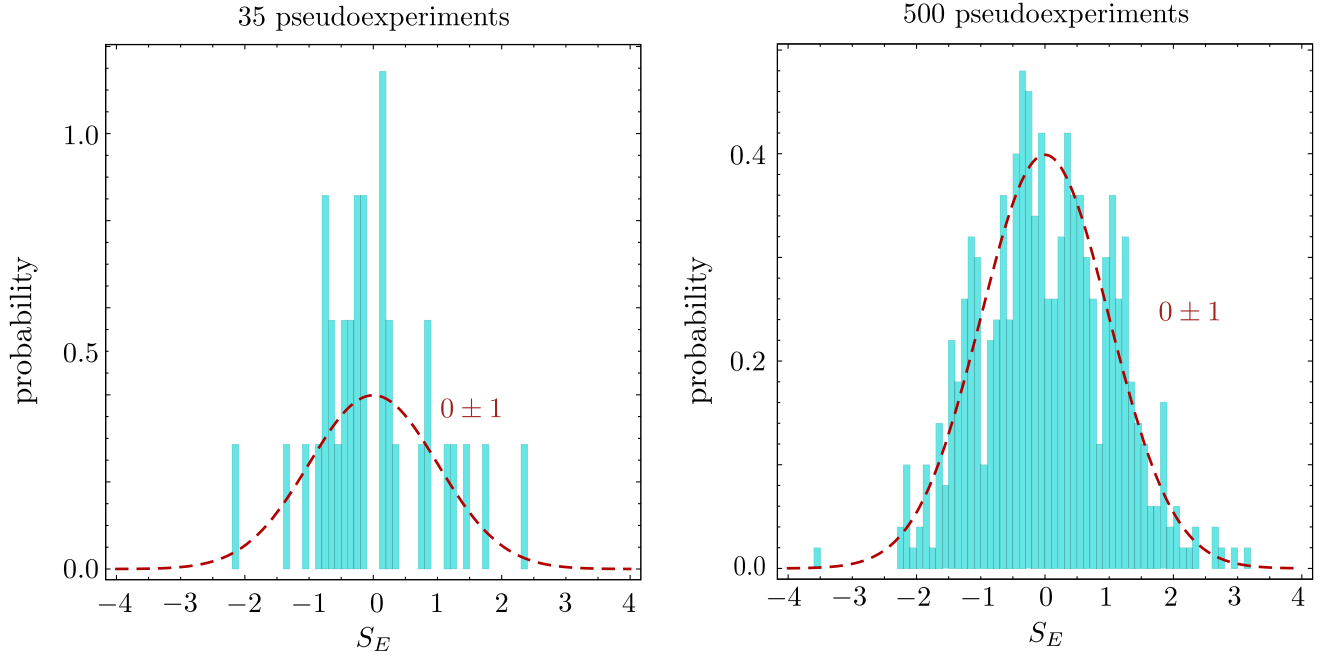


FIG. 8. Probability distributions of  $S_E = \sqrt{2\chi^2(D(E), a_{\text{fit}})} - \sqrt{2N_E - 1}$  for 35 and 500 random pseudoexperiments, each of which has the number  $N_E$  of data chosen at random in the range  $0 \leq N_E \leq 3000$ . The red dashed line shows the  $\mathcal{N}(0, 1)$  Gaussian distribution, which describes the observed probabilities well.

$$\rho(S_E) \approx (2\pi)^{-1/2} \exp(-S_E^2/2). \quad (158)$$

Note that the  $S_E$  distribution is independent of  $N_E$ . The original  $N_E$  dependence for the distribution of  $\chi^2$  is absorbed into the definition of  $S_E$ .

To test the quality of the fit, we can plot a histogram of the  $S_E$  values for all of the experiments (or datasets)  $E$  contributing to the fit. The histogram should match the Gaussian distribution (158).

To see how this should work, we can generate  $S_E$  values for a number of randomly generated pseudoexperiments. The number  $N_E$  of data for each pseudoexperiment is chosen at random between 0 and 3000. For each pseudoexperiment, we generate a value of  $\chi_E^2$  at random according to the standard  $\chi^2$  distribution with  $N_E$  degrees of freedom. Then we define  $S_E$  for that pseudoexperiment by  $S_E = \sqrt{2\chi_E^2} - \sqrt{2N_E - 1}$ . In the left-hand plot in Fig. 8, we show the resulting histogram of  $S_E$  values obtained for 35 pseudoexperiments, along with the expected Gaussian distribution (158). In the right-hand plot in Fig. 8, we show the analogous histogram for 500 pseudoexperiments.

The histogram in Fig. 8 for 500 pseudoexperiments is fairly close to the expected distribution (158). For the histogram for 35 pseudoexperiments, this is not clear merely by eye. For a quantitative estimate, we can apply the standard Anderson-Darling test, described in Sec. IV.E, of the probability that the observed distribution matches  $\mathcal{N}(0, 1)$ . For the left-hand histogram in Fig. 8, we find that  $P_{A-D} = 0.53$ , and for the right-hand histogram we find that  $P_{A-D} = 0.44$ . These values indicate that it is plausible that the  $S_E$  values were generated from  $\mathcal{N}(0, 1)$ , which, to a good approximation, they were.

Now we turn to the distributions of  $S_E$  from recent NNLO global analyses shown in Fig. 9. The parameters  $S_E$  in the

histograms are computed using Eq. (157) from  $\chi^2$  and  $N_E$  values listed in Table 5 of Harland-Lang *et al.* (2015) for MMHT 2014 NNLO, Tables 2.1–2.3 and 3.1 of Ball *et al.* (2017) for NNPDF3.1 and NNPDF3.0, and internal tables of  $\chi^2$  for CT14HERA2 NNLO (Hou, Sayipjamal *et al.*, 2017). Variations in  $S_E$  are broader than the standard normal distribution expected in an ideal fit to all experiments, both in the positive and negative directions. We can estimate the differences by the mean and standard deviation for each observed distribution. For CT14HERA2 and MMHT2014 fits, the means are close to zero, indicating that while some experiments are not fit well, the other experiments are fit too well. On the other hand, for the NNPDF3.0 and NNPDF3.1 analyses, the observed mean is of the order of 0.8 to 1: more experiments are not fitted well than fitted too well. For the four fits, the probability values for matching the expected  $\mathcal{N}(0, 1)$  distribution according to the Anderson-Darling test are

$$\begin{aligned} P_{A-D} &= 6.4 \times 10^{-3}, & \text{CT14HERA2 NNLO,} \\ P_{A-D} &= 6.4 \times 10^{-3}, & \text{MMHT2014 NNLO,} \\ P_{A-D} &= 2.6 \times 10^{-3}, & \text{NNPDF3.0 NNLO,} \\ P_{A-D} &= 1.6 \times 10^{-3}, & \text{NNPDF3.1 NNLO.} \end{aligned} \quad (159)$$

In all four cases, it is highly unlikely that the observed distribution came from the expected Gaussian distribution.

We emphasize that none of the four PDF fits described here are good fits according to the  $P_{A-D}$  values obtained by breaking the data into smaller datasets, even though each fit is acceptable according to its total  $\chi^2$  value.

The expectation that the  $S_E$  distribution should match an  $\mathcal{N}(0, 1)$  distribution is based in part on the assumption that the

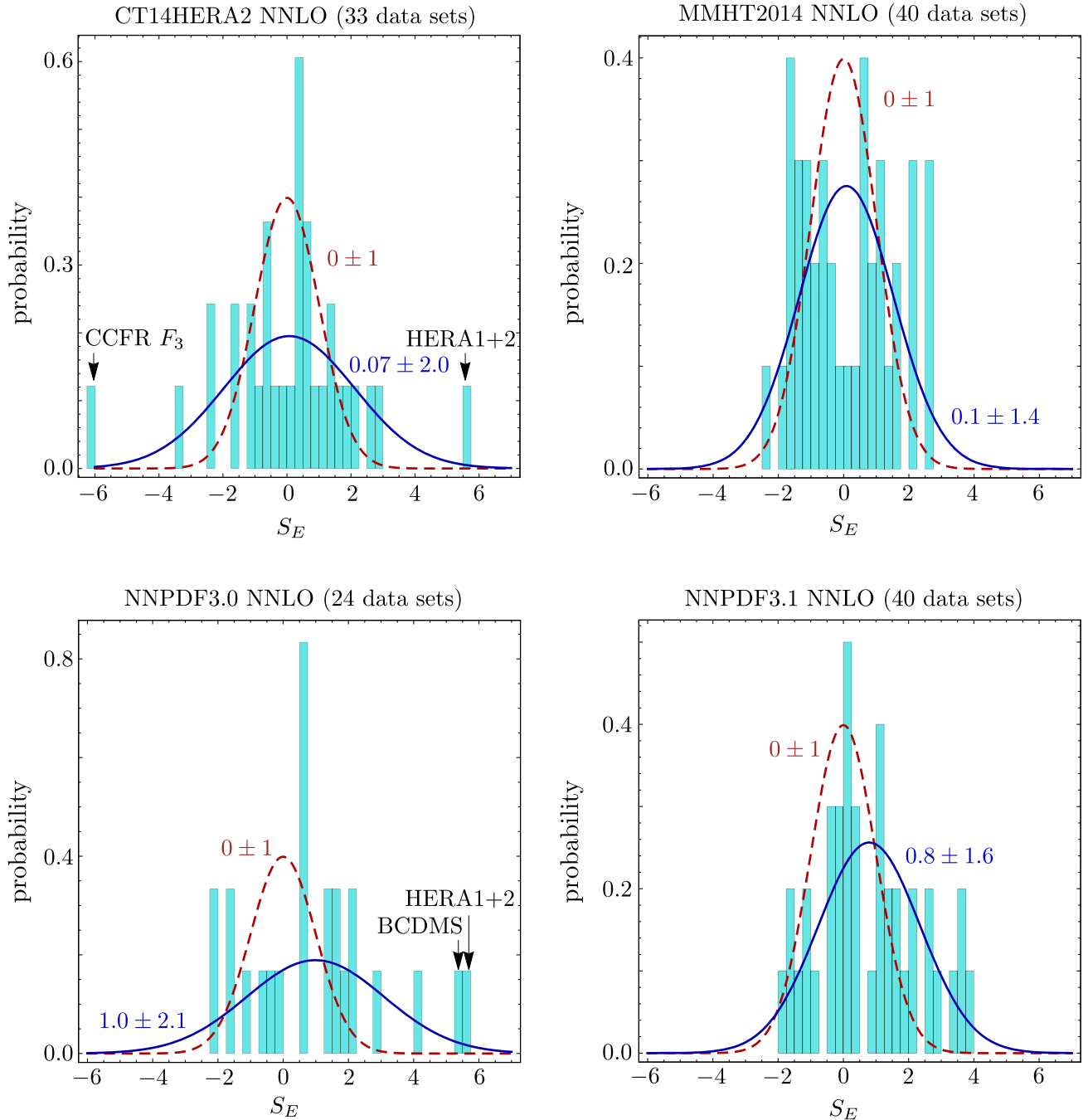


FIG. 9. Probability distributions in the effective Gaussian variable  $S_E$  for  $\chi^2$  values of the fitted datasets from the NNLO fits CT14HERA2, MMHT2014, NNPDF3.0, and NNPDF3.1.

parameters  $R_k$  representing imperfections in the theory are negligible in Eq. (155). The evident failure of the distributions in Fig. 9 to match  $\mathcal{N}(0, 1)$  may indicate that the theory is not precise enough to match extremely precise experiments. In fact, some elevated  $S_E$  values are contributed by the most precise experiments, such as the combined HERA 1 + 2 DIS data (Abramowicz *et al.*, 2015) and some LHC measurements. These experiments test QCD at unprecedented (NNLO) precision and thus may reveal evidence for new dynamical mechanisms. For instance,  $S_E \approx 5.5$  for HERA 1 + 2 DIS data can be reduced to  $S_E \approx 3$  by including small- $x$  resummation in DIS or by evaluating NNLO DIS cross sections with an

$x$ -dependent factorization scale; see the discussion in Sec. I.F. Similarly, the description of HERA 1 + 2 DIS and fixed-target DIS data such as BCDMS (Benvenuti *et al.*, 1990) is improved in the NNPDF3.1 analysis compared to NNPDF3.0, in part by introducing the “fitted charm,” an independent and possibly process-dependent nonperturbative function that is similar to power-suppressed (“higher-twist”) terms in DIS. In Sec. II.C, we briefly review the rationale for optionally including fitted charm in some PDF fits and its possible interpretation.

A different class of concerns arises when too many datasets have negative  $S_E$ , indicating that the datasets are

systematically described better than would be expected in a good fit in which the data fluctuations around the best-fit theory are random. When too many experiments have negative  $S_E$ , the reduced global  $\chi^2$  may hide some problems, such as overestimated experimental uncertainties or overfitting of the statistical fluctuations because of using too flexible theory. Sometimes, a low global  $\chi^2 \ll N_D - N_P$  is taken as evidence for perfect theory that justifies aggressive estimates for PDF uncertainties based exclusively on the  $\Delta\chi^2 = 1$  criterion. In fact, the resulting small PDF uncertainties would be wrong: not only it is improbable that the low  $\chi^2$  is caused by random fluctuations consistent with the experimental errors, but the parametrization uncertainty also needs to be estimated by trying other PDF parametrization forms that render  $\chi^2$  of up to about  $N_D - N_P$ .

### I. Test of consistency between experiments

We can carry this analysis further by asking whether different experiments have imposed consistent constraints on PDF parameters  $a$ . To this end, we consider an observable  $\sigma(a)$  that depends on the parton parameters, as in Sec. III.G. The observable could be a cross section, as suggested by the notation, or, extending the notion of an observable a bit, it could be the value of the PDF  $f_{a/p}(x, \mu^2)$  for a specific flavor at a particular momentum fraction  $x$  and scale  $\mu$ .

As in Sec. III.G, as long as we consider parameters  $a$  that are not far from the best-fit parameters  $a_{\text{fit}}$ , we can apply a linear approximation for the evaluation of  $\sigma(a)$

$$\sigma(a) = \sigma(a_{\text{fit}}) + (a_\alpha - a_\alpha^{\text{fit}})\sigma'_\alpha, \quad (160)$$

with  $\sigma'_\alpha = \partial\sigma(a)/\partial a_\alpha$ . Furthermore, if we define a special vector  $e(\sigma)$  according to Eq. (116),

$$e(\sigma)_\alpha = \frac{H_{\alpha\beta}^{-1}\sigma'_\beta}{\sqrt{\sigma'_\gamma H_{\gamma\delta}^{-1}\sigma'_\delta}}, \quad (161)$$

then as we found in Sec. III.G, we can evaluate  $\sigma(a) - \sigma(a_{\text{fit}})$  by setting

$$a = a_{\text{fit}} + te(\sigma) \quad (162)$$

in Eq. (160). Variations of  $a - a_{\text{fit}}$  in the orthogonal directions  $\{e(\sigma)^{(2)}, e(\sigma)^{(3)}, \dots\}$  do not contribute to  $\sigma(a)$ . That is,

$$\begin{aligned} & \sigma\left(a_{\text{fit}} + te(\sigma) + \sum_{n \geq 2} t_n e(\sigma)^{(n)}\right) \\ &= \sigma(a_{\text{fit}}) + te(\sigma)_\alpha \sigma'_\alpha \\ &= \sigma(a_{\text{fit}}) + \sqrt{\sigma'_\gamma H_{\gamma\delta}^{-1}\sigma'_\delta} t. \end{aligned} \quad (163)$$

The result is independent of the parameters  $t_n$ . That is,  $t$  directly measures  $\sigma(a)$ . We denote the corresponding cross section by  $\sigma_0(t)$ .

The parameters  $a_{\text{fit}}$  correspond to the minimum of the global  $\chi^2$  so that, according to Eq. (119),

$$\chi^2(D, a_{\text{fit}} + te(\sigma)) = \chi^2(D, a_{\text{fit}}) + t^2. \quad (164)$$

Furthermore, if we evaluate Eq. (119) at a general point  $a_{\text{fit}} + te(\sigma) + \sum_n t_n e(\sigma)^{(n)}$ , we get

$$\chi^2\left(D, a_{\text{fit}} + te(\sigma) + \sum_{n=2}^{N_P} t_n e(\sigma)^{(n)}\right) = \chi^2(D, a_{\text{fit}}) + t^2 + \sum_{n=2}^{N_P} t_n^2. \quad (165)$$

If we regard parameter points  $a$  as distributed at random according to a probability density proportional to  $\exp[-(\chi^2 - \chi^2_{\text{min}})/2]$ , then, to find the probability  $\rho$  of  $a$  lying in a plane of constant  $\sigma(a) = \sigma_0(t)$ , we simply integrate over the other variables  $t_n$  as follows:

$$\begin{aligned} \rho &= (2\pi)^{-N_P/2} \int dt_2 \cdots dt_{N_P} \exp\left[-\left(t^2 + \sum_{n \geq 2} t_n^2\right)/2\right] \\ &= (2\pi)^{-1/2} \exp[-t^2/2] \\ &= (2\pi)^{-1/2} \exp\{-[\chi^2(D, a_{\text{fit}} + te) - \chi^2(D, a_{\text{fit}})]/2\}. \end{aligned} \quad (166)$$

That is,  $\chi^2(D, a_{\text{fit}} + te(\sigma))$  gives both the probability of  $a$  lying at position  $t$  along the line  $a_{\text{fit}} + te(\sigma)$  and the probability of  $a$  lying in the plane  $\sigma(a) = \sigma_0(t)$  that intersects this line at position  $t$ .

It may be useful to note that one can find the direction of  $e(\sigma)$  simply. Up to its normalization,  $e(\sigma)$  is the vector from  $a_{\text{fit}}$  to the point on the surface  $\sigma(a) = \sigma_0(t)$  that minimizes  $\chi^2$  on this surface. The standard Lagrange multiplier method (Stump *et al.*, 2001) produces this vector.

After this introduction, we explore the role of a single experiment  $E$  in the fit. Consider  $\chi^2(D, a)$  for an  $a$  that varies along the line  $a = a_{\text{fit}} + te(\sigma)$ . The parameter  $t$  labels the distance along this line. We use one of three sets of data  $D$ : all of the data,  $D(\text{all})$  or all of the data except for the data from experiment  $E$ ,  $D(\text{no } E)$  or the data from experiment  $E$  alone  $D(E)$ . We are interested in how the function  $\chi^2(D, a_{\text{fit}} + te(\sigma))$  depends on  $t$  when we make these different choices for which dataset  $D$  we use in computing  $\chi^2$ .<sup>14</sup>

We ask two questions concerning the role of experiment  $E$  in determining  $t$ . The first question is: does experiment  $E$  exert a substantial pull on the value of  $t$ ? To answer this question, we ask what would happen if we omitted the data from experiment  $E$  from the evaluation of  $\chi^2$ . Then the minimum value of  $\chi^2(D(\text{no } E), a_{\text{fit}} + te(\sigma))$  occurs at a value  $t(\text{no } E)$  that typically differs from the value  $t(\text{all}) = 0$  that we get using all of the data. The  $1\sigma$  uncertainty in  $t(\text{all})$  is  $\Delta t(\text{all}) = 1$ . If the difference between  $t(\text{no } E)$  and  $t(\text{all}) = 0$  is smaller than this uncertainty, then we may conclude that experiment  $E$  does not exert a substantial pull on  $t$ . That is, for experiment  $E$  to substantially pull  $t$ , we need

<sup>14</sup>The dependence of  $\chi^2$  on the position of the parameters  $a$  along the line  $a_{\text{fit}} + te(\sigma)$  near  $a = a_{\text{fit}}$  is easily determined. The dependence on  $a$  over the entire plane  $\sigma(a) = \sigma_0$  requires knowledge of  $\chi^2$  in the entire parameter space. This is not as meaningful if we use just a small subset of the data  $D(E)$ .

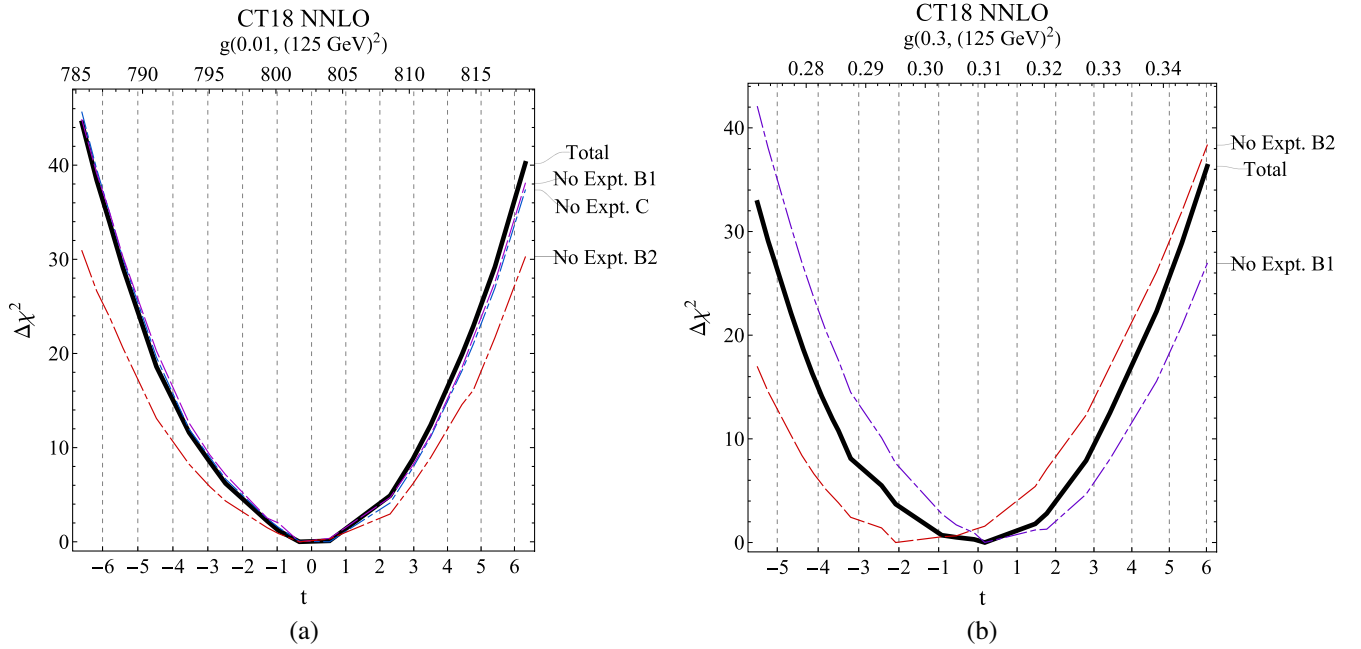


FIG. 10. (a) Dependence of  $\chi^2$  in a CT18 NNLO fit as a function of distance  $t$  in parameter space corresponding to changes in  $g(0.01, (125 \text{ GeV})^2)$ . The black solid curve shows the total  $\chi^2$ , while the remaining three curves show  $\chi^2$  as a function of  $t$  with particular experiments removed from the dataset. (b)  $\chi^2$  is as before, but along a line corresponding to changes in  $g(0.3, (125 \text{ GeV})^2)$ .

$$|t(\text{no } E)| > f, \quad (167)$$

where  $f$  is a parameter we could pick, perhaps  $f = 1$ .

This is illustrated in Fig. 10, which is based on the CT18 NNLO fit (Hou *et al.*, 2019). For this illustration, we choose a conservative value of  $f$ :  $f = 0.5$ .

In Fig. 10(a), we choose the gluon distribution at  $x = 0.01$  and  $\mu = 125 \text{ GeV}$  as our observable  $\sigma$ . The heavy black curve is the difference of  $\chi^2(D(\text{all}), a_{\text{fit}} + te(\sigma))$  and its minimum value as a function of the parameter  $t$ . The corresponding values of  $g(0.01, (125 \text{ GeV})^2)$  are shown along the top of the plot. We also show curves for the differences of  $\chi^2(D(\text{no } E), a_{\text{fit}} + te(\sigma))$  and their minimum values for three choices of datasets  $E$ , labeled as experiments B1, B2, and C. (The experiments are taken from an actual global fit.) We see that the minima of all of these curves lie in the range  $-0.5 < t < 0.5$ , indicating that none of these datasets exerts a substantial pull on  $t$  in the sense of Eq. (167) with  $f = 0.5$ .

In Fig. 10(b), we choose the gluon distribution at  $x = 0.3$  and  $\mu = 125 \text{ GeV}$  as our observable  $\sigma$ . Again, the heavy black curve is constructed from  $\chi^2(D(\text{all}), a_{\text{fit}} + te(\sigma))$ . We also show curves for  $\chi^2(D(\text{no } E), a_{\text{fit}} + te(\sigma))$  for two choices of datasets  $E$ : experiments B1 and B2. These correspond to two datasets obtained from the same experiment B for two different collision energies  $\sqrt{s}$ . We see that the minima of these curves lie outside the range  $-0.5 < t < 0.5$ , indicating that both of these datasets exert a substantial pull on  $t$  in the sense of Eq. (167) with  $f = 0.5$ .

Now suppose that experiment  $E$  does exert a substantial pull on  $t$ . Then we need to check whether the global fit solution  $t(\text{all})$  is consistent with what experiment  $E$  says. We define

$$\Delta\chi_E^2(t) = \chi^2(D(E), a_{\text{fit}} + te(\sigma)) - \chi^2(D(E), a_{\text{fit}}). \quad (168)$$

According to experiment  $E$  alone, the best-fit  $t(E)$  is obtained by minimizing  $\Delta\chi_E^2(t)$ . The  $1\sigma$  uncertainty range for  $t(E)$  is given by  $|\Delta\chi_E^2(t) - \Delta\chi_E^2(t(E))| = 1$ . Thus, the result  $t = 0$  from the full fit is consistent with the result from experiment  $E$  alone if

$$|\Delta\chi_E^2(0) - \Delta\chi_E^2(t(E))| < n^2, \quad (169)$$

where  $n$  is a parameter that we could pick, perhaps  $n = 2$  for consistency within a 95% confidence interval.

This is illustrated in Fig. 11, again based on the CT18 NNLO fit and the observable  $\sigma = g(0.3, (125 \text{ GeV})^2)$ . In Fig. 10(b), we see that two datasets B1 and B2 make at least a marginal difference in the overall fit. Now we plot  $\Delta\chi_{\text{total}}^2(t)$  for the overall fit as a heavy black line and also  $\Delta\chi_E^2(t)$  for  $E = \text{B1}$  and  $E = \text{B2}$ : the datasets obtained in the same experiment B that was repeated at two collider energies  $\sqrt{s}$ . We also exhibit the  $\Delta\chi_E^2(t)$  curves for three other datasets. We see that  $\Delta\chi_{\text{B1}}^2(t)$  is about 10 units higher at  $t = 0$  than it is at its minimum. Thus, the consistency condition (169) with  $n = 2$  is violated in experiment B1. On the other hand,  $\Delta\chi_{\text{B2}}^2(t)$  is only about 3 units higher at  $t = 0$  than it is at its minimum. Thus, the consistency condition (169) with  $n = 2$  is satisfied in experiment B2.

At this level of inconsistency for experiment B1, it is not credible that we are simply looking at statistical fluctuations. One simple but crude way to remove the inconsistency would be to increase the error estimates for the discrepant dataset(s). To illustrate how this might work, in Fig. 12 we refit the PDFs by assuming increased quoted errors for the dataset B1; namely, we refit after multiplying the quoted errors of B1 by a constant factor  $\sqrt{2}$ . That is, in Eq. (59) we multiply  $C_{ij}$  for  $i, j \in \text{B1}$  by a common factor  $1/2$ .

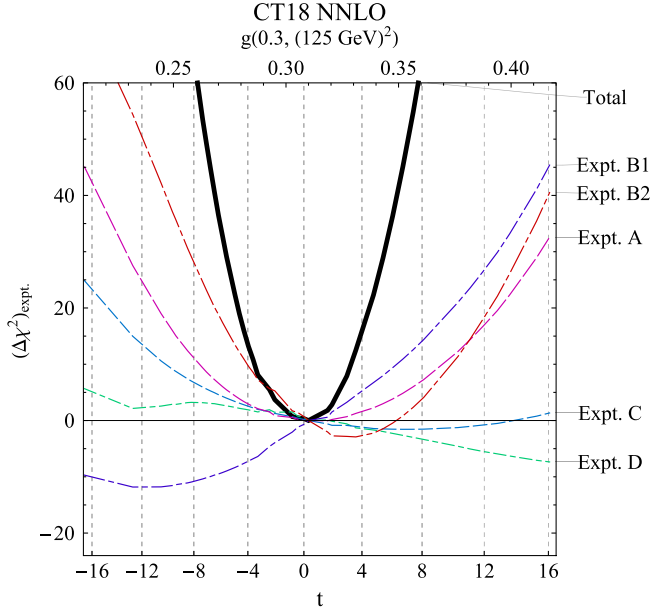


FIG. 11. Dependence of  $\chi^2$  in a CT18 NNLO fit as a function of distance  $t$  in parameter space corresponding to changes in  $g(0.3, (125 \text{ GeV})^2)$ . The black solid curve shows the total  $\chi^2$ , while the remaining curves show  $\chi^2$  as a function of  $t$  in certain experiments  $E$  alone.

Figure 12(a) shows differences of  $\chi^2$  from their minimum values as functions of  $t$  for the full dataset  $D(\text{all})$  and then for  $D(\text{no } E)$  with  $E = \text{B1}$  (with the rescaled errors) and B2, as in Fig. 10(b). In general, we expect that increasing the estimated errors from certain datasets would change the position of the best fit for the observable  $\sigma$  and increase the estimated error on the prediction for  $\sigma$ . In this case, for  $g(0.3, (125 \text{ GeV})^2)$ ,

neither the position of the minimum nor the estimated error changes significantly.

In Fig. 12(a), we see that the minima of the two curves occur well within the region  $-f < t < f$ , even for  $f = 0.5$ . Since the criterion (167) no longer indicates that these two datasets exert substantial pulls on  $t$ , we need not examine the criterion (169) for a discrepancy between a dataset and the overall fit.

If we do examine the criterion (169), we obtain Fig. 12(b), where we show  $\Delta\chi_E^2(t)$  for the full dataset  $D(\text{all})$  and then for  $D(E)$  with  $E = \text{B1}$  (rescaled errors) and B2. We see that  $\Delta\chi_E^2(t)$  is less than 4 units higher at  $t = 0$  than it is at its minimum for both B1 and B2. These represent less than  $2\sigma$  discrepancies, which are not nearly as alarming as the discrepancy for experiment B1 that we saw in Fig. 11.

In summary, this analysis gives us criteria for checking to see whether there is a problem associated with the data from experiment  $E$  in determining  $t$ . There is a problem if experiment  $E$  exerts a substantial pull on the value of  $t$  [Eq. (167)], and if the fit based only on experiment  $E$  is inconsistent with the global fit [Eq. (169)]. There is one set of criteria for each independent direction  $e(\sigma)$  corresponding to an observable  $\sigma$  and for each experiment  $E$ .

We have explored how one can make the results from a dataset  $E$  more consistent with the rest of the data by simply rescaling the errors for this dataset. This is a crude method. We do not recommend using it for finding the best fit. In Sec. IV.J, we explore a more subtle method.

A less precise alternative is to leave the disagreeing experiment(s) and best fit based on the experiment(s) unchanged but increase the PDF uncertainty to reflect the incompatibility in the experimental constraints. This possibility is discussed in Sec. IV.L.

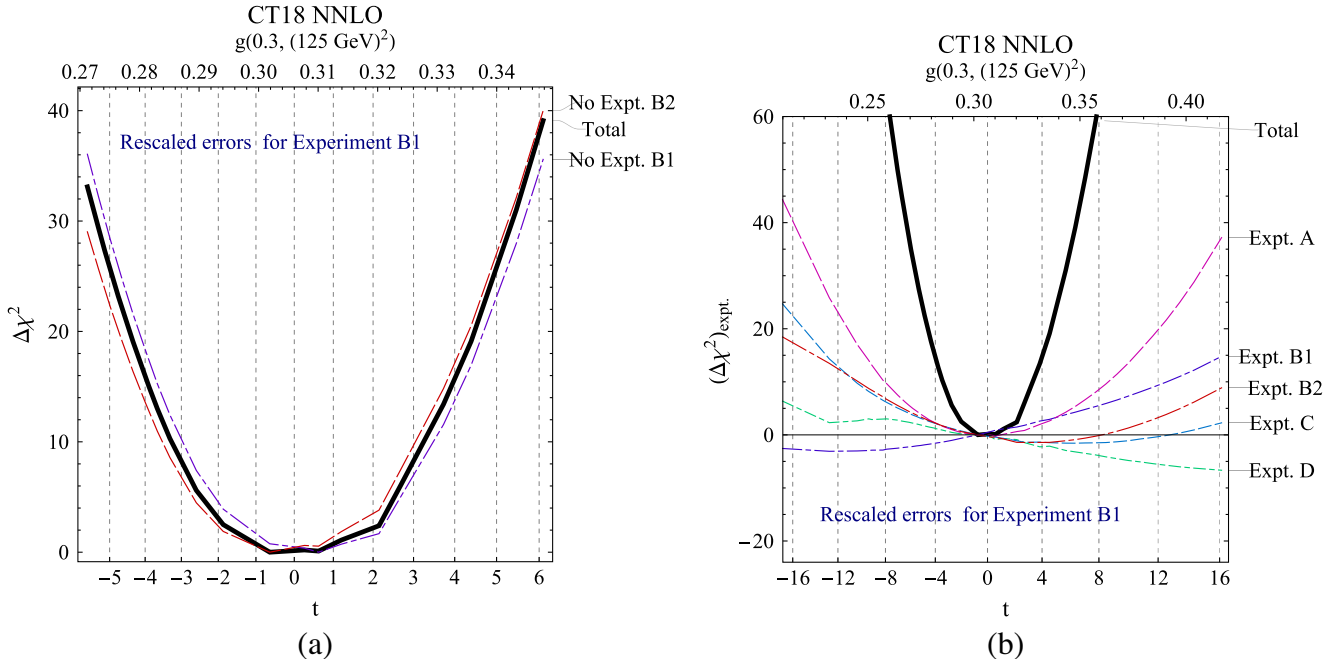


FIG. 12. Same as Figs. 10(b) and 11 if the estimated errors for the experiment B1 are multiplied by  $\sqrt{2}$ .

### J. A more conservative way to adjust the errors

Suppose that the experiment  $E$  exerts a substantial pull on the fit value of the observable  $\sigma$  according to the criterion (167), and  $\chi^2$  for dataset  $E$  along the line  $a = a_{\text{fit}} + te(\sigma)$  is not consistent according to the criterion (169) with its value at  $t = 0$ , the best-fit value of  $t$  according to the fit to all data. Then it may be helpful to increase the experimental errors for experiment  $E$ . We stated that simply increasing the total error estimate for experiment  $E$  by dividing  $C_{ij}$  for  $i, j \in E$  by a common factor is a crude strategy. A more focused strategy would be to add a systematic error of the form of Eq. (52) such that it resolves the inconsistency of the constraints on our observable  $\sigma$  but does not affect the fit in any other way. We parametrize this systematic error as

$$\sigma_k \beta_{k,\text{new}} = \xi \bar{\beta}_k, \quad (170)$$

where  $\xi$  is a constant that we can adjust and

$$\bar{\beta}_k = \theta(k \in E) \frac{T_{k\beta} e(\sigma)_\beta}{[e(\sigma)_\alpha H(E)_{\alpha\beta} e(\sigma)_\beta]^{1/2}}. \quad (171)$$

In the normalization factor,  $H(E)_{\alpha\beta}$  is the Hessian matrix, as in Eq. (86), but includes only the data from dataset  $E$

$$H(E)_{\alpha\beta} = \sum_{i,j \in E} T_{i\alpha} T_{j\beta} C_{ij}. \quad (172)$$

With this factor,  $\bar{\beta}_k$  is independent of the normalization of the vector  $e(\sigma)$ . The vector  $e(\sigma)$  does have a definite normalization (161), but in Eq. (170) only the direction of  $e(\sigma)$  matters. A simple relation for the normalization factor  $e(\sigma)_\alpha H(E)_{\alpha\beta} e(\sigma)_\beta$  is given in Eq. (188).

We examine whether adding a new systematic error of this form can repair the incompatibility between dataset  $E$  and the rest of the data while not significantly affecting the remaining fit.

From Eq. (56) with the systematic error in Fig. (170) added, the covariance matrix becomes

$$C(\xi^2)_{ij}^{-1} = \sigma_i \sigma_j \delta_{ij} + \sum_j \sigma_i \beta_{i,j} \sigma_j \beta_{j,j} + \xi^2 \bar{\beta}_i \bar{\beta}_j. \quad (173)$$

Thus,

$$\frac{d}{d\xi^2} C(\xi^2)_{ij}^{-1} = \bar{\beta}_i \bar{\beta}_j. \quad (174)$$

Using  $dC/d\xi^2 = -C[dC^{-1}/d\xi^2]C$ , this is

$$\frac{d}{d\xi^2} C(\xi^2)_{ij} = -\sum_k C(\xi^2)_{ik} \bar{\beta}_k \sum_l C(\xi^2)_{jl} \bar{\beta}_l. \quad (175)$$

It is straightforward to solve this differential equation to obtain

$$C(\xi^2)_{ij} = C_{ij} - \frac{\xi^2 \sum_k C_{ik} \bar{\beta}_k \sum_l C_{jl} \bar{\beta}_l}{1 + \xi^2 \sum_{kl} \bar{\beta}_k C_{kl} \bar{\beta}_l}. \quad (176)$$

Here  $C_{ij} = C(0)_{ij}$  is the covariance matrix without the added systematic error. With the definition in Eq. (171) of  $\bar{\beta}_k$ , we have

$$\sum_{kl} \bar{\beta}_k C_{kl} \bar{\beta}_l = 1, \quad (177)$$

so

$$C(\xi^2)_{ij} = C_{ij} - \frac{\xi^2}{1 + \xi^2} \sum_k C_{ik} \bar{\beta}_k \sum_l C_{jl} \bar{\beta}_l. \quad (178)$$

What is the effect on  $\chi^2$  for experiment  $E$  of adding this systematic error? Consider  $\chi^2$  for experiment  $E$ , for parameters

$$a_\alpha = a_\alpha^{\text{fit}} + te_\alpha, \quad (179)$$

where the  $a_\alpha^{\text{fit}}$  are the parameters from the global fit before adding the extra systematic error and the vector  $e$  could be the special vector  $e(\sigma)$  for observable  $\sigma$ , but it could also be any other vector in the many-dimensional space of parameters. We have, from Eqs. (59) and (77),

$$\begin{aligned} \chi^2(D(E), a_{\text{fit}} + te, \xi) \\ = \sum_{i,j \in E} [D_i - T_i(a_{\text{fit}}) - tT_{i\alpha} e_\alpha][D_j - T_j(a_{\text{fit}}) - tT_{j\beta} e_\beta] C_{ij}(\xi). \end{aligned} \quad (180)$$

We can write this as

$$\begin{aligned} \chi^2(D(E), a_{\text{fit}} + te, \xi) \\ = \chi^2(D(E), a_{\text{fit}}, \xi) - 2tB(E, \xi)_\beta e_\beta + t^2 e_\alpha H(E, \xi)_{\alpha\beta} e_\beta, \end{aligned} \quad (181)$$

where

$$B(E, \xi)_\beta = \sum_{i,j \in E} [D_i - T_i(a_{\text{fit}})] C_{ij}(\xi) T_{j\beta} \quad (182)$$

gives the contribution linear in  $t$  and

$$H(E, \xi)_{\alpha\beta} = \sum_{i,j \in E} T_{i\alpha} T_{j\beta} C_{ij}(\xi) \quad (183)$$

is the Hessian matrix including the added systematic error, but just for the data from experiment  $E$ . Following the notation from Eq. (172), we define

$$H(E)_{\alpha\beta} \equiv H(E, 0)_{\alpha\beta}, \quad B(E)_{\alpha\beta} \equiv B(E, 0)_{\alpha\beta}. \quad (184)$$

Without the added systematic error, we have

$$\begin{aligned} \chi^2(D(E), a_{\text{fit}} + te, 0) \\ = \chi^2(D(E), a_{\text{fit}}, 0) - 2tB(E)_\gamma e_\gamma + t^2 e_\alpha H(E)_{\alpha\beta} e_\beta. \end{aligned} \quad (185)$$

When we add the new systematic error, the result changes to

$$\begin{aligned}
 & \chi^2(D(E), a_{\text{fit}} + te, \xi) \\
 &= \chi^2(D(E), a_{\text{fit}} + te, 0) - \frac{\xi^2}{1 + \xi^2} \frac{[B(E)_\beta e(\sigma)_\beta]^2}{e(\sigma)_\alpha H(E)_{\alpha\beta} e(\sigma)_\beta} \\
 &+ 2t \frac{\xi^2}{1 + \xi^2} \frac{[B(E)_\gamma e(\sigma)_\gamma][e(\sigma)_\alpha H(E)_{\alpha\beta} e_\beta]}{e(\sigma)_\alpha H(E)_{\alpha\beta} e(\sigma)_\beta} \\
 &- t^2 \frac{\xi^2}{1 + \xi^2} \frac{[e(\sigma)_\alpha H(E)_{\alpha\beta} e_\beta]^2}{e(\sigma)_\alpha H(E)_{\alpha\beta} e(\sigma)_\beta}. \quad (186)
 \end{aligned}$$

Here we use Eq. (178) to separate from  $\chi^2(D(E), a, 0)$  the extra terms resulting from the new systematic error (proportional to  $\xi^2$ ).

We now show that the new systematic error reduces an apparent tension between dataset  $E$  and the rest of the data along the line associated with  $\sigma$  (our observable of interest). It does not modify constraints in the other directions. In the last two terms of Eq. (186), the numerators contain  $[e(\sigma)_\alpha H(E)_{\alpha\beta} e_\beta]$ , the inner product between the unit vector  $e(\sigma)$  defining the direction associated with  $\sigma$ , and another (possibly orthogonal) unit vector  $e$  that defines the line  $a_{\text{fit}} + te$  along which we choose to scan  $\chi^2(D(E), a_{\text{fit}} + te, \xi)$ . When we scan along the direction  $e = e(\sigma)$ , the  $\chi^2(D(E), a_{\text{fit}} + te, \xi)$  function changes as

$$\begin{aligned}
 & \chi^2(D(E), a_{\text{fit}} + te(\sigma), \xi) \\
 &= \chi^2(D(E), a_{\text{fit}}, 0) - \frac{\xi^2}{1 + \xi^2} \frac{[B(E)_\beta e(\sigma)_\beta]^2}{e(\sigma)_\alpha H(E)_{\alpha\beta} e(\sigma)_\beta} \\
 &- 2t \frac{1}{1 + \xi^2} [B(E)_\gamma e(\sigma)_\gamma] + t^2 \frac{1}{1 + \xi^2} [e(\sigma)_\alpha H(E)_{\alpha\beta} e(\sigma)_\beta]. \quad (187)
 \end{aligned}$$

If we turn off the systematic error (set  $\xi = 0$ ) for a moment, from Eq. (187) we can numerically find the factor  $e(\sigma)_\alpha H(E)_{\alpha\beta} e(\sigma)_\beta$  that appears here and in the definition in Eq. (171):

$$e(\sigma)_\alpha H(E)_{\alpha\beta} e(\sigma)_\beta = \frac{1}{2} \frac{d^2}{dt^2} \chi^2(D(E), a_{\text{fit}} + te(\sigma), 0). \quad (188)$$

If we turn the systematic error back on by choosing  $\xi \neq 0$ , the second term makes  $\chi^2$  at  $t = 0$  smaller as  $\xi$  increases without affecting the shape of  $\chi^2$  as a function of  $t$ . In the remaining terms, the coefficients of  $t$  and  $t^2$  are reduced by the same factor as  $\xi$  increases. The shape of  $\chi^2$  versus  $t$  simply becomes shallower as desired, reducing the tension between dataset  $E$  and the rest of the data along the line  $a = a_{\text{fit}} + te(\sigma)$ .

On the other hand, in the directions that are orthogonal to the direction of variation of  $\sigma$ , the unit vectors  $e$  obey

$$e(\sigma)_\alpha H(E)_{\alpha\beta} e_\beta = 0. \quad (189)$$

If, for example, the parameter space is 26 dimensional, then there is a 25-dimensional vector space in which  $e$  could lie and satisfy this condition. In this case, Eq. (186) gives

$$\begin{aligned}
 & \chi^2(D(E), a_{\text{fit}} + te, \xi) \\
 &= \chi^2(D(E), a_{\text{fit}} + te, 0) - \frac{\xi^2}{1 + \xi^2} \frac{[B(E)_\beta e(\sigma)_\beta]^2}{e(\sigma)_\alpha H(E)_{\alpha\beta} e(\sigma)_\beta}. \quad (190)
 \end{aligned}$$

That is,  $\chi^2$  for dataset  $E$  at  $t = 0$  is smaller when  $\xi > 0$ , but the shape of  $\chi^2$  as a function of  $t$  is not changed at all. Thus, the new systematic error is a conservative choice in that it alleviates the incompatibility problem while having a minimal effect on the rest of the fit.

We see how this prescription resolves the tension between experiment B1 and other experiments that we observe in Fig. 11. There we examine the parameter fit along a direction  $e(\sigma)$  corresponding to  $\sigma = g(0.3, (125 \text{ GeV})^2)$ . We see that the curve of

$$\Delta\chi^2_E(t) = \chi^2(D(E), a_{\text{fit}} + te(\sigma)) - \chi^2(D(E), a_{\text{fit}}) \quad (191)$$

for a certain experiment  $E = \text{B1}$  is not consistent with the choice of  $t = 0$  that minimizes the total  $\chi^2$ . To alleviate this problem, we add the new systematic error in Eq. (170) for experiment B1 with  $\xi$  defined by  $1/(1 + \xi^2) \approx 4/10$ . With this error added, the new  $\Delta\chi^2_{\text{B1}}(t)$  for B1 now satisfies

$$\Delta\chi^2_{\text{B1,new}}(t) = \frac{\Delta\chi^2_{\text{B1,old}}(t)}{1 + \xi^2} \approx \frac{4}{10} \Delta\chi^2_{\text{B1,old}}(t). \quad (192)$$

The new  $\Delta\chi^2_{\text{B1}}(t)$  curve becomes flat enough that consistency with the rest of the data along the line  $a_{\text{fit}} + te(\sigma)$  is no longer a problem.

We can now perform the global fit again with the modified systematic error for experiment B1. Then the best-fit parameters  $a_{\text{fit}}$  change. The direction vector  $e(\sigma)$  corresponding to the observable  $g(0.3, (125 \text{ GeV})^2)$  also changes. The new fit gives us a new plot analogous to Fig. 11 in which all of the  $\chi^2$  curves have changed. The result is shown in Fig. 13. In the new fit,  $\chi^2$  for experiment B1 is flat, indicating that experiment B1 is now not significantly affecting the determination of  $g(0.3, (125 \text{ GeV})^2)$ . The best-fit value of  $g(0.3, (125 \text{ GeV})^2)$  changes from 0.309 to 0.312. The estimated error on the fit value of  $g(0.3, (125 \text{ GeV})^2)$ , determined by the second derivative of the total  $\chi^2$  curve with respect to  $g(0.3, (125 \text{ GeV})^2)$ , is about 2% larger.

### K. Summary of measures of goodness of fit

The statistical analysis that we present in Sec. III relies on several assumptions. One assumes that experimental systematic errors are adequately represented by a Gaussian distribution described by a covariance matrix given by the experiments. Typically, one also assumes that theory errors are small enough to be neglected. Additionally, one assumes that the theoretical prediction is, to a good approximation, a linear function of the PDF parameters in the region  $a_{\text{fit}} \pm \delta a$  near the best-fit parameter choice. Some of these assumptions could be wrong.

We have argued that we should not simply trust these assumptions but should instead test them by using a strong set of goodness-of-fit criteria. Taken together, these tests are

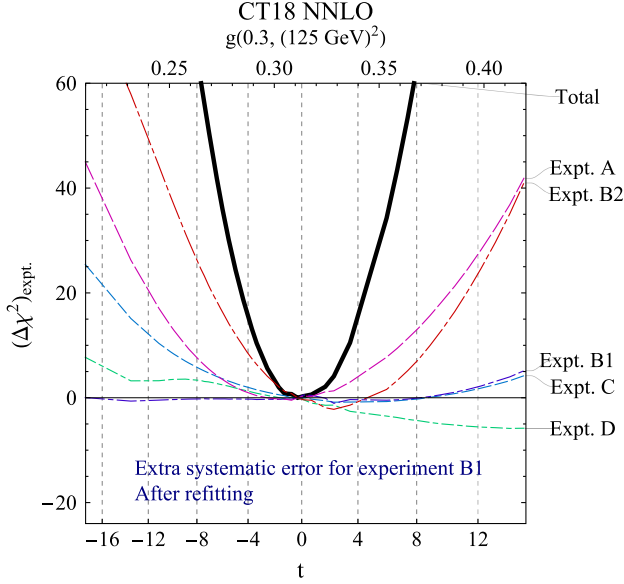


FIG. 13.  $\chi^2$  curves for individual experiments as in Fig. 11 but with an extra systematic error added for experiment B1 according to Eqs. (170) and (171). The fit is repeated with the new systematic error for experiment B1. The new fit gives a new best-fit choice  $a_{\text{fit}}$ . Now the observable  $g(0.3, (125 \text{ GeV})^2)$  defines a new direction  $e(\sigma)$  in parameter space. This plot uses the new  $a_{\text{fit}}$ ,  $e(\sigma)$ , and the total  $\chi^2$  values after the fit.

much more stringent than that obtained by simply noting the global  $\chi^2$  value. If the fit passes all of these tests, we can have some confidence in the results and the errors on the results. If the fit does not pass all tests, then remediation is needed. We do not offer a fixed prescription, but we have pointed out some possibilities.

The choice that has usually been made is to leave the fit as it is but to use larger error estimates on the final PDFs than those found with the parameter-fitting criterion that requires  $\Delta\chi^2 = 4$  at the 95% probability level. Estimation of trustworthy PDF errors in such an imperfect situation can be difficult and sometimes controversial. Before a new generation of PDFs is published, it may undergo many months of multifaceted PDF testing to establish the realistic estimates for PDF uncertainties. The increase over the nominal PDF errors that results from this procedure is often referred to as applying tolerance to the PDF uncertainty.

#### L. Global and dynamic tolerance

Tolerance is relevant at the stage of determination of PDF uncertainties after the best-fit PDF has been found. In its simplest realization, tolerance defines an allowed range for the variation

$$a = a_{\text{fit}} + te \quad (193)$$

of the parameters  $a$  along a direction  $e$  and at a probability level  $v$ . Often,  $e$  is one of the eigenvectors of  $H$ ,  $H_{\alpha\beta}e_\beta = he_\alpha$ . However, any direction  $e$  is a possible choice. We define the normalization of  $e$  using the Hessian matrix, as in Eq. (117):  $e_\alpha H_{\alpha\beta} e_\beta = 1$ . Then the dependence of  $\chi^2$  on  $t$  is given by Eq. (119) as

$$\chi^2(D, a_{\text{fit}} + te) = \chi^2(D, a_{\text{fit}}) + t^2. \quad (194)$$

According to Eq. (111), if the experiments that determine  $a_{\text{fit}}$  are repeated many times, then the component  $t$  of  $a_{\text{fit}} - \bar{a}$  in the direction  $e$  is distributed according to  $\mathcal{N}(0, 1)$ . Thus, if we pick a probability  $v$  and ask that

$$-t_{\text{lim}}(v) < t < t_{\text{lim}}(v) \quad (195)$$

with probability  $v$ , the limiting value  $t_{\text{lim}}(v)$  is determined by

$$\int_{-t_{\text{lim}}(v)}^{t_{\text{lim}}(v)} d\bar{t} p(\bar{t}) = v, \quad (196)$$

where  $p(\bar{t})$  is the Gaussian distribution  $\mathcal{N}(0, 1)$ . Then  $t_{\text{lim}}(0) \approx 0$ ,  $t_{\text{lim}}(0.68) \approx 1$ ,  $t_{\text{lim}}(0.8) \approx 1.3$ , and  $t_{\text{lim}}(0.95) \approx 2$ .

When we believe that this procedure misestimates the true uncertainty on  $t$ , we could try to find a better probability distribution  $p(\bar{t})$  to use in Eq. (196). For example, we could use  $\mathcal{N}(0, T)$  with  $T^2 > 1$  as our  $p(\bar{t})$  so that

$$p(\bar{t}) = \frac{1}{\sqrt{2\pi}} \exp[-\bar{t}^2/(2T^2)]. \quad (197)$$

Here we use the same value  $T$  for every direction vector  $e$  (Pumplin *et al.*, 2001). The value  $T^2$  in this case is referred to as the global tolerance. With  $v = 0.68$ , the allowed variation of PDF parameters is constrained to satisfy  $-T < t < T$  along any vector direction with this prescription. With  $v = 0.95$ , the allowed variation of PDF parameters is constrained to satisfy  $-4T < t < 4T$ .

The dynamic tolerance introduced by the Martin-Stirling-Thorne-Watt group (Martin *et al.*, 2009) is determined by a similar consideration by constructing  $p(\bar{t})$  from  $\chi^2$  distributions for individual experiments  $E$ . If  $P_N(\chi^2)$  is the  $\chi^2$  distribution with  $N$  degrees of freedom, we can define  $\xi(N, v)$  by

$$\int_{-\infty}^{\xi(N, v)} d\chi^2 P_N(\chi^2) = v \quad (198)$$

so that  $\chi^2 < \xi(N, v)$  with probability  $v$ . Note that we choose a one-sided limit here. Since according to Eq. (157)  $S = \sqrt{2\chi^2} - \sqrt{2N-1}$  closely obeys the  $\mathcal{N}(0, 1)$  distribution, we can relate  $\xi(N, v)$  to  $t_{\text{lim}}(v)$  to the following good approximation:

$$\xi(N, v) \approx \frac{1}{2}[\sqrt{2N-1} + t_{\text{lim}}(2v-1)]^2. \quad (199)$$

With this information, the ‘‘dynamic tolerance’’ prescription of Martin *et al.* (2009) assigns an allowed interval

$$T_{\text{min}} < t < T_{\text{max}} \quad (200)$$

for some eigenvector direction in the following way.

We define  $\chi^2(D(E), a)$  as the part of  $\chi^2$  coming only from the data in dataset  $E$ , as in Sec. IV.I. We use  $\chi^2(D(E), a_{\text{fit}} + te)$  to define limits  $T_{\text{min}}(E)$  and  $T_{\text{max}}(E)$  arising from dataset  $E$ , as explained later. Then we set



$$\begin{aligned} T_{\min} &= \max_E T_{\min}(E), \\ T_{\max} &= \min_E T_{\max}(E). \end{aligned} \quad (201)$$

For every  $E$ , Martin *et al.* (2009) defined the range  $T_{\min}(E) < t < T_{\max}(E)$  by the criterion

$$\frac{\chi^2(D(E), a_{\text{fit}} + te)}{\chi^2(D(E), a_{\text{fit}})} < \frac{\xi(N_E, v)}{\xi(N_E, 1/2)}, \quad (202)$$

where  $N_E$  is the number of data in dataset  $E$ . To understand the result of applying this criterion, it is helpful to use some approximations.

First, using Eq. (199) gives

$$\frac{\chi^2(D(E), a_{\text{fit}} + te)}{\chi^2(D(E), a_{\text{fit}})} < \left[ 1 + \frac{t_{\text{lim}}(2v - 1)}{\sqrt{2N_E - 1}} \right]^2. \quad (203)$$

Noting that  $t_{\text{lim}}(2v - 1)$  is of the order of 1, we see that for  $N_E \gg 1$  this is

$$\frac{\chi^2(D(E), a_{\text{fit}} + te)}{\chi^2(D(E), a_{\text{fit}})} < 1 + \frac{\sqrt{2}t_{\text{lim}}(2v - 1)}{\sqrt{N_E}}. \quad (204)$$

Now we examine the left-hand side of Eq. (203).  $\chi^2(D(E), a_{\text{fit}} + te)$  is a quadratic function of  $t$  given as

$$\chi^2(D(E), a_{\text{fit}} + te) = \chi^2(D(E), a_{\text{fit}}) + A_1(E)t + A_2(E)t^2, \quad (205)$$

with the coefficients  $A_1(E)$  and  $A_2(E)$  given in Eq. (185). We note that  $0 < A_2(E) < 1$  because

$$\begin{aligned} A_2(E) &= \sum_{i,j \in E} e_\alpha T_{ia} C_{ij} T_{j\beta} e_\beta < \sum_{i,j} e_\alpha T_{ia} C_{ij} T_{j\beta} e_\beta \\ &= e_\alpha H_{\alpha\beta} e_\beta = 1. \end{aligned} \quad (206)$$

The coefficient  $A_1(E)$  could have either sign and could be large.

Inserting Eq. (205) into Eq. (204) gives

$$1 + \frac{A_1(E)t + A_2(E)t^2}{\chi^2(D(E), a_{\text{fit}})} < 1 + \frac{\sqrt{2}t_{\text{lim}}(2v - 1)}{\sqrt{N_E}}. \quad (207)$$

The large terms, 1 here, cancel exactly. This gives

$$A_1(E)t + A_2(E)t^2 < \frac{\chi^2(D(E), a_{\text{fit}})}{\sqrt{N_E}} \sqrt{2}t_{\text{lim}}(2v - 1). \quad (208)$$

To understand this, we can estimate  $\chi^2(D(E), a_{\text{fit}})$  by its expectation value, which according to Eq. (137) is approximately  $N_E$ . This gives

$$A_1(E)t + A_2(E)t^2 < \sqrt{2N_E}t_{\text{lim}}(2v - 1). \quad (209)$$

This gives upper and lower limits on  $t$  for each experiment  $E$ . If, for example, we take  $v = 0.9$ , then  $t_{\text{lim}}(2v - 1) = t_{\text{lim}}(0.8) \approx 1.3$ . For simplicity, consider the case that  $A_1(E)$  is

small. If  $A_2(E)$  is also small, then this inequality restricts  $t$  only weakly. That is,  $|T_{\min}(E)|$  and  $|T_{\max}(E)|$  are large. We always have  $A_2(E) < 1$ . If  $A_2(E)$  is close to 1, then this inequality can provide a significant restriction on  $t$ . However, the restriction is only significant if  $\sqrt{2N_E}$  is not too large. For datasets with large  $N_E$ , the restriction is always weak. Thus, the most restrictive values of  $|T_{\min}(E)|$  and  $|T_{\max}(E)|$ , and thus the overall values of  $|T_{\min}|$  and  $|T_{\max}|$ , are likely to come from datasets in which  $\sqrt{2N_E}$  is not too large and  $A_2(E)$  is not too small. For most experiments, the values of  $|T_{\min}(E)|$  or  $|T_{\max}(E)|$  tend to be substantially greater than 1.

We do not attempt to justify the definition (202) of the range for  $t$  or its approximate version (209). We note, however, that the factor  $\sqrt{2N_E}$  in Eq. (209) is familiar: it is the standard deviation for the distribution of  $\chi^2$ , as in Eq. (138) for  $N_D \gg N_p$ .

We have described one approach to defining a tolerance factor. There are several other approaches. It is beyond our scope to explore them in detail.

## V. PARTON DISTRIBUTIONS FOR HEAVY IONS

The concepts discussed in this review can be applied to nuclear parton distribution functions (nPDFs), nonperturbative QCD functions that are increasingly employed to model the structure of heavy nuclei in high-energy scattering. The concept of collinear QCD factorization that is central to describing scattering of free hadrons is also relevant for the growing number of measurements in collisions of heavy nuclei. The experimental data available for constraining the nPDFs is still limited in their span over  $x$  and  $\mu^2$ . They are anticipated to grow quickly as the Large Hadron Collider and, especially, the envisioned Electron-Ion Collider produce new results. We now review the key features of the nPDFs and refer the interested reader to the original publications by nuclear PDF analysis groups EPPS (Eskola *et al.*, 2017), nCTEQ (Kovařík *et al.*, 2016), DSSZ (de Florian *et al.*, 2012), HKN (Hirai, Kumano, and Nagai, 2007), KA (Khanpour and Atashbar Tehrani, 2016), and NNPDF (Abdul Khalek, Ethier, and Rojo, 2019) for more details.

### A. Universality of nuclear PDFs

As in the standard case of protons, the structure functions and cross sections in collisions involving one or more nuclei are related to nPDFs via perturbative QCD factorization (Collins, Soper, and Serman, 1983, 1985, 1988; Bodwin, 1985). Even though inclusive collinear factorization has been proven in only a few cases (e.g., deeply inelastic scattering on hadrons and lepton pair production in hadron collisions), it has formed the basis for the analysis of proton PDFs. Equations (35) and (36) express QCD factorization for free-nucleon QCD observables. We now show how to extend these factorization formulas to high-energy scattering of heavy nuclei. We denote a nucleus by  $N$  and its atomic number by  $A$ .

We assume that, at sufficiently large  $\sqrt{s}$ , the QCD observables of interest are dominated by independent parton scatterings, in which only one parton per initial-state nucleus contributes to the hard scattering. The interactions between

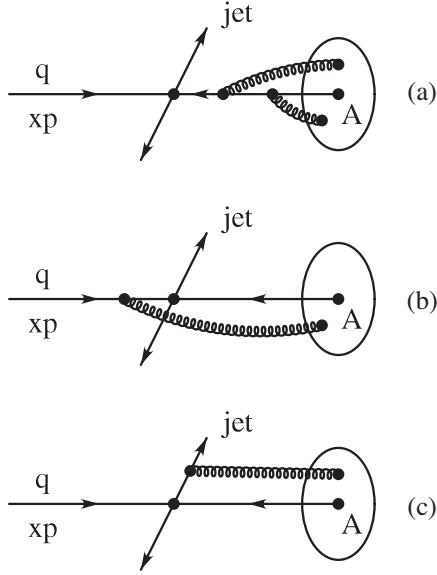


FIG. 14. Interactions of the nucleus with the initial- and final-state partons. From Qiu, 2003.

a parton and a nucleus can be classified in three categories (Qiu, 2003; Accardi *et al.*, 2004), as shown in Fig. 14. Figure 14(c) illustrates jet quenching, the changes in a jet as it moves through a large nucleus. Figure 14(b) illustrates two partons from the nucleus participating in the hard interaction. These effects contribute to the power-suppressed corrections to the factorization formulas (35) and (36) for nuclei, but they can be important if the nucleus is large enough and the scattering is not too hard.

The additional interactions between the nucleus and the initial-state parton within the same nucleus [shown in Fig. 14(a)] change the parton distributions of the nucleus and do not affect the hard scattering, thereby leaving intact the form of factorization given by Eqs. (35) and (36). The only change in the prescription replaces the free-proton PDFs with the nPDFs, which accounts for the additional initial-state effects and can be defined as

$$f_{i/N}(\xi_N, \mu^2) = \frac{1}{4\pi} \int dy^- e^{-i\xi_N P_N^+ y^-} \times \langle N | \bar{\psi}_i(0, y^-, \mathbf{0}) \gamma^+ W(y^-, 0) \psi_i(0) | N \rangle. \quad (210)$$

This definition is analogous to the one in Eq. (19), but the proton matrix element of the number density operator is replaced by the nuclear one. This nPDF is defined with respect to the entire nucleus with + momentum  $P_N^+$ . Accordingly, the parton described by this parton distribution function carries the + momentum  $p^+ = \xi_N P_N^+$ . The momentum fraction  $\xi_N$  is defined as

$$\xi_N = \frac{p^+}{P_N^+}, \quad \text{with } 0 \leq \xi_N < 1. \quad (211)$$

The modified factorization prescription for a cross section in collisions of nuclei  $N_1$  and  $N_2$  can be written using the nPDFs (210) as

$$\sigma[F] = \sum_{a,b} \int_0^1 d\xi_a \int_0^1 d\xi_b f_{a/N_1}(\xi_a, \mu_F^2) f_{b/N_2}(\xi_b, \mu_F^2) \times \hat{\sigma}_{a,b,\xi_a,\xi_b,\mu_F^2}[F] + \mathcal{O}(M/Q). \quad (212)$$

Even in collisions of nuclei, hard processes such as the production of muon pairs or sufficiently high- $p_T$  jets are dominated by the leading-power contributions in Eq. (212). Thus, these processes can be well described using  $f_{i/N}(\xi_N, \mu^2)$ , where the dependence on the scale is still governed by DGLAP equations. However, the environment of the nuclear collisions is much different from the free-nucleon collisions. For example, at small values of  $\xi_N$  the parton momenta, as viewed in the nucleus rest frame, are small, so the parton wave functions spread over the entire nucleus and beyond. Then ‘‘saturation’’ (Golec-Biernat and Wusthoff, 1999; Mueller, 1999; Bartels, Golec-Biernat, and Kowalski, 2002; Hautmann and Soper, 2007) or ‘‘shadowing’’ (Armesto, 2006) can substantially modify the nPDFs. The nPDFs can incorporate these and other initial-state nuclear effects, such as the ‘‘EMC’’ effect (named for the European Muon Collaboration) (Aubert *et al.*, 1983; Geesaman, Saito, and Thomas, 1995; Malace *et al.*, 2014), and still be universal.

On the other hand, jets produced in a partonic scattering can be altered by their passage through nuclear matter, as in Fig. 14(c), unless the jet’s transverse momentum is exceedingly large. This ‘‘jet quenching’’ can affect jet cross sections beyond what is predicted by Eq. (212) (Aad *et al.*, 2010; Chatrchyan *et al.*, 2011; Adam *et al.*, 2015). Jet-quenched contributions do not factorize in the same way.

The nPDFs depend on the number of protons  $Z$  and number of neutrons  $A - Z$ . Highly nontrivial  $A$  dependence arises from strong interactions of partons inside the nucleus. There is also trivial  $A$  dependence that is present even if the nucleons are free. Consider a simple model in which a nucleus with + momentum  $P_N^+$  is just a collection of comoving independent protons and neutrons in which each nucleon carries the same fraction  $\xi_{p,n} = 1/A$  of the total momentum  $P_N^+$ .

In this model, one could write the nPDF of the entire nucleus as

$$f_{a/N}(\xi_N, \mu^2) d\xi_N = [Z f_{a/p}(\xi_A, \mu^2) + (A - Z) f_{a/n}(\xi_A, \mu^2)] d\xi_A, \quad (213)$$

where  $f_{a/p}$  and  $f_{a/n}$  are the parton distributions in the free proton and neutron, and  $\xi_A$  is the momentum fraction of the + momentum of the parton with respect to the + momentum of the nucleon. The + momentum  $p^+$  of parton  $a$  is

$$p^+ = \xi_N P_N^+ = \xi_A \xi_{p,n} P_N^+ \quad (214)$$

so that the momentum fractions  $\xi_A$  and  $\xi_N$  are related via

$$\xi_A = \frac{\xi_N}{\xi_{p,n}}, \quad \text{or } \xi_A = A \xi_N \quad \text{if } \xi_{p,n} = \frac{1}{A}. \quad (215)$$

We can use this relation to rewrite Eq. (213) in terms of momentum fraction  $\xi_N$  as

$$f_{a/N}(\xi_N, \mu^2) d\xi_N = [Z f_{a/p}(\xi_N, \mu^2) + (A-Z) f_{a/n}(\xi_N, \mu^2)] d(\xi_N). \quad (216)$$

In this model,  $\xi_N$  is constrained to be in the range  $0 \leq \xi_N \leq 1/A$ , since free-nucleon PDFs vanish for  $\xi_A > 1$ , and each nucleon carries exactly the fraction  $1/A$  of the + momentum of the nucleus.

In reality, one nucleon can carry any fraction of the nucleus' + momentum since the nucleons participate in Fermi motion relative to each other (Bodek and Ritchie, 1981; Saito and Uchiyama, 1985). We still find it helpful to use the momentum fraction  $\xi_A \equiv A\xi_N$ . We now define it as the fraction of the average + momentum  $P_N^+ / A$  of a bound nucleon. The variable  $\xi_A$  now takes values in the interval  $0 \leq \xi_A \leq A$ , with contributions at  $1 < \xi_A < A$  arising from in-nucleus motion.

The PDFs of bound nucleons in the nucleus do not coincide with the free-nucleon PDFs. However, if nuclear modifications are moderate, we can start from Eq. (213) to get a reasonable ansatz for the parametrizations of nuclear PDFs.

We define a nuclear PDF of an average nucleon in a nucleus with atomic number  $A$ , denoted by  $f_a^A(\xi_A, \mu^2)$ . This nPDF has the form

$$f_a^A(\xi_A, \mu^2) = \frac{Z}{A} f_{a/p}^A(\xi_A, \mu^2) + \frac{(A-Z)}{A} f_{a/n}^A(\xi_A, \mu^2). \quad (217)$$

In Eq. (217),  $f_{a/p}^A(\xi_A, \mu^2)$  and  $f_{a/n}^A(\xi_A, \mu^2)$  are the PDFs in the bound proton and bound neutron. They are different from the free-nucleon PDFs  $f_{a/p,n}(\xi_A, \mu^2)$ . They depend on the previously defined momentum fraction  $\xi_A$ . We can relate the two types of nPDFs that we just discussed as

$$f_{a/N}(\xi_N, \mu^2) d\xi_N = A f_a^A(\xi_A, \mu^2) d\xi_A. \quad (218)$$

Both the nPDFs  $f_{a/N}(\xi_N, \mu^2)$  in the nucleus and the nPDFs  $f_a^A(\xi_A, \mu^2)$  for an average nucleon are acceptable for use in QCD calculations. But “trivial”  $A$  dependence makes it difficult to compare the nPDFs of the first kind  $f_{a/N}(\xi_N, \mu^2)$  for two different nuclei.

For example, consider the prominent feature of proton PDFs: the peaks of the up- and down-quark distributions at  $\xi \approx 1/3$ . Similar peaks are found in the respective nPDFs  $f_{u/N}(\xi_N)$  and  $f_{d/N}(\xi_N)$  at  $\xi_N \sim 1/(3A)$ ; i.e., the position of the peaks in these nPDFs depends on the nucleus. In addition, the respective valence-quark distributions are normalized by the sum rules in the following nucleus-dependent way:

$$\begin{aligned} \int_0^1 [f_{u/N}(\xi_N, \mu^2) - f_{\bar{u}/N}(\xi_N, \mu^2)] d\xi_N &= A + Z, \\ \int_0^1 [f_{d/N}(\xi_N, \mu^2) - f_{\bar{d}/N}(\xi_N, \mu^2)] d\xi_N &= 2A - Z. \end{aligned} \quad (219)$$

In contrast, not only do the nPDFs for an average nucleon take into account the trivial  $A$  dependence, they also correctly incorporate the specific ratio of protons to neutrons. The nPDFs  $f_{a/p}^A(\xi_A, \mu^2)$  of a bound proton satisfy the sum rules

$$\begin{aligned} \int_0^A [f_{u/p}^A(\xi_A, \mu^2) - f_{\bar{u}/p}^A(\xi_A, \mu^2)] d\xi_A &= 2, \\ \int_0^A [f_{d/p}^A(\xi_A, \mu^2) - f_{\bar{d}/p}^A(\xi_A, \mu^2)] d\xi_A &= 1, \end{aligned} \quad (220)$$

which are much like the sum rules for the free proton.

Experimental analyses of nuclear DIS account for the trivial  $A$  dependence by presenting the cross sections or DIS structure functions not for the entire nucleus but rather per nucleon. Similarly, for collisions between two nuclei with atomic numbers  $A_1$  and  $A_2$ , cross sections  $\bar{\sigma}[F] \equiv \sigma[F]/(A_1 A_2)$  per nucleon are usually quoted. The cross section  $\bar{\sigma}[F]$  can be expressed using either type of nPDF:

$$\begin{aligned} \bar{\sigma}[F] &= \frac{1}{A_1 A_2} \sum_{a,b} \int_0^1 d\xi_a d\xi_b f_{a/N_1}(\xi_a, \mu_F^2) f_{b/N_2}(\xi_b, \mu_F^2) \\ &\quad \times \hat{\sigma}_{a,b,\xi_a,\xi_b,\mu_F^2}[F] + \mathcal{O}(M/Q) \\ &= \sum_{a,b} \int_0^{A_1} d\xi'_a \int_0^{A_2} d\xi'_b f_a^{A_1}(\xi'_a, \mu_F^2) f_b^{A_2}(\xi'_b, \mu_F^2) \\ &\quad \times \hat{\sigma}_{a,b,\xi'_a,\xi'_b,\mu_F^2}[F] + \mathcal{O}(M/Q). \end{aligned} \quad (221)$$

One should note that even in the case of noninteracting nucleons the cross section per nucleon  $\bar{\sigma}[F]$  differs from the cross section in proton-proton collisions due to the differing flavor decomposition.

To summarize, the trivial  $A$  dependence reflecting the sheer number of nucleons can be captured by using the ansatz (217) for the nPDF  $f_a^A(\xi_A, \mu^2)$  per average bound nucleon. On the right-hand side of Eq. (217), we introduce the PDFs  $f_{a/p}^A$  and  $f_{a/n}^A$  for bound protons and neutrons that acquire nontrivial  $A$  dependence from a combination of nuclear effects. Their parametrization at the input scale  $\mu_0$  is discussed in Sec. V.B.

## B. Parametrizing the $A$ dependence

In principle, one can extract the nPDFs  $f_{a/p}^A(\xi_A, \mu^2)$  of a bound proton from experimental data for each nucleus separately, without constructing a comprehensive model for initial-state nuclear effects. The current nuclear scattering data, however, are insufficient to determine the complete set of nPDFs for any single nucleus. The dependence of nuclear effects on  $\xi$  (also denoted as  $x$ ),  $A$ , and  $Z$  is assumed to be unknown from first principles. Thus, it must be determined in a global fit to experimental data. To assemble all scattering data taken on various nuclei within a common global analysis, a number of simplifying assumptions need to be made.

First, given that the nuclear modifications in the bound-proton PDFs  $f_{a/p}^A$  are expected to be small, it makes sense to use the free-proton PDFs  $f_{a/p}$  as the baseline for the parametrization of  $f_{a/p}^A$ .

Second, to use the available data, one makes an assumption that the bulk of the nuclear corrections depends only on  $A$ , the total number of nucleons of either isospin.

Third, the current data are not sufficient to constrain the nPDFs for momentum fractions  $\xi_A > 1$ , so all nPDF analyses

assume that  $0 < \xi_A < 1$ . However, future fits may be able to include the region  $1 < \xi_A$ .<sup>15</sup>

Fourth, we need to decide how to introduce the  $A$  dependence in  $f_{a/p}^A$ . In practice, one of two approaches is taken.

The first approach introduces nuclear-correction factors  $R_a(x, A)$  at the input scale  $\mu_0^2$  as follows:

$$f_{a/p}^A(x, A, \mu_0^2) = R_a(x, A) f_{a/p}(x, \mu_0^2),$$

$$\text{for } a = u_v, d_v, g, \bar{u} + \bar{d}, s, \bar{s}, \bar{d}/\bar{u}. \quad (222)$$

In Eq. (222)  $f_{a/p}(x, \mu_0^2)$ , the corresponding PDF for a free proton, is held fixed during any nPDF analysis. The PDF  $f_{a/n}^A(x, \mu_0^2)$  of a bound neutron is related to  $f_{a/p}^A(x, \mu_0^2)$  by charge symmetry. All free parameters associated with the nuclear modification are contained in  $R_a$ . For example, the EPPS16 analysis (Eskola *et al.*, 2017) used the following piecewise expression:

$$R_a(x, A) = \begin{cases} a_0 + a_1(x - x_a)^2, & 0 \leq x \leq x_a, \\ b_0 + b_1 x^\alpha + b_2 x^{2\alpha} + b_3 x^{3\alpha}, & x_a \leq x \leq x_e, \\ c_0 + (c_1 - c_2 x)(1 - x)^{-\beta}, & x_e \leq x \leq 1, \end{cases} \quad (223)$$

where  $\alpha = 10x_a$  and all parameters  $a_k$ ,  $b_k$  and  $c_k$  implicitly depend on the atomic number  $A$  and the PDF flavor  $a$ . A similar approach that employed a nuclear-correction factor was followed by HKN07 (Hirai, Kumano, and Nagai, 2007) and DSSZ (de Florian *et al.*, 2012). Each analysis used a different proton baseline; cf. Hirai, Kumano, and Nagai (2007), de Florian *et al.* (2012), and Eskola *et al.* (2017).

The second approach presented by Kovařík *et al.* (2016) did not operate with the nuclear-correction factors  $R_a$ . It instead parametrized the entire nPDF  $f_{a/p}^A(x, \mu_0^2)$  with a flexible functional form used for the free-proton PDF  $f_{a/p}(x, \mu_0^2)$ , but with  $A$ -dependent free parameters. As an example, in the nCTEQ15 analysis the explicit parametrization at the input scale was

$$x f_{a/p}^A(x, A, \mu_0^2) = c_0 x^{c_1} (1 - x)^{c_2} e^{c_3 x} (1 + e^{c_4 x})^{c_5},$$

$$\text{for } a = u_v, d_v, g, \bar{u} + \bar{d}, s, \bar{s}. \quad (224)$$

As in the underlying CTEQ6 parametrization (Pumplin *et al.*, 2002), the parton combination  $\bar{d}/\bar{u}$  was given by a different form

$$f_{\bar{d}/\bar{u}}^A(x, A, \mu_0^2) / f_{\bar{u}/\bar{d}}^A(x, A, \mu_0^2)$$

$$= c_0 x^{c_1} (1 - x)^{c_2} + (1 + c_3 x)(1 - x)^{c_4}. \quad (225)$$

All free parameters  $c_k$  depend on the atomic number

$$c_k(A) = c_{k,0} + c_{k,1}(1 - A^{-c_{k,2}}), \quad k = 1, \dots, 5. \quad (226)$$

The coefficient  $c_{k,0} = c_k(A = 1)$  is the underlying proton coefficient. It was held constant during the nCTEQ analysis. In the case of the nCTEQ15 analysis, the underlying proton coefficients  $c_{k,0}$  were set to the coefficients from the proton analysis performed by Owens *et al.* (2007).

Over the years, there have been other approaches to including nuclear effects in analyses of parton distribution functions. For example, the first next-to-leading-order analysis of nuclear parton distribution functions performed by de Florian and Sassot (2004) used a convolution approach to relate the nuclear PDF of an average nucleon to the one of a free nucleon as

$$f_{a/p}^A(x, \mu_0^2) = \int_x^A \frac{dy}{y} W_i(y, A, Z) f_{a/p}\left(\frac{x}{y}, \mu_0^2\right). \quad (227)$$

The analysis of proton PDFs using the neural networks was also extended to the analysis of nuclear PDFs (Abdul Khalek, Ethier, and Rojo, 2019). In that case the dependence on the number of nucleons was introduced as an additional parameter to the neural network. We do not go into detail for any of these approaches, as they have not yet been used in a global analysis of all nuclear scattering data.

### C. Comparisons of nuclear PDFs

The nPDFs  $f_{a/p}^A(x, A, \mu_0^2)$  of the bound proton are determined from experimental datasets taken on many different nuclei. Most of the data are still coming from deeply inelastic scattering and are provided in the form of nuclear-correction factors

$$R_{\text{DIS}}(x, \mu^2) = \frac{F_2^A(x, \mu^2)}{F_2^2(x, \mu^2)}. \quad (228)$$

The more recent data from neutrino DIS are provided as double-differential cross sections  $d^2\sigma/(dx dQ^2)$ . The collider data from Fermilab, RHIC, and the LHC are also provided as differential cross sections (per nucleon). The coverage of the relevant nuclear world data is nowhere close to that of the data available for free-nucleon PDFs. Many features of nPDFs are still poorly known, especially outside of the interval  $0.01 \lesssim x \lesssim 0.5$ . Most notably, no data constrain the nuclear gluon PDF at low momentum fractions.

The comparison of different nuclear PDF ensembles is a little trickier than comparing free-proton PDFs. The deficit of precise data introduces strong sensitivity to the prior and methodological assumptions, such as the kinematic cuts, nPDF parametrization form, and the choice of the baseline free-proton PDFs.

The methods introduced in Sec. IV can illustrate the differences between the various nPDF analyses. First, in Fig. 15 we show the distributions of  $S_E$ , defined in Sec. IV.H, from four recent NLO global nPDF analyses. As in the case of the proton analyses shown in Fig. 9, the distributions of  $S_E$  for the nPDF analyses are broader than the standard normal distribution  $\mathcal{N}(0, 1)$  expected from an ideal fit. Looking at the means and standard deviations of the distributions of  $S_E$  shown in Fig. 15, we see that all means are negative except for the HKN07 analysis, indicating that more experiments were fitted too well. This indicates possibly overestimated

<sup>15</sup>In the following text we denote the momentum fraction  $\xi_A$  restricted to the interval  $\xi_A \in (0, 1)$  for simplicity by  $x$ .

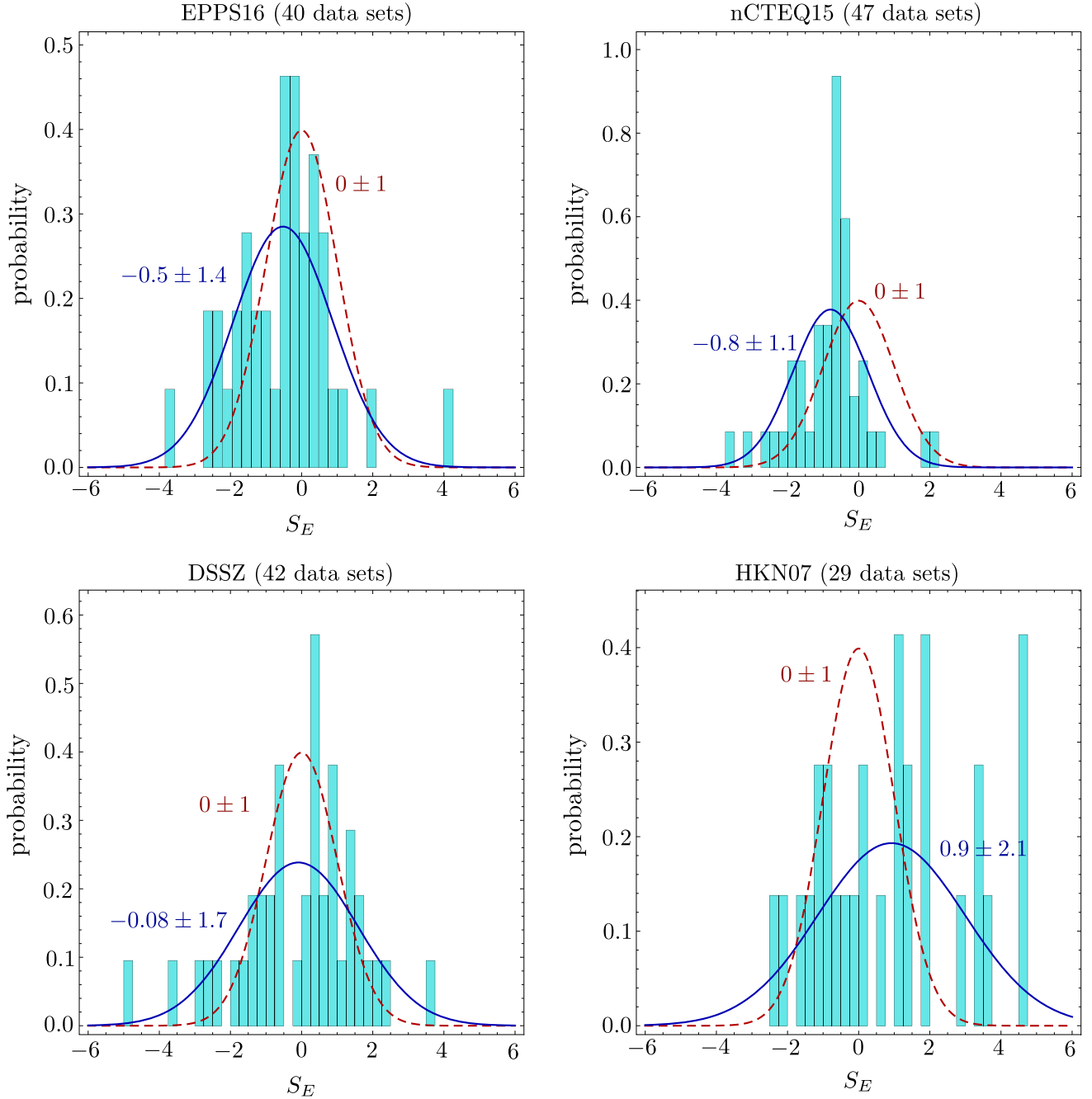


FIG. 15. Probability distributions in the effective Gaussian variable  $S_E$  for  $\chi^2$  values of the fitted datasets from the NLO nuclear PDF fits EPPS16, nCTEQ15, DSSZ, and HKN07.

uncertainties in multiple nuclear experiments. The prior assumptions made in the HKN07 analysis did not allow for a good description of many Drell-Yan total cross-section measurements by E772 and E866 experiments at Fermilab. Consequently, the  $S_E$  distribution for HKN07 had its mean shifted to the right, and it was wider. Some caution is needed when comparing the  $S_E$  distributions between the analyses in detail. For example, one entry with high  $S_E$  in the EPPS16 analysis is the double-differential neutrino DIS cross section from the CHORUS Collaboration. This experiment is not included in the nCTEQ15 and HKN07 analyses. In the DSSZ, it is included only in the form of the structure functions  $F_2$ .

We can quantify the observation that the  $S_E$  distributions are far from the ideal  $\mathcal{N}(0, 1)$  distribution using the Anderson-Darling test. The probability values that the distributions for the four nPDF analyses were drawn from  $\mathcal{N}(0, 1)$  are

$$\begin{aligned}
 P_{A-D} &= 6.8 \times 10^{-4}, & \text{EPPS16,} \\
 P_{A-D} &= 1.3 \times 10^{-5}, & \text{nCTEQ15,} \\
 P_{A-D} &= 1.4 \times 10^{-2}, & \text{DSSZ,} \\
 P_{A-D} &= 2.1 \times 10^{-5}, & \text{HKN07.}
 \end{aligned} \tag{229}$$

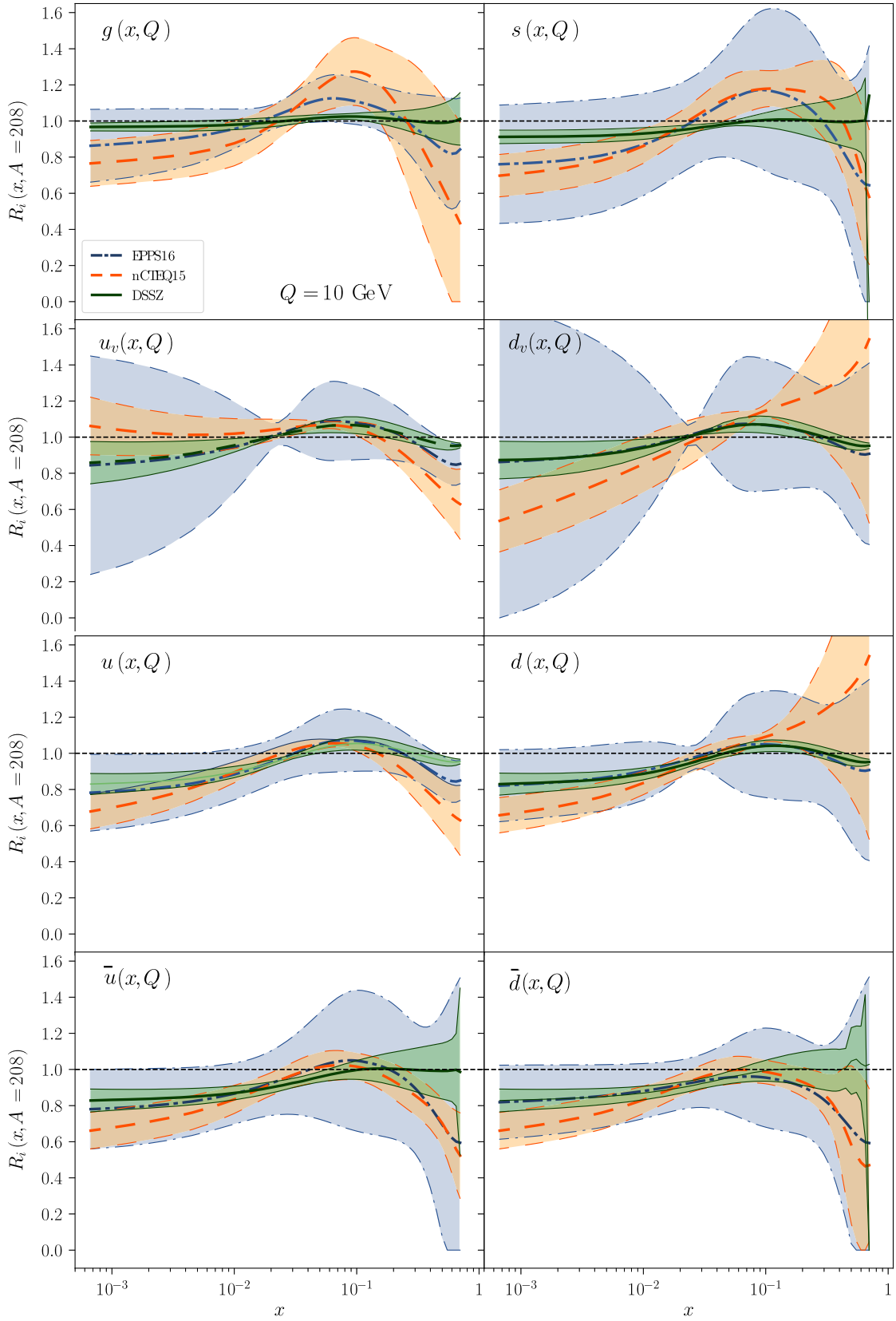


FIG. 16. Nuclear-correction factor  $R_i(x, A) = f_{i/p}^A(x, A, Q^2)/f_{i/p}(x, Q^2)$  for lead ( $A = 208$ ) and the partons  $i = g, s, u_v, d_v, \bar{u}, \bar{d}$  and at the momentum transfer  $Q = 10$  GeV.

With the possible exception of the DSSZ distribution, the Anderson-Darling test confirms that it is highly unlikely that the distributions in question come from the expected Gaussian

distribution. This is reminiscent of what we find in the proton case in Eq. (159); however, in three cases out of four, the nuclear data are fitted too well rather than too poorly.

The momentum fraction dependence of nPDFs is often examined by plotting scale-dependent nuclear-correction factors,

$$R_i(x, \mu^2, A) = \frac{f_{i/p}^A(x, A, \mu^2)}{f_{i/p}(x, \mu^2)}, \quad (230)$$

where  $f_{i/p}(x, \mu^2)$  is the baseline free-proton PDF. In Fig. 16, we turn to a comparison of the EPPS16, nCTEQ15, and DSSZ nuclear PDFs presented as these nuclear-correction factors. We show  $R_i(x, Q^2, A)$  at  $Q = 10$  GeV for lead ( $A = 208$ ), for which the nuclear effects are the largest. Broadly speaking, we can conclude that all three nPDF families are consistent with each other within the indicated uncertainties. Upon closer inspection, we see that the central values of  $R_i$  differ substantially among the three nPDF sets, in large part due to the strong dependence on the previously mentioned methodological assumptions, and most prominently due to the choice of the parametrization form. Furthermore, even though it is conventional to compare the ratios  $R_i$  rather than the nPDFs themselves, this quantity artificially introduces a dependence on the proton baseline. Much of the dependence on the baseline is absent when one compares the bound-proton PDFs  $f_{a/p}^A(x, A, \mu^2)$  directly.

The other notable difference among the results in Fig. 16 is their strikingly different uncertainties. One source of the differences is the various definitions of the uncertainties. All nPDF analyses employ some version of the global tolerance criterion that is based solely on the global  $\chi^2$ ; cf. Sec. IV.L. The DSSZ analysis uses the simplest version of the tolerance: their uncertainties correspond to varying the underlying parameters along the eigenvector directions [see Eq. (193)] by  $t = T = \sqrt{30}$ . Both the nCTEQ15 and EPPS16 analyses first examine a version of the dynamical tolerance, as described in Sec. IV.L, to estimate proper global tolerances for their final nPDF uncertainties. They determine the limits  $T_{\min}^i$  and  $T_{\max}^i$  according to Eq. (201) using the probability  $v = 0.90$  for each eigenvector direction  $e_i$ . Then, a global tolerance is constructed by averaging the changes in  $\chi^2$  over all eigenvector directions as

$$\begin{aligned} T^2 &= \sum_{i=1}^{N_p} \frac{\chi^2(a_{\text{fit}} + T_{\max}^i e_i) + \chi^2(a_{\text{fit}} + T_{\min}^i e_i) - 2\chi_0^2}{2n} \\ &= \sum_{i=1}^{N_p} \frac{(T_{\max}^i)^2 + (T_{\min}^i)^2}{2n}. \end{aligned} \quad (231)$$

The nCTEQ15 analysis has 16 free parameters ( $N_p = 16$ ) and generates the error PDFs in a standard manner for the global tolerance of  $T^2 = 35$ .

The EPPS16 analysis was the first to include the LHC data from proton-lead collisions. It uses 20 free parameters, and their prescription given by Eq. (231) yields the global tolerance of  $T^2 = 52$ .

If all nuclear PDF analyses were to use the same nuclear data in a specific range of momentum fractions and all analyses had a flexible parametrization form, the uncertainties would be similar. At present, the compared nPDF analyses do not fit the same data. Furthermore, as the four nPDF analyses

rely on the traditional minimization of global  $\chi^2$ , introducing more free nPDF parameters that can be constrained by the nuclear data would lead to unstable global fits. In Secs. IV.A and IV.B, we show how one can find the optimal number  $N_p$  of free parameters needed to obtain a stable fit to a given set of hadronic data. For the current nPDF analyses, the optimal number of free parameters appears to be no more than 15–20. Adding new data, for example, the LHC data that are included in the EPPS16 analysis, allows one to expand the constraints to a wider range of momentum fractions or new parton flavor combinations. With more LHC data expected in the near future, it will be possible to open up additional free parameters in the initial nPDF parametrizations, leading to a more realistic estimate of uncertainties on nuclear PDFs.

## VI. CONCLUSIONS

We have reviewed certain aspects of the fitting of collinear parton distribution functions (PDFs) to data. This is a large field. We have concentrated on just a few areas that could be of interest to the readers who use the PDFs or are interested in the rich subject of the global QCD analysis.

First, we have described the basic definition of what parton distribution functions are, and how they relate to the description of data. We have also provided definitions and a brief description for parton distributions in nuclei instead of just protons and neutrons.

Second, we have described the basic statistical treatment needed to fit the PDFs using what is often called the Hessian method. Our description is simplified compared to what is actually used in current PDF fits. We have assumed that, in a small enough neighborhood of the best fit, the theory predictions  $T_k(a)$  are approximately linear functions of the parameters  $a$ . This is not exactly the case, but this approximation is reasonably good when the PDF uncertainties are small and allows us to derive results in a closed form. Working within this framework, we have explored the statistical reasoning behind the fitting procedure and have derived analytic expressions for the key results of a PDF fit, such as expectation values and uncertainties.

We have then provided a battery of tests to critically examine whether the statistical assumptions are consistent with certain statistical measures that result from the fit. With these tests, one can identify specific features or data subsets in the multidimensional QCD fits that may indicate discrepancies between experimental measurements and theoretical predictions. Without insisting on a concrete recipe, we have presented some idea of what one can do in the case of inconsistency.

## ACKNOWLEDGMENTS

We are grateful to our CTEQ colleagues for the insightful discussions. We thank the Kavli Institute for Theoretical Physics at University of California in Santa Barbara, supported by National Science Foundation Grant No. PHY11-25915, for its hospitality during the start of the work on this review. We also thank the Aspen Center for Physics, supported by National Science Foundation Grant No. PHY-1607611, for its hospitality during a later stage of our work. Work at the

University of Oregon was supported by the U.S. Department of Energy under Grant No. DE-SC0011640. Work at Southern Methodist University was supported by the U.S. Department of Energy under Grant No. DE-SC0010129.

#### APPENDIX: TRANSFORMATION FOR $\chi^2(D, a, \lambda)$

In this appendix, we relate the form of Eq. (66) for  $\chi^2(D, a, \lambda)$  to the form of Eq. (68), in which it is apparent that the minimum of  $\chi^2(D, a, \lambda)$  with respect to the variables  $\lambda$  is  $\chi^2(D, a)$ . We begin with  $\chi^2(D, a, \lambda)$ , as given in Eq. (66),

$$\begin{aligned} \chi^2(D, a, \lambda) &= \sum_k \left[ \frac{D_k - T_k(a)}{\sigma_k} - \sum_I \beta_{kI} \lambda_I \right]^2 + \sum_J \lambda_J^2 \\ &= \sum_k \frac{[D_k - T_k(a)]^2}{\sigma_k^2} - 2 \sum_J \rho_J \lambda_J + \sum_{IJ} \lambda_I \lambda_J B_{IJ}, \end{aligned} \quad (\text{A1})$$

where  $B_{IJ}$  was defined in Eq. (69) as

$$B_{IJ} = \delta_{IJ} + \sum_k \beta_{kI} \beta_{kJ} \quad (\text{A2})$$

and

$$\rho_J \equiv \sum_k \frac{D_k - T_k(a)}{\sigma_k} \beta_{kJ}. \quad (\text{A3})$$

Completing the square in the variables  $\lambda$  gives

$$\begin{aligned} \chi^2(D, a, \lambda) &= \sum_k \frac{[D_k - T_k(a)]^2}{\sigma_k^2} - \sum_{IJ} \rho_I \rho_J B_{IJ}^{-1} \\ &\quad + \sum_{IJ} \left[ \lambda_I - \sum_K \rho_K B_{KI}^{-1} \right] B_{IJ} \left[ \lambda_J - \sum_L \rho_L B_{JL}^{-1} \right]. \end{aligned} \quad (\text{A4})$$

Define the shifted variables  $\lambda$  as

$$\lambda'_I = \lambda_I - \sum_K B_{IK}^{-1} \rho_K \quad (\text{A5})$$

and the matrix  $\tilde{C}_{ij}$  as

$$\tilde{C}_{ij} = \frac{1}{\sigma_i \sigma_j} \left\{ \delta_{ij} - \sum_{IJ} \beta_{iI} B_{IJ}^{-1} \beta_{jJ} \right\}. \quad (\text{A6})$$

This gives

$$\begin{aligned} \chi^2(D, a, \lambda) &= \sum_{ij} [D_i - T_k(a)][D_j - T_k(a)] \tilde{C}_{ij} \\ &\quad + \sum_{IJ} \lambda'_I \lambda'_J B_{IJ}. \end{aligned} \quad (\text{A7})$$

The matrix  $\tilde{C}_{ij}$  is in fact the covariance matrix  $C_{ij}$ . To prove this, use the definition from Eq. (56) of  $C_{ij}^{-1}$ , calculate  $\sum_j \tilde{C}_{ij} C_{jk}^{-1}$ , and simplify the product using  $\sum_j \beta_{jI} \beta_{jL} = B_{IL} - \delta_{IL}$ . The calculation gives

$$\sum_j \tilde{C}_{ij} C_{jk}^{-1} = \delta_{ik} \quad (\text{A8})$$

so that  $\tilde{C}_{ij} = C_{ij}$ .

We arrive at the following form of  $\chi^2(D, a, \lambda)$  given in Eq. (68):

$$\chi^2(D, a, \lambda) = \sum_{ij} [D_i - T_k(a)][D_j - T_k(a)] C_{ij} + \sum_{IJ} \lambda'_I \lambda'_J B_{IJ}. \quad (\text{A9})$$

It is clear that minimizing  $\chi^2(D, a, \lambda)$  with respect to  $\lambda$ , which is equivalent to setting  $\lambda'_{I,J} = 0$ , leaves only the first term in Eq. (A9), which is  $\chi^2(D, a)$  according to Eq. (59).

#### REFERENCES

- Aad, Georges, *et al.* (ATLAS Collaboration), 2010, ‘‘Observation of a Centrality-Dependent Dijet Asymmetry in Lead-Lead Collisions at  $\sqrt{s_{NN}} = 2.77$  TeV with the ATLAS Detector at the LHC,’’ *Phys. Rev. Lett.* **105**, 252303.
- Abdolmaleki, Hamed, *et al.* (xFitter Developers’ Team), 2018, ‘‘Impact of low- $x$  resummation on QCD analysis of HERA data,’’ *Eur. Phys. J. C* **78**, 621.
- Abdul Khalek, Rabah, Shaun Bailey, Jun Gao, Lucian Harland-Lang, and Juan Rojo, 2018, ‘‘Towards ultimate parton distributions at the high-luminosity LHC,’’ *Eur. Phys. J. C* **78**, 962.
- Abdul Khalek, Rabah, Jacob J. Ethier, and Juan Rojo (NNPDF Collaboration), 2019, ‘‘Nuclear parton distributions from lepton-nucleus scattering and the impact of an electron-ion collider,’’ *Eur. Phys. J. C* **79**, 471.
- Abdul Khalek, Rabah, *et al.*, 2019a, ‘‘A first determination of parton distributions with theoretical uncertainties,’’ [arXiv:1905.04311](https://arxiv.org/abs/1905.04311).
- Abdul Khalek, Rabah, *et al.* (NNPDF Collaboration), 2019b, ‘‘Parton distributions with theory uncertainties: General formalism and first phenomenological studies,’’ *Eur. Phys. J. C* **79**, 931.
- Ablinger, J., A. Behring, J. Blümlein, A. De Freitas, A. von Manteuffel, and C. Schneider, 2015, ‘‘The 3-loop pure singlet heavy flavor contributions to the structure function  $F_2(x, Q^2)$  and the anomalous dimension,’’ *Nucl. Phys.* **B890**, 48–151.
- Ablinger, J., J. Blümlein, A. De Freitas, A. Hasselhuhn, C. Schneider, and F. Wißbrock, 2017, ‘‘Three loop massive operator matrix elements and asymptotic Wilson coefficients with two different masses,’’ *Nucl. Phys.* **B921**, 585–688.
- Ablinger, J., J. Blümlein, S. Klein, C. Schneider, and F. Wißbrock, 2011, ‘‘The  $\mathcal{O}(\alpha_s^3)$  Massive operator matrix elements of  $\mathcal{O}(n_f)$  for the structure function  $F_2(x, Q^2)$  and transversity,’’ *Nucl. Phys.* **B844**, 26–54.
- Abramowicz, H., *et al.* (ZEUS and H1 Collaborations), 2015, ‘‘Combination of measurements of inclusive deep inelastic  $e^\pm p$  scattering cross sections and QCD analysis of HERA data,’’ *Eur. Phys. J. C* **75**, 580.
- Accardi, A., L. T. Brady, W. Melnitchouk, J. F. Owens, and N. Sato, 2016, ‘‘Constraints on large- $x$  parton distributions from new weak boson production and deep-inelastic scattering data,’’ *Phys. Rev. D* **93**, 114017.
- Accardi, A., M. E. Christy, C. E. Keppel, P. Monaghan, W. Melnitchouk, J. G. Morfin, and J. F. Owens, 2010, ‘‘New parton distributions from large- $x$  and low- $Q^2$  data,’’ *Phys. Rev. D* **81**, 034016.



- Accardi, A., W. Melnitchouk, J. F. Owens, M. E. Christy, C. E. Keppel, L. Zhu, and J. G. Morfin, 2011, “Uncertainties in determining parton distributions at large  $x$ ,” *Phys. Rev. D* **84**, 014008.
- Accardi, A., *et al.*, 2016, “A critical appraisal and evaluation of modern PDFs,” *Eur. Phys. J. C* **76**, 471.
- Accardi, Alberto, *et al.*, 2004, “Hard probes in heavy ion collisions at the LHC: PDFs, shadowing, and  $pA$  collisions,” [arXiv:hep-ph/0308248](https://arxiv.org/abs/hep-ph/0308248).
- Adam, Jaroslav, *et al.* (ALICE Collaboration), 2015, “Measurement of jet suppression in central Pb-Pb collisions at  $\sqrt{s_{NN}} = 2.76$  TeV,” *Phys. Lett. B* **746**, 1.
- Aitchison, I. J. R., and A. J. G. Hey, 2012, *Gauge Theories in Particle Physics: A Practical Introduction. Vol. 2: Non-Abelian Gauge Theories: QCD and the Electroweak Theory* (CRC Press, Boca Raton).
- Aivazis, M. A. G., John C. Collins, Fredrick I. Olness, and Wu-Ki Tung, 1994, “Leptoproduction of heavy quarks. II. A unified QCD formulation of charged and neutral current processes from fixed target to collider energies,” *Phys. Rev. D* **50**, 3102.
- Alekhin, S., J. Blumlein, S. Klein, and S. Moch, 2010, “3-, 4-, and 5-flavor next-to-next-to-leading order parton distribution functions from deep-inelastic-scattering data and at hadron colliders,” *Phys. Rev. D* **81**, 014032.
- Alekhin, S., J. Blumlein, and S. Moch, 2012, “Parton distribution functions and benchmark cross sections at next-to-next-to-leading order,” *Phys. Rev. D* **86**, 054009.
- Alekhin, S., J. Blümlein, S. Moch, and R. Placakyte, 2017, “Parton distribution functions,  $\alpha_s$ , and heavy-quark masses for LHC Run II,” *Phys. Rev. D* **96**, 014011.
- Alekhin, Sergei, 1999, “Extraction of parton distributions and  $\alpha_s$  from DIS data within the Bayesian treatment of systematic errors,” *Eur. Phys. J. C* **10**, 395.
- Alekhin, Sergey, *et al.*, 2011, “The PDF4LHC Working Group interim report,” [arXiv:1101.0536](https://arxiv.org/abs/1101.0536).
- Altarelli, Guido, and G. Parisi, 1977, “Asymptotic freedom in parton language,” *Nucl. Phys.* **B126**, 298–318.
- Amati, D., R. Petronzio, and G. Veneziano, 1978, “Relating hard QCD processes through universality of mass singularities. II,” *Nucl. Phys.* **B146**, 29–49.
- Anastasiou, Charalampos, Claude Duhr, Falko Dulat, Elisabetta Furlan, Franz Herzog, and Bernhard Mistlberger, 2015, “Soft expansion of double-real-virtual corrections to Higgs production at N<sup>3</sup>LO,” *J. High Energy Phys.* **08**, 051.
- Anderson, T. W., and D. A. Darling, 1952, “Asymptotic theory of certain goodness of fit criteria based on stochastic processes,” *Ann. Math. Stat.* **23**, 193–212.
- Armesto, Nestor, 2006, “Nuclear shadowing,” *J. Phys. G* **32**, R367–R394.
- ATLAS and CMS Collaborations, 2019, “Report on the physics at the HL-LHC and perspectives for the HE-LHC,” [arXiv:1902.10229](https://arxiv.org/abs/1902.10229).
- Aubert, J. J., *et al.* (European Muon Collaboration), 1983, “The ratio of the nucleon structure functions  $F_2^n$  for iron and deuterium,” *Phys. Lett.* **123B**, 275–278.
- Ball, Richard D., Valerio Bertone, Marco Bonvini, Stefano Carrazza, Stefano Forte, Alberto Guffanti, Nathan P. Hartland, Juan Rojo, and Luca Rottoli (NNPDF Collaboration), 2016, “A determination of the charm content of the proton,” *Eur. Phys. J. C* **76**, 647.
- Ball, Richard D., Valerio Bertone, Marco Bonvini, Simone Marzani, Juan Rojo, and Luca Rottoli, 2018, “Parton distributions with small- $x$  resummation: Evidence for BFKL dynamics in HERA data,” *Eur. Phys. J. C* **78**, 321.
- Ball, Richard D., Valerio Bertone, Stefano Carrazza, Luigi Del Debbio, Stefano Forte, Alberto Guffanti, Nathan P. Hartland, and Juan Rojo (NNPDF Collaboration), 2013, “Parton distributions with QED corrections,” *Nucl. Phys.* **B877**, 290–320.
- Ball, Richard D., Valerio Bertone, Francesco Cerutti, Luigi Del Debbio, Stefano Forte, Alberto Guffanti, Nathan P. Hartland, Jose I. Latorre, Juan Rojo, and Maria Ubiali, 2012, “Reweighting and unweighting of parton distributions and the LHC  $W$  lepton asymmetry data,” *Nucl. Phys.* **B855**, 608–638.
- Ball, Richard D., Valerio Bertone, Francesco Cerutti, Luigi Del Debbio, Stefano Forte, Alberto Guffanti, Jose I. Latorre, Juan Rojo, and Maria Ubiali (NNPDF Collaboration), 2011, “Reweighting NNPDFs: The  $W$  lepton asymmetry,” *Nucl. Phys.* **B849**, 112–143.
- Ball, Richard D., Stefano Carrazza, Luigi Del Debbio, Stefano Forte, Zahari Kassabov, Juan Rojo, Emma Slade, and Maria Ubiali (NNPDF Collaboration), 2018, “Precision determination of the strong coupling constant within a global PDF analysis,” *Eur. Phys. J. C* **78**, 408.
- Ball, Richard D., Luigi Del Debbio, Stefano Forte, Alberto Guffanti, Jose I. Latorre, Juan Rojo, and Maria Ubiali, 2010, “A first unbiased global NLO determination of parton distributions and their uncertainties,” *Nucl. Phys.* **B838**, 136–206.
- Ball, Richard D., *et al.*, 2013a, “Parton distributions with LHC data,” *Nucl. Phys.* **B867**, 244–289.
- Ball, Richard D., *et al.*, 2013b, “Parton distribution benchmarking with LHC data,” *J. High Energy Phys.* **04**, 125.
- Ball, Richard D., *et al.* (NNPDF Collaboration), 2017, “Parton distributions from high-precision collider data,” *Eur. Phys. J. C* **77**, 663.
- Ball, Richard D., *et al.* (NNPDF Collaboration), 2015, “Parton distributions for the LHC Run II,” *J. High Energy Phys.* **04**, 040.
- Bartels, J., Krzysztof J. Golec-Biernat, and H. Kowalski, 2002, “A modification of the saturation model: Dokshitzer-Gribov-Lipatov-Altarelli-Parisi evolution,” *Phys. Rev. D* **66**, 014001.
- Benvenuti, A. C., *et al.* (BCDMS Collaboration), 1990, “A high statistics measurement of the deuteron structure functions  $F_2(x, Q^2)$  and  $R$  from deep inelastic muon scattering at high  $Q^2$ ,” *Phys. Lett. B* **237**, 592–598.
- Berger, Edmond L., Jun Gao, Chong Sheng Li, Ze Long Liu, and Hua Xing Zhu, 2016, “Charm-Quark Production in Deep-Inelastic Neutrino Scattering at Next-to-Next-to-Leading Order in QCD,” *Phys. Rev. Lett.* **116**, 212002.
- Bertone, V., *et al.* (xFitter Developers’ Team), 2017, “Impact of the heavy quark matching scales in PDF fits,” *Eur. Phys. J. C* **77**, 837.
- Bertone, Valerio, Stefano Carrazza, Nathan P. Hartland, and Juan Rojo (NNPDF Collaboration), 2017, “Illuminating the photon content of the proton within a global PDF analysis,” [arXiv:1712.07053](https://arxiv.org/abs/1712.07053).
- Bertone, Valerio, Stefano Carrazza, and Juan Rojo, 2014, “APFEL: A PDF evolution library with QED corrections,” *Comput. Phys. Commun.* **185**, 1647–1668.
- Bertone, Valerio, Rikkert Frederix, Stefano Frixione, Juan Rojo, and Mark Sutton, 2014, “aMCfast: Automation of fast NLO computations for PDF fits,” *J. High Energy Phys.* **08**, 166.
- Bertone, Valerio, Alexandre Glazov, Alexander Mitov, Andrew Papanastasiou, and Maria Ubiali, 2018, “Heavy-flavor parton distributions without heavy-flavor matching prescriptions,” *J. High Energy Phys.* **04**, 046.
- Bierenbaum, Isabella, Johannes Blumlein, and Sebastian Klein, 2007, “Two-loop massive operator matrix elements and unpolarized heavy flavor production at asymptotic values  $Q^2 \gg m^2$ ,” *Nucl. Phys.* **B780**, 40–75.
- Bierenbaum, Isabella, Johannes Blumlein, and Sebastian Klein, 2009, “Mellin moments of the  $\mathcal{O}(\alpha_s^3)$  heavy flavor contributions

- to unpolarized deep-inelastic scattering at  $Q^2 \gg m^2$  and anomalous dimensions,” *Nucl. Phys.* **B820**, 417–482.
- Binoth, T., *et al.* (SM and NLO Multileg Working Group), 2010, “The SM and NLO Multileg Working Group: Summary report,” in *Proceedings of the 6th Les Houches Workshop on Physics at TeV Colliders, Les Houches, France, 2009*, edited by G. Belanger, F. Boudjema, and J.-P. Guillet (Laboratoire d’Annecy de Physique des Particules, Annecy, France), pp. 21–189.
- Bjorken, J. D., 1969, “Asymptotic sum rules at infinite momentum,” *Phys. Rev.* **179**, 1547–1553.
- Bjorken, J. D., John B. Kogut, and Davison E. Soper, 1971, “Quantum electrodynamics at infinite momentum: Scattering from an external field,” *Phys. Rev. D* **3**, 1382.
- Bodek, A., and J. L. Ritchie, 1981, “Fermi motion effects in deep inelastic lepton scattering from nuclear targets,” *Phys. Rev. D* **23**, 1070.
- Bodwin, Geoffrey T., 1985, “Factorization of the Drell-Yan cross section in perturbation theory,” *Phys. Rev. D* **31**, 2616.
- Botje, M., 2011, “QCDNUM: Fast QCD evolution and convolution,” *Comput. Phys. Commun.* **182**, 490–532.
- Boughezal, Radja, John M. Campbell, R. Keith Ellis, Christfried Focke, Walter Giele, Xiaohui Liu, Frank Petriello, and Ciaran Williams, 2017, “Color singlet production at NNLO in MCFM,” *Eur. Phys. J. C* **77**, 7.
- Bourilkov, D., R. C. Group, and M. R. Whalley, 2006, “LHAPDF: PDF use from the Tevatron to the LHC,” [arXiv:hep-ph/0605240](https://arxiv.org/abs/hep-ph/0605240).
- Brock, Raymond, *et al.* (CTEQ Collaboration), 1995, “Handbook of perturbative QCD: Version 1.0,” *Rev. Mod. Phys.* **67**, 157–248.
- Buckley, Andy, James Ferrando, Stephen Lloyd, Karl Nordström, Ben Page, Martin Rüfenacht, Marek Schönherr, and Graeme Watt, 2015, “LHAPDF6: Parton density access in the LHC precision era,” *Eur. Phys. J. C* **75**, 132.
- Butterworth, Jon, *et al.*, 2016, “PDF4LHC recommendations for LHC Run II,” *J. Phys. G* **43**, 023001.
- Buza, M., Y. Matiounine, J. Smith, R. Migneron, and W. L. van Neerven, 1996, “Heavy quark coefficient functions at asymptotic values  $Q^2 \gg m^2$ ,” *Nucl. Phys.* **B472**, 611–658.
- Buza, M., Y. Matiounine, J. Smith, and W. L. van Neerven, 1998, “Charm electroproduction viewed in the variable flavor number scheme versus fixed order perturbation theory,” *Eur. Phys. J. C* **1**, 301–320.
- Buza, M., and W. L. van Neerven, 1997, “ $O(\alpha_s^2)$  contributions to charm production in charged-current deep-inelastic lepton-hadron scattering,” *Nucl. Phys.* **B500**, 301–324.
- Cacciari, Matteo, and Nicolas Houdeau, 2011, “Meaningful characterisation of perturbative theoretical uncertainties,” *J. High Energy Phys.* **09**, 039.
- Campbell, John, Joey Huston, and Frank Krauss, 2017, *The Black Book of Quantum Chromodynamics* (Oxford University Press, New York).
- Campbell, John M., and R. K. Ellis, 2010, “MCFM for the Tevatron and the LHC,” *Nucl. Phys. B, Proc. Suppl.* **205–206**, 10–15.
- Caola, Fabrizio, Stefano Forte, and Juan Rojo, 2010, “Deviations from NLO QCD evolution in inclusive HERA data,” *Phys. Lett. B* **686**, 127–135.
- Carli, Tancredi, Dan Clements, Amanda Cooper-Sarkar, Claire Gwenlan, Gavin P. Salam, Frank Siegert, Pavel Starovoitov, and Mark Sutton, 2010, “A *posteriori* inclusion of parton density functions in NLO QCD final-state calculations at hadron colliders: The APPLGRID Project,” *Eur. Phys. J. C* **66**, 503–524.
- Carrazza, Stefano, Stefano Forte, Zahari Kassabov, Jose Ignacio Latorre, and Juan Rojo, 2015, “An unbiased Hessian representation for Monte Carlo PDFs,” *Eur. Phys. J. C* **75**, 369.
- Carrazza, Stefano, José I. Latorre, Juan Rojo, and Graeme Watt, 2015, “A compression algorithm for the combination of PDF sets,” *Eur. Phys. J. C* **75**, 474.
- Catani, Stefano, Leandro Cieri, Giancarlo Ferrera, Daniel de Florian, and Massimiliano Grazzini, 2009, “Vector Boson Production at Hadron Colliders: A Fully Exclusive QCD Calculation at Next-to-Next-to-Leading Order,” *Phys. Rev. Lett.* **103**, 082001.
- Catani, Stefano, Daniel de Florian, and German Rodrigo, 2012, “Space-like (versus time-like) collinear limits in QCD: Is factorization violated?,” *J. High Energy Phys.* **07**, 026.
- Catani, Stefano, and Massimiliano Grazzini, 2007, “Next-to-Next-to-Leading-Order Subtraction Formalism in Hadron Collisions and Its Application to Higgs-Boson Production at the Large Hadron Collider,” *Phys. Rev. Lett.* **98**, 222002.
- Chatrchyan, Serguei, *et al.* (CMS Collaboration), 2011, “Observation and studies of jet quenching in PbPb collisions at nucleon-nucleon center-of-mass energy = 2.76 TeV,” *Phys. Rev. C* **84**, 024906.
- Chetyrkin, K. G., Johann H. Kuhn, and M. Steinhauser, 2000, “RunDec: A *Mathematica* package for running and decoupling of the strong coupling and quark masses,” *Comput. Phys. Commun.* **133**, 43–65.
- Chuvakin, A., J. Smith, and W. L. van Neerven, 2000, “Comparison between variable flavor number schemes for charm quark electroproduction,” *Phys. Rev. D* **61**, 096004.
- Collins, John, 2013, *Foundations of perturbative QCD* (Cambridge University Press, Cambridge, England).
- Collins, John, and Jianwei Qiu, 2007, “ $k_T$  factorization is violated in production of high-transverse-momentum particles in hadron-hadron collisions,” *Phys. Rev. D* **75**, 114014.
- Collins, John C., 1998, “Hard scattering factorization with heavy quarks: A general treatment,” *Phys. Rev. D* **58**, 094002.
- Collins, John C., and Jon Pumplin, 2001, “Tests of goodness of fit to multiple data sets,” [arXiv:hep-ph/0105207](https://arxiv.org/abs/hep-ph/0105207).
- Collins, John C., and Davison E. Soper, 1982, “Parton distribution and decay functions,” *Nucl. Phys.* **B194**, 445–492.
- Collins, John C., Davison E. Soper, and George F. Sterman, 1983, “Factorization for one loop corrections in the Drell-Yan process,” *Nucl. Phys.* **B223**, 381–421.
- Collins, John C., Davison E. Soper, and George F. Sterman, 1985, “Factorization for short distance hadron-hadron scattering,” *Nucl. Phys.* **B261**, 104–142.
- Collins, John C., Davison E. Soper, and George F. Sterman, 1988, “Soft gluons and factorization,” *Nucl. Phys.* **B308**, 833–856.
- Collins, John C., and Wu-Ki Tung, 1986, “Calculating heavy quark distributions,” *Nucl. Phys.* **B278**, 934.
- Collins, John C., Frank Wilczek, and A. Zee, 1978, “Low-energy manifestations of heavy particles: Application to the neutral current,” *Phys. Rev. D* **18**, 242.
- Currie, J., E. W. N. Glover, and J. Pires, 2017, “Next-to-Next-to-Leading Order QCD Predictions for Single Jet Inclusive Production at the LHC,” *Phys. Rev. Lett.* **118**, 072002.
- Currie, James, E. W. N. Glover, Thomas Gehrmann, Aude Gehrmann-De Ridder, Alexander Huss, and Joao Pires, 2017, “Single jet inclusive production for the individual jet  $p_T$  scale choice at the LHC,” *Acta Phys. Pol. B* **48**, 955–967.
- Cybenko, G., 1989, “Approximation by superpositions of a sigmoidal function,” *Math. Control Signals Syst.* **2**, 303.
- de Florian, D., and R. Sassot, 2004, “Nuclear parton distributions at next-to-leading order,” *Phys. Rev. D* **69**, 074028.
- de Florian, D., *et al.* (LHC Higgs Cross Section Working Group), 2016, “Handbook of LHC Higgs Cross Sections: 4. Deciphering

- the nature of the Higgs sector,” arXiv:1610.07922, <https://doi.org/10.23731/CYRM-2017-002>.
- de Florian, Daniel, Rodolfo Sassot, Pia Zurita, and Marco Stratmann, 2012, “Global analysis of nuclear parton distributions,” *Phys. Rev. D* **85**, 074028.
- Devenish, Robin, and Amanda Cooper-Sarkar, 2004, *Deep Inelastic Scattering* (Oxford University Press, New York).
- Dokshitzer, Yuri L., 1977, “Calculation of the structure functions for deep inelastic scattering and  $e^+e^-$  annihilation by perturbation theory in quantum chromodynamics,” *Sov. Phys. JETP* **46**, 641–653.
- Duke, D. W., and J. F. Owens, 1984, “ $Q^2$  Dependent parametrizations of parton distribution functions,” *Phys. Rev. D* **30**, 49–54.
- Dulat, Sayipjamal, Tie-Jiun Hou, Jun Gao, Marco Guzzi, Joey Huston, Pavel Nadolsky, Jon Pumplin, Carl Schmidt, Daniel Stump, and C.-P. Yuan, 2016, “New parton distribution functions from a global analysis of quantum chromodynamics,” *Phys. Rev. D* **93**, 033006.
- Eichten, E., I. Hinchliffe, Kenneth D. Lane, and C. Quigg, 1984, “Super collider physics,” *Rev. Mod. Phys.* **56**, 579–707.
- Ellis, R. Keith, Howard Georgi, Marie Machacek, H. David Politzer, and Graham G. Ross, 1979, “Perturbation theory and the parton model in QCD,” *Nucl. Phys. B* **152**, 285–329.
- Eskola, Kari J., Petja Paakkinen, Hannu Paukkunen, and Carlos A. Salgado, 2017, “EPPS16: Nuclear parton distributions with LHC data,” *Eur. Phys. J. C* **77**, 163.
- Feynman, R. P., 1972, *Photon-Hadron Interactions* (CRC Press, Boca Raton).
- Fisher, Ronald A., 1925, *Statistical methods for research workers* (Oliver and Boyd, Edinburgh), Chap. 4.
- Forshaw, Jeffrey R., Michael H. Seymour, and Andrzej Siodmok, 2012, “On the breaking of collinear factorization in QCD,” *J. High Energy Phys.* **11**, 066.
- Forte, Stefano, Lluís Garrido, Jose I. Latorre, and Andrea Piccione, 2002, “Neural network parametrization of deep inelastic structure functions,” *J. High Energy Phys.* **05**, 062.
- Forte, Stefano, Andrea Isgro, and Gherardo Vita, 2014, “Do we need  $N^3$ LO parton distributions?,” *Phys. Lett. B* **731**, 136–140.
- Forte, Stefano, Eric Laenen, Paolo Nason, and Juan Rojo, 2010, “Heavy quarks in deep-inelastic scattering,” *Nucl. Phys. B* **834**, 116–162.
- Forte, Stefano, and Graeme Watt, 2013, “Progress in the determination of the partonic structure of the proton,” *Annu. Rev. Nucl. Part. Sci.* **63**, 291–328.
- Gao, Jun, 2011, “Implementation of higher-order contributions as systematic errors in the PDF analysis” (unpublished).
- Gao, Jun, Marco Guzzi, and Pavel M. Nadolsky, 2013, “Charm quark mass dependence in a global QCD analysis,” *Eur. Phys. J. C* **73**, 2541.
- Gao, Jun, Lucian Harland-Lang, and Juan Rojo, 2018, “The structure of the proton in the LHC precision era,” *Phys. Rep.* **742**, 1–121.
- Gao, Jun, and Pavel Nadolsky, 2014, “A meta-analysis of parton distribution functions,” *J. High Energy Phys.* **07**, 035.
- Gavin, Ryan, Ye Li, Frank Petriello, and Seth Quackenbush, 2011, “FEWZ2.0: A code for hadronic Z production at next-to-next-to-leading order,” *Comput. Phys. Commun.* **182**, 2388–2403.
- Gavin, Ryan, Ye Li, Frank Petriello, and Seth Quackenbush, 2013, “W physics at the LHC with FEWZ2.1,” *Comput. Phys. Commun.* **184**, 208–214.
- Geesaman, Donald F., K. Saito, and Anthony William Thomas, 1995, “The nuclear EMC effect,” *Annu. Rev. Nucl. Part. Sci.* **45**, 337.
- Gehrmann-De Ridder, A., T. Gehrmann, E. W. N. Glover, A. Huss, and T. A. Morgan, 2016, “Precise QCD Predictions for the Production of a Z Boson in Association with a Hadronic Jet,” *Phys. Rev. Lett.* **117**, 022001.
- Gehrmann-De Ridder, A., T. Gehrmann, E. W. N. Glover, A. Huss, and D. M. Walker, 2018, “Next-to-Next-to-Leading-Order QCD Corrections to the Transverse Momentum Distribution of Weak Gauge Bosons,” *Phys. Rev. Lett.* **120**, 122001.
- Giele, W., et al., 2002, “The QCD / SM Working Group: Summary report,” arXiv:hep-ph/0204316.
- Giele, Walter T., and Stephane Keller, 1998, “Implications of hadron collider observables on parton distribution function uncertainties,” *Phys. Rev. D* **58**, 094023.
- Giele, Walter T., Stephane A. Keller, and David A. Kosower, 2001, “Parton distribution function uncertainties,” arXiv:hep-ph/0104052.
- Giuli, F., et al. (xFitter Developers’ Team), 2017, “The photon PDF from high-mass Drell-Yan data at the LHC,” *Eur. Phys. J. C* **77**, 400.
- Gizhko, A., et al., 2017, “Running of the charm-quark mass from HERA deep-inelastic scattering data,” *Phys. Lett. B* **775**, 233–238.
- Glazov, Alexander, Sven Moch, and Voica Radescu, 2011, “Parton distribution uncertainties using smoothness prior,” *Phys. Lett. B* **695**, 238–241.
- Gluck, M., R. M. Godbole, and E. Reya, 1988, “Heavy-flavor production at high-energy  $ep$  colliders,” *Z. Phys. C* **38**, 441.
- Gluck, M., E. Hoffmann, and E. Reya, 1982, “Scaling violations and the gluon distribution of the nucleon,” *Z. Phys. C* **13**, 119.
- Golec-Biernat, Krzysztof J., and M. Wusthoff, 1999, “Saturation in diffractive deep inelastic scattering,” *Phys. Rev. D* **60**, 114023.
- Gribov, V. N., and L. N. Lipatov, 1972, “Deep inelastic  $ep$  scattering in perturbation theory,” *Sov. J. Nucl. Phys.* **15**, 438–450.
- Guzzi, Marco, Pavel M. Nadolsky, Hung-Liang Lai, and C.-P. Yuan, 2012, “General-mass treatment for deep inelastic scattering at two-loop accuracy,” *Phys. Rev. D* **86**, 053005.
- Harland-Lang, L. A., A. D. Martin, P. Motylinski, and R. S. Thorne, 2015, “Parton distributions in the LHC era: MMHT 2014 PDFs,” *Eur. Phys. J. C* **75**, 204.
- Harland-Lang, L. A., A. D. Martin, P. Motylinski, and R. S. Thorne, 2016, “The impact of the final HERA combined data on PDFs obtained from a global fit,” *Eur. Phys. J. C* **76**, 186.
- Harland-Lang, L. A., and R. S. Thorne, 2019, “On the consistent use of scale variations in PDF fits and predictions,” *Eur. Phys. J. C* **79**, 225.
- Harriman, P. N., Alan D. Martin, W. James Stirling, and R. G. Roberts, 1990, “Parton distributions extracted from data on deep inelastic lepton scattering, prompt photon production and the Drell-Yan process,” *Phys. Rev. D* **42**, 798–810.
- Hautmann, F., and Davison E. Soper, 2007, “Parton distribution function for quarks in an  $s$ -channel approach,” *Phys. Rev. D* **75**, 074020.
- Hirai, M., S. Kumano, and T. H. Nagai, 2007, “Determination of nuclear parton distribution functions and their uncertainties in next-to-leading order,” *Phys. Rev. C* **76**, 065207.
- Hobbs, T. J., Bo-Ting Wang, Pavel M. Nadolsky, and Fredrick I. Olness, 2019, “Charting the coming synergy between lattice QCD and high-energy phenomenology,” *Phys. Rev. D* **100**, 094040.
- Hornik, Kurt, 1991, “Approximation capabilities of multilayer feedforward networks,” *Neural Netw.* **4**, 251.
- Hornik, Kurt, Maxwell Stinchcombe, and Halbert White, 1990, “Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks,” *Neural Netw.* **3**, 551.
- Hou, Tie-Jiun, Sayipjamal Dulat, Jun Gao, Marco Guzzi, Joey Huston, Pavel Nadolsky, Jon Pumplin, Carl Schmidt, Daniel Stump, and C.-P. Yuan, 2017, “CTEQ-TEA parton distribution

- functions and HERA Run I and II combined data,” *Phys. Rev. D* **95**, 034003.
- Hou, Tie-Jiun, Sayipjamal Dulat, Jun Gao, Marco Guzzi, Joey Huston, Pavel Nadolsky, Carl Schmidt, Jan Winter, Keping Xie, and C.-P. Yuan, 2018, “CT14 intrinsic charm parton distribution functions from CTEQ-TEA global analysis,” *J. High Energy Phys.* **02**, 059.
- Hou, Tie-Jiun, *et al.*, 2019, “New CTEQ global analysis of quantum chromodynamics with high-precision data from the LHC,” [arXiv:1912.10053](https://arxiv.org/abs/1912.10053).
- Hou, Tie-Jiun, *et al.*, 2017, “Reconstruction of Monte Carlo replicas from Hessian parton distributions,” *J. High Energy Phys.* **03**, 099.
- James, F., and M. Roos, 1975, “MINUIT: A system for function minimization and analysis of the parameter errors and correlations,” *Comput. Phys. Commun.* **10**, 343–367.
- Jimenez-Delgado, P., T. J. Hobbs, J. T. Londergan, and W. Melnitchouk, 2015, “New Limits on Intrinsic Charm in the Nucleon from Global Analysis of Parton Distributions,” *Phys. Rev. Lett.* **114**, 082002.
- Khanpour, Hamzeh, and S. Atashbar Tehrani, 2016, “Global analysis of nuclear parton distribution functions and their uncertainties at next-to-next-to-leading order,” *Phys. Rev. D* **93**, 014026.
- Kniehl, B. A., G. Kramer, I. Schienbein, and H. Spiesberger, 2005a, “Collinear subtractions in hadroproduction of heavy quarks,” *Eur. Phys. J. C* **41**, 199–212.
- Kniehl, B. A., G. Kramer, I. Schienbein, and H. Spiesberger, 2005b, “Inclusive  $D^{*\pm}$  production in  $p\bar{p}$  collisions with massive charm quarks,” *Phys. Rev. D* **71**, 014018.
- Kogut, John B., and Davison E. Soper, 1970, “Quantum electrodynamics in the infinite momentum frame,” *Phys. Rev. D* **1**, 2901–2913.
- Kovářik, K., *et al.*, 2016, “nCTEQ15—Global analysis of nuclear parton distributions with uncertainties in the CTEQ framework,” *Phys. Rev. D* **93**, 085037.
- Kramer, G., and H. Spiesberger, 2004, “Inclusive photoproduction of  $D^*$  mesons with massive charm quarks,” *Eur. Phys. J. C* **38**, 309–318.
- Krämer, Michael, Fredrick I. Olness, and Davison E. Soper, 2000, “Treatment of heavy quarks in deeply inelastic scattering,” *Phys. Rev. D* **62**, 096007.
- Kunszt, Zoltan, and Davison E. Soper, 1992, “Calculation of jet cross-sections in hadron collisions at order  $\alpha_s^3$ ,” *Phys. Rev. D* **46**, 192–221.
- Kusina, A., F. I. Olness, I. Schienbein, T. Jezo, K. Kovarik, T. Stavreva, and J. Y. Yu, 2013, “Hybrid scheme for heavy flavors: Merging the fixed flavor number scheme and variable flavor number scheme,” *Phys. Rev. D* **88**, 074032.
- Lai, Hung-Liang, Marco Guzzi, Joey Huston, Zhao Li, Pavel M. Nadolsky, Jon Pumplin, and C.-P. Yuan, 2010, “New parton distributions for collider physics,” *Phys. Rev. D* **82**, 074024.
- Lai, Hung-Liang, Joey Huston, Zhao Li, Pavel Nadolsky, Jon Pumplin, Daniel Stump, and C.-P. Yuan, 2010, “Uncertainty induced by QCD coupling in the CTEQ global analysis of parton distributions,” *Phys. Rev. D* **82**, 054021.
- Lewis, Toby, 1988, “A simple improved-accuracy normal approximation for  $\chi^2$ ,” *Aust. N.Z. J. Stat.* **30A**, 160.
- Li, Ye, and Frank Petriello, 2012, “Combining QCD and electroweak corrections to dilepton production in the framework of the FEWZ simulation code,” *Phys. Rev. D* **86**, 094034.
- Lin, Huey-Wen, *et al.*, 2018, “Parton distributions and lattice QCD calculations: A community white paper,” *Prog. Part. Nucl. Phys.* **100**, 107–160.
- Malace, Simona, David Gaskell, Douglas W. Higinbotham, and Ian Cloet, 2014, “The challenge of the EMC effect: Existing data and future directions,” *Int. J. Mod. Phys. E* **23**, 1430013.
- Mangano, M. L., *et al.*, 2017, “Physics at a 100 TeV  $pp$  collider: Standard model processes,” *CERN Yellow Rep. Monogr.* **3**, 1–254.
- Manohar, Aneesh, Paolo Nason, Gavin P. Salam, and Giulia Zanderighi, 2016, “How Bright is the Proton? A Precise Determination of the Photon Parton Distribution Function,” *Phys. Rev. Lett.* **117**, 242002.
- Manohar, Aneesh V., Paolo Nason, Gavin P. Salam, and Giulia Zanderighi, 2017, “The photon content of the proton,” *J. High Energy Phys.* **12**, 046.
- Martin, A. D., R. G. Roberts, W. J. Stirling, and R. S. Thorne, 2004, “Uncertainties of predictions from parton distributions. II: Theoretical errors,” *Eur. Phys. J. C* **35**, 325–348.
- Martin, A. D., W. J. Stirling, R. S. Thorne, and G. Watt, 2009, “Parton distributions for the LHC,” *Eur. Phys. J. C* **63**, 189–285.
- Martin, A. D., A. J. Th. M. Mathijssen, W. J. Stirling, R. S. Thorne, B. J. A. Watt, and G. Watt, 2013, “Extended Parameterisations for MSTW PDFs and their effect on lepton charge asymmetry from  $W$  decays,” *Eur. Phys. J. C* **73**, 2318.
- Martin, Alan D., R. G. Roberts, and W. James Stirling, 1988, “Structure-function analysis and  $\psi$ , jet,  $W$ ,  $Z$  production: Pinning down the gluon,” *Phys. Rev. D* **37**, 1161.
- Martin, Alan D., R. G. Roberts, and W. James Stirling, 1989, “Improved parton distributions and  $W$ ,  $Z$  production at  $p\bar{p}$  colliders,” *Mod. Phys. Lett. A* **04**, 1135.
- Moch, S., and M. Rogal, 2007, “Charged current deep-inelastic scattering at three loops,” *Nucl. Phys.* **B782**, 51–78.
- Moch, S., J. A. M. Vermaseren, and A. Vogt, 2004, “The three loop splitting functions in QCD: The nonsinglet case,” *Nucl. Phys.* **B688**, 101–134.
- Moch, S., J. A. M. Vermaseren, and A. Vogt, 2005, “The longitudinal structure function at the third order,” *Phys. Lett. B* **606**, 123–129.
- Moch, S., J. A. M. Vermaseren, and A. Vogt, 2009, “Third-order QCD corrections to the charged-current structure function  $F_3$ ,” *Nucl. Phys.* **B813**, 220–258.
- Morfin, Jorge G., and Wu-Ki Tung, 1991, “Parton distributions from a global QCD analysis of deep inelastic scattering and lepton pair production,” *Z. Phys. C* **52**, 13–30.
- Mueller, Alfred H., 1999, “Parton saturation at small  $x$  and in large nuclei,” *Nucl. Phys.* **B558**, 285–303.
- Nadolsky, Pavel M., and Z. Sullivan, 2001, “PDF uncertainties in  $WH$  production at Tevatron,” *eConf C* **010630**, P510.
- Nathvani, Ricky, Robert Thorne, Lucian Harland-Lang, and Alan Martin, 2018, “*Ad lucem*: The photon in the MMHT PDFs,” *Proc. Sci., DIS2018*, 029 [arXiv:1807.07846].
- Olness, Fredrick I., and Davison E. Soper, 2010, “Correlated theoretical uncertainties for the one-jet inclusive cross section,” *Phys. Rev. D* **81**, 035018.
- Owens, J. F., 1991, “An updated set of parton distribution parameterizations,” *Phys. Lett. B* **266**, 126–130.
- Owens, J. F., J. Huston, C. E. Keppel, S. Kuhlmann, J. G. Morfin, F. Olness, J. Pumplin, and D. Stump, 2007, “The impact of new neutrino DIS and Drell-Yan data on large- $x$  parton distributions,” *Phys. Rev. D* **75**, 054030.
- Paukkunen, Hannu, and Pia Zurita, 2014, “PDF reweighting in the Hessian matrix approach,” *J. High Energy Phys.* **12**, 100.
- Peskin, Michael E., and Daniel V. Schroeder, 1995, *An introduction to quantum field theory* (Addison-Wesley, Reading, MA).
- Pumplin, J., D. Stump, R. Brock, D. Casey, J. Huston, J. Kalk, H.-L. Lai, and W.-K. Tung, 2001, “Uncertainties of predictions from

- parton distribution functions. II. The Hessian method,” *Phys. Rev. D* **65**, 014013.
- Pumplin, J., D. R. Stump, J. Huston, H.-L. Lai, Pavel M. Nadolsky, and W.-K. Tung, 2002, “New generation of parton distributions with uncertainties from global QCD analysis,” *J. High Energy Phys.* **07**, 012.
- Pumplin, Jon, 2010, “Parametrization dependence and  $\Delta\chi^2$  in parton distribution fitting,” *Phys. Rev. D* **82**, 114020.
- Qiu, Jianwei, 2003, “QCD factorization and rescattering in proton nucleus collisions,” [arXiv:hep-ph/0305161](https://arxiv.org/abs/hep-ph/0305161).
- Qiu, Jianwei, and George F. Sterman, 1991a, “Power corrections in hadronic scattering. I. Leading  $1/Q^2$  corrections to the Drell-Yan cross section,” *Nucl. Phys.* **B353**, 105–136.
- Qiu, Jianwei, and George F. Sterman, 1991b, “Power corrections to hadronic scattering. II. Factorization,” *Nucl. Phys.* **B353**, 137–164.
- Rothstein, Ira Z., and Iain W. Stewart, 2016, “An effective field theory for forward scattering and factorization violation,” *J. High Energy Phys.* **08**, 025.
- Saito, Koichi, and Toshihiro Uchiyama, 1985, “Effect of the Fermi motion on nuclear structure functions and the EMC effect,” *Z. Phys. A* **322**, 299.
- Salam, Gavin P., and Juan Rojo, 2009, “A higher order perturbative parton evolution toolkit (HOPPET),” *Comput. Phys. Commun.* **180**, 120–156.
- Sato, Nobuo, J. F. Owens, and Harrison Prosper, 2014, “Bayesian reweighting for global fits,” *Phys. Rev. D* **89**, 114020.
- Schienbein, Ingo, *et al.*, 2008, “A review of target mass corrections,” *J. Phys. G* **35**, 053101.
- Schmidt, Carl, Jon Pumplin, Daniel Stump, and C.-P. Yuan, 2016, “CT14QED parton distribution functions from isolated photon production in deep inelastic scattering,” *Phys. Rev. D* **93**, 114015.
- Schmidt, Carl, Jon Pumplin, and C.-P. Yuan, 2018, “Updating and optimizing error parton distribution function sets in the Hessian approach,” *Phys. Rev. D* **98**, 094005.
- Schwartz, Matthew D., 2014, *Quantum Field Theory and the Standard Model* (Cambridge University Press, Cambridge, England).
- Schwartz, Matthew D., Kai Yan, and Hua Xing Zhu, 2018, “Factorization violation and scale invariance,” *Phys. Rev. D* **97**, 096017.
- Srednicki, M., 2007, *Quantum Field Theory* (Cambridge University Press, Cambridge, England).
- Sterman, George F., and Steven Weinberg, 1977, “Jets from Quantum Chromodynamics,” *Phys. Rev. Lett.* **39**, 1436.
- Stump, D., J. Pumplin, R. Brock, D. Casey, J. Huston, J. Kalk, H.-L. Lai, and W.-K. Tung, 2001, “Uncertainties of predictions from parton distribution functions. I. The Lagrange multiplier method,” *Phys. Rev. D* **65**, 014012.
- Tanabashi, M., *et al.* (Particle Data Group), 2018, “Review of particle physics,” *Phys. Rev. D* **98**, 030001.
- Thorne, R. S., 2006, “Variable-flavor number scheme for next-to-next-to-leading order,” *Phys. Rev. D* **73**, 054019.
- Thorne, R. S., 2014, “The effect on PDFs and  $\alpha_S(M_Z^2)$  due to changes in flavour scheme and higher twist contributions,” *Eur. Phys. J. C* **74**, 2958.
- Thorne, R. S., and R. G. Roberts, 1998, “An ordered analysis of heavy-flavor production in deep inelastic scattering,” *Phys. Rev. D* **57**, 6871–6898.
- Thorne, Robert S., Shaun Bailey, T. Cridge, Lucian A. Harland-Lang, A. D. Martin, and Ricky Nathvani, 2019, “Updates of PDFs using the MMHT framework,” *Proc. Sci. DIS2019*, 036 [[arXiv:1907.08147](https://arxiv.org/abs/1907.08147)].
- Thorne, Robert S., Lucian A. Harland-Lang, and Alan D. Martin, 2018, “MMHT Updates, LHC jets and  $\alpha_S$ ,” *Proc. Sci. DIS2018*, 030.
- Ueda, Takahiro, 2018, “Towards the N<sup>3</sup>LO evolution of parton distributions,” *Proc. Sci. DIS2018*, 034.
- van Neerven, W. L., and E. B. Zijlstra, 1991, “Order  $\alpha_s^2$  contributions to the deep inelastic Wilson coefficient,” *Phys. Lett. B* **272**, 127–133.
- Vermaseren, J. A. M., A. Vogt, and S. Moch, 2005, “The third-order QCD corrections to deep-inelastic scattering by photon exchange,” *Nucl. Phys.* **B724**, 3–182.
- Vogt, A., 2005, “Efficient evolution of unpolarized and polarized parton distributions with QCD-PEGASUS,” *Comput. Phys. Commun.* **170**, 65–92.
- Vogt, A., S. Moch, and J. A. M. Vermaseren, 2004, “The Three-loop splitting functions in QCD: The singlet case,” *Nucl. Phys.* **B691**, 129–181.
- Watt, G., and R. S. Thorne, 2012, “Study of Monte Carlo approach to experimental uncertainty propagation with MSTW 2008 PDFs,” *J. High Energy Phys.* **08**, 052.
- Whalley, M. R., D. Bourilkov, and R. C. Group, 2005, “The Les Houches accord PDFs (LHAPDF) and LHAGLUE,” [arXiv:hep-ph/0508110](https://arxiv.org/abs/hep-ph/0508110).
- Wilson, Kenneth G., 1969, “Non-Lagrangian models of current algebra,” *Phys. Rev.* **179**, 1499–1512.
- Wobisch, M., D. Britzger, T. Kluge, K. Rabbertz, and F. Stober (fastNLO Collaboration), 2011, “Theory-data comparisons for jet measurements in hadron-induced processes,” [arXiv:1109.1310](https://arxiv.org/abs/1109.1310).
- Zijlstra, E. B., and W. L. van Neerven, 1991, “Contribution of the second order gluonic Wilson coefficient to the deep inelastic structure function,” *Phys. Lett. B* **273**, 476–482.