Parametrization dependence and $\Delta \chi^2$ in parton distribution fitting

Jon Pumplin

Michigan State University, East Lansing MI 48824, USA

(Dated: March 15, 2010)

Abstract

Parton distribution functions (PDFs), which describe probability densities of quarks and gluons in the proton, are essential for interpreting the data from high energy colliders. The PDFs are measured by approximating them with functional forms that contain many adjustable parameters. Those parameters are determined by fitting a wide variety of experimental data. This paper examines the uncertainty that arises in this procedure from the choice of parametrization, and shows how that uncertainty can be reduced by a technique based on Chebyshev polynomials.
1. INTRODUCTION

Interpreting the data from high energy colliders such as the Tevatron and LHC requires parton distribution functions (PDFs), which describe the probability density for quarks and gluons in the proton, as a function of light-cone momentum fraction $x$ and QCD factorization scale $\mu$. In current practice [1, 2], the PDFs are measured by parametrizing them at a low scale $\mu_0$, using functional forms in $x$ that contain many adjustable parameters. The distributions at higher $\mu$ are calculated using QCD renormalization group equations, and the best-fit parameter values are found by a “global analysis,” in which data from a variety of experiments are simultaneously fitted by minimizing a $\chi^2$ measure of fit quality.

In the Hessian method [3], the uncertainty range of the PDFs is estimated by accepting all fits for which $\chi^2$ is not more than some fixed constant $\Delta\chi^2$ above the best-fit value. Traditionally, $\Delta\chi^2 = 100$ [4] or $\Delta\chi^2 = 50$ [5] have been used. When these “large” $\Delta\chi^2$ values are used, weight factors or penalties are included to maintain an adequate fit to each individual data set [1]; or the uncertainty range along each eigenvector direction can instead be estimated as the region where the fit to every experiment is acceptable [2]. The large $\Delta\chi^2$ tolerance criterion has long been a source of controversy, and some PDF groups [6, 7] eschew it in favor of the $\Delta\chi^2 = 1$ for 68% confidence ($\Delta\chi^2 = 2.7$ for 90% confidence) that would be expected from Gaussian statistics.

A potential motivation for large $\Delta\chi^2$ is based on conflicts among the input data sets, which signal unknown systematic errors in the experiments, or important theoretical errors introduced by reliance on leading-twist NLO perturbation theory. It would make sense to scale up the experimental errors to allow for such conflicts [8]. However, the conflicts among experiments were recently shown [9, 10] to be fairly small: measured discrepancies between the implications from each experiment with respect to all of the others suggest a minimum $\Delta\chi^2 \approx 10$ for 90% confidence, but supply no incentive for $\Delta\chi^2 \approx 100$.

Another source of uncertainty in PDF determination is the parametrization dependence error, which comes from representing PDFs at $\mu_0$—unknown continuous functions—by expressions that are adjustable only through a finite number of free parameters. In traditional practice, one adds flexibility to the parametrizing functions one parameter at a time, until the resulting minimum $\chi^2$ ceases to decrease “significantly.” However, wherever one chooses to stop adding fitting
parameters, a further small decreases in $\chi^2$ remain possible. This aspect of the PDF problem, namely that the number of fitting parameters is not uniquely defined, spoils the normal rules such as $\Delta \chi^2 = 1$ for 68% confidence, which would otherwise follow from standard Gaussian statistics. This point is illustrated in Sect. 2 by some hypothetical examples. A method that invokes Chebyshev polynomials to increase the freedom of the parametrization is introduced in Sect. 3 and applied to a typical PDF fit in Sect. 4. Conclusions are presented in Sect. 5.

2. HYPOTHETICAL EXAMPLES

Let $z$ be the displacement from the minimum point in $\chi^2$ along some specific direction in parameter space. It can be normalized such that

$$\chi^2 = z^2 + C$$

in the neighborhood of the minimum. The parameter $z$ can be any one of the eigenvector coefficients $z_i$ that are discussed in [9] or [10]. Or by means of a suitable linear transformation $X = \alpha + \beta z$, $z$ can represent the prediction for some cross section $X$ that depends on the PDFs; or simply a PDF itself for some specific flavor, $x$, and $\mu$. According to standard statistics, Eq. (1) would imply $z = 0 \pm 1$ at 68% confidence and $0 \pm 1.64$ at 90% confidence. If we assume instead—guided by [10]—that $\Delta \chi^2 = 10$ for 90% confidence, we would expect $z = 0 \pm 3.16$ at that confidence. However, the following argument shows that the uncertainty range could actually be much broader than that.

Suppose that, in order to reduce the dependence on the choice of parametrization, we introduce additional flexibility into the PDF model through a new parameter $y$, which is defined such that $\chi^2$ reduces to Eq. (1) at $y = 0$. (To achieve a substantially improved fit, it will likely be necessary to increase the flexibility in more than one flavor, and therefore it will be necessary to introduce several new fitting parameters. The parameter $y$ therefore represents displacement in a direction that is defined by some particular linear combination of new and old parameters.)

Figure 1 shows two hypothetical contour plots for $\chi^2$ as a function of $y$ and $z$. The contour interval is 10. In each case, introducing the new parameter reveals that $z = 10$ is a better estimate of the true value of $z$, so the prediction according to $y = 0$, that $z = 0 \pm 3.16$ at 90%
FIG. 1: Contours of $\chi^2 = 3010, 3020, 3030, \ldots$ in two hypothetical examples. In each case, the best fit has $\chi^2 = 3000$ at $z = 10$. Meanwhile, if the fit is confined to $y = 0$, the $\Delta \chi^2 = 10$ error limits appear to be $z = 0 \pm 3.16$, which is far too restrictive.

...confidence, is incorrect. In the scenario of the left panel, the additional freedom measured by $y$ has reduced the best-fit $\chi^2$ by 50; while in the right panel, the reduction is only 5—a change so small that one might easily have been content to mistakenly settle for $y = 0$.

In the hypothetical examples of Fig. 1, $\Delta \chi^2 = 10$ yields an estimate of uncertainty for fits with $y = 0$ that is far too narrow—even though in one case, the additional freedom only allows $\chi^2$ to be lowered by 5. Appendix I shows that the qualitative form of the dependence of $\chi^2$ on $y$ and $z$ shown in Fig. 1 arises rather generally, when additional freedom is introduced into the parametrizations. However, it remains to be seen whether such large quantitative changes arise in typical PDF fitting. A new parametrization method introduced in the next Section will be used to answer that question.
3. CHEBYSHEV PARAMETRIZATIONS

In a recent typical PDF fit (CT09 [1]), the gluon distribution at $\mu_0$ was parametrized by

$$x g(x) = a_0 x^{a_1} (1 - x)^{a_2} e^{p(x)}$$

(2)

where

$$p(x) = a_3 \sqrt{x} + a_4 x + a_5 x^2.$$  

(3)

The same form was used—with different parameters of course—for $u_v = u - \bar{u}$ and $d_v = d - \bar{d}$, except that $a_3$ was set to 0 in $d_v$, because that distribution is less constrained by data.

To provide greater flexibility in the parametrization, it would be natural to replace $p(x)$ by a general polynomial in $\sqrt{x}$:

$$p(x) = \sum_{j=1}^{n} b_j x^{j/2}.$$  

(4)

This form has several attractive features:

1. The power-law dependence in $x$ at $x \to 0$, with subleading terms suppressed by additional powers of approximately $x^{0.5}$, is expected from Regge theory.

2. The power-law suppression in $(1 - x)$ at $x \to 1$ is expected from spectator counting arguments.

3. The exponential form $e^{p(x)}$ allows for the possibility of a large ratio between the coefficients of the power-law behaviors at $x \to 0$ and $x \to 1$, without requiring large coefficients. It also conveniently guarantees that $g(x)$ is positive definite—although that could in principle be an unnecessarily strong assumption, since the $\overline{\text{MS}}$ parton distributions are not directly observable, so it is only required that predictions for all possible cross sections be positive.

4. Restricting the order $n$ of the polynomial in Eq. (4) can express the assumed smoothness of the parton distributions—although if $n$ is large, some additional condition must still be imposed to prevent rapid variation.

The constraints on smoothness and behavior at $x \to 0$ and $x \to 1$ are important. Without them, the momentum sum rule and the valence quark number sum rules would have no power,
because mildly singular contributions near $x = 0$ and $x = 1$ could make arbitrary contributions to those integrals over $x$, without otherwise affecting any predictions.

In past practice, only a small number of non-zero parameters $b_j$ have been retained in (4), as exemplified by the typical choice (3). The number of parameters can be increased to add flexibility, and thereby reduce the dependence on the choice of form for parametrization. However, that quickly runs into a technical difficulty: as more fitting parameters are included, the numerical procedure to find the minimum of $\chi^2$ becomes unstable, with large coefficients and strong cancellations arising in $p(x)$. The resulting best fits, if they can be found at all, contain implausibly rapid variations.

This technical difficulty can be overcome using a method based on Chebyshev polynomials. These polynomials have a long tradition in numerical analysis, but have only recently begun to be applied in PDF studies \[11\]. They are defined—and conveniently calculated—by recursion:

$$
T_0(y) = 1, \quad T_1(y) = y \\
T_{n+1}(y) = 2yT_n(y) - T_{n-1}(y).
$$

(5)

Since $T_j(y)$ is a polynomial of order $j$ in $y$, the parametrization (4) can be rewritten as

$$
p(x) = \sum_{j=1}^{n} c_j T_j(y),
$$

(6)

where $y = 1 - 2\sqrt{x}$ conveniently maps the physical region $0 < x < 1$ to $-1 < y < 1$.

The parameters $c_1, \ldots, c_n$ are formally equivalent to the parameters $b_1, \ldots, b_n$; but they are more convenient for fitting, because the requirement for smoothness in the input PDFs can be expressed by forcing the $c_j$ parameters to be reasonably small, particularly at large order. This can be seen from the following property of the Chebyshev polynomials:

$$
T_j(y) = \cos(j\theta) \quad \text{where} \quad y = \cos \theta.
$$

(7)

With the mapping $y = 1 - 2\sqrt{x}$ (or more generally $y = 1 - 2x^a$), $T_j(y)$ has extreme values of $\pm 1$ at the endpoints and at $j - 1$ points in the interior of the physical region $0 < x < 1$. Chebyshev polynomials of increasingly large $j$ thus model structure at an increasingly fine scale in $x$. A particular way to force the coefficients $c_j$ to be small, especially at large $j$, in order to enforce smoothness, is described in Appendix II.
Using the Chebyshev polynomial method, it becomes possible to obtain fits, and even eigenvector uncertainty sets, with two to three times more free parameters than were tractable in previous PDF fitting. The method is applied in the next Section to evaluate the parametrization error in a traditional fit.

4. FITS USING THE CHEBYSHEV METHOD

![Graph showing PDF uncertainties](image)

**FIG. 2:** Uncertainty of gluon distribution, and $d$ and $u$ quark distributions, calculated at $\Delta \chi^2 = 10$ using the CTEQ6.6 form of parametrization with 22 parameters, compared to fits using the Chebyshev polynomial method with 26 parameters (long dashed), 31 parameters (medium dashed), and 36 parameters (short dashed). The two versions of each curve (with or without dots) correspond to different choices for the smoothness penalty.

Figure 2 shows the fractional PDF uncertainty obtained using the parametrization method of CTEQ6.6 [4], which has 22 free parameters. The uncertainty limit is defined here by $\Delta \chi^2 = 10$, which is the range suggested by observed conflicts among the input data sets [10]. The curves show the ratio to that central fit, of fits made using the Chebyshev polynomial method defined specifically in Appendix II, with 26, 31, and 36 free parameters. Two fits are shown for each of these choices, corresponding to different strengths of the penalty that is added to $\chi^2$ to enforce smoothness. Introducing more flexible parametrizations is seen to shift the best-fit estimate of the PDFs by an amount that is in some places comparable to or somewhat outside the range of
uncertainty that was estimated on the basis of conflicts among the input data sets.

Figure 3 displays the same results over a wider range in the ratio, using a linear scale to reveal the behavior at large \( x \). In the region \( x > 0.8 \), where the PDFs are very poorly constrained, the improved parametrizations sometimes lead to fits that lie far outside the original uncertainty band. This comes about because the uncertainty range was obtained using the Hessian eigenvector method, which is based on assuming quadratic behavior (second-order Taylor series) of \( \chi^2 \) as a function of the fitting parameters. For quantities that are weakly constrained (i.e., along “flat directions” in parameter space), this approximation can easily break down before \( \chi^2 \) has increased very much.

Table I lists \( \chi^2 \) for the fits shown in Figs. 2 and 3. All of these fits would be acceptable according to a crude “hypothesis testing” criterion, since \( \chi^2 \) is close to the total number of data points (minus the number of fitting parameters)—well within the range \( \pm \sqrt{2N_{\text{pts}}} \). The additional parametrization freedom allowed by the Chebyshev method produces a distinctly lower overall \( \chi^2 \), and it therefore can be assumed to provide a global best fit that is a better approximation to the true PDFs. This is analogous to moving from the minimum \( \chi^2 \) along \( y = 0 \) to the minimum \( \chi^2 \) in the \((z, y)\) plane depicted in Fig. 1. Using more than 36 parameters for this data set offers negligible further improvement in the fit, since beyond that number, the further decrease in \( \chi^2 \) becomes smaller than the number of added parameters, so \( \chi^2 \) per degree of freedom no longer decreases.
5. CONCLUSION

Improving the flexibility of the input parametrizations used in PDF analysis, using a method based on Chebyshev polynomials, reveals a significant uncertainty in previous PDF determinations caused by parametrization dependence. In regions where the parton distributions are fairly accurately constrained by experiment, the uncertainty due to parametrization is roughly comparable to the uncertainty suggested by observed conflicts between different data sets that are used in the global analysis. In regions where the experimental constraints are weak, such as the large-$x$ regions shown in Fig. 3, the parametrization-dependence uncertainty is considerably larger. This comes about because the uncertainty there comes mainly from parametrization dependence, and the reliance of the Hessian method on the quadratic approximation naturally underestimates such uncertainties.

The parametrization uncertainty must be combined with the uncertainty evidenced by conflicts among the input data sets. The latter has been previously estimated to be given by $\Delta \chi^2 = 10$ at 90% confidence [10]. The parametrization uncertainty found here is on the same order, so combining the two uncertainties naïvely leads to a prescription of $\Delta \chi^2 = 20$ or $\Delta \chi^2 = 40$, depending on whether the uncertainties are added in quadrature or linearly. Because there are
additional uncertainties that have not been treated here, such as the choice of the effective mass of the charm quark, and the choice of which data sets to include in the analysis, the use of $\Delta \chi^2$ in the range $50 - 100$ for 90% confidence continues to appear reasonable. This estimate is also supported by recent fitting, which will be reported elsewhere, in which the many DIS data sets from the H1 and ZEUS experiments at HERA are replaced by a single joint data set in which the experimentalists have reduced the systematic errors by combining results from the two experiments. The best fit using the new data set is different from the best fit using the old data by an amount corresponding to $\Delta \chi^2 = 55$.

Further confirmation of the size of the uncertainties from a practical point of view is shown in Fig. 4, which compares fits shown in Fig. 2 with the original CTEQ6.6 for $d(x)$ at $\mu = 100$ GeV. The wide band is the CTEQ6.6 uncertainty (weighted $\Delta \chi^2 = 100$). The narrow band and solid curve are the new fit using the CTEQ6.6 parametrization (22 parameters). The central fit differs from CTEQ6.6 because of additions to the input data (inclusive jet and $Z^0$ rapidity distribution data from run II at CDF and D0), the introduction of a mild cut, $(1-x)Q^2 > 9$ GeV$^2$, on fixed-target muon DIS data as suggested by results of [10], and because all experiments are assigned weight 1 in the new fit. The width of the uncertainty is smaller, mainly because $\Delta \chi^2 = 10$ was assumed here. The dot-dashed curve is the most extreme of the Chebyshev fits shown in Fig. 2 (36 parameters). Note that the difference between this curve and the result of the CTEQ6.6 parametrization is nearly as large as the uncertainty originally estimated in CTEQ6.6. This proves that the CTEQ6.6 parametrization form induces an uncertainty due to parametrization at the level of $\Delta \chi^2 = 100$. It is reasonable to expect that the parametrization uncertainty is somewhat smaller than that in the case of the Chebyshev fits, since they are more flexible.

An alternative to the Hessian method is provided by the NNPDF approach [13], which represents the PDFs at $\mu_0$ by a Neural Network model that contains a very large number of effective parameters, and thereby avoids parametrization dependence almost entirely. Broadly speaking, the uncertainties estimated by NNPDF are similar in magnitude to the uncertainties estimated previously by the Hessian method using $\Delta \chi^2 \sim 100$. A more detailed comparison, which requires a discussion of differences in assumptions based on nonperturbative hadronic physics—such as positivity and behaviors in $x \to 0$ and $x \to 1$ limits—will be undertaken in a
FIG. 4: Uncertainty of $d$ at $\mu = 100$ GeV, relative to CTEQ6.6 \[4\]. Wide band is the CTEQ6.6 uncertainty (weighted $\Delta \chi^2 = 100$). Narrow band and solid curve are the fit using CTEQ6.6 parametrization (22 parameters, $\Delta \chi^2 = 10$) shown in Fig. 2. Dot-dashed curve is a fit with 36 parameters, also shown in Fig. 2.

The sensitivity to parametrization choice demonstrated in Figs. 2 and 3 shows once again that uncertainty estimates based on the classical Gaussian $\Delta \chi^2 = 1$ estimate, with a fixed choice of parametrization, are overly optimistic—especially since the groups who apply that criterion generally use PDF parametrizations with substantially less flexibility than any of those used here. (In some recent studies, the parametrization dependence has been incorporated in a simple way, by expanding the estimated uncertainty range to include results from each of a variety of specific choices for the parametrization forms \[14\].)

The parametrization effects found here are comparable to, or slightly larger than, the uncertainty estimated by a $\Delta \chi^2 = 10$ analysis. They are not as large as the hypothetical examples whose spectre was raised in Section 2 except perhaps in regions where the uncertainties are already very large. In future analyses, the Chebyshev method may be combined with the Hessian
method, to compute PDF uncertainties that are less subject to parametrization dependence.

Acknowledgments

I thank my TEA (Tung et al.) colleagues J. Huston, H. L. Lai, P. M. Nadolsky, and C.–P. Yuan for discussions of these issues. I thank Stefano Forte and Robert Thorne for helpful correspondence and discussions. This research was supported by National Science Foundation grants PHY-0354838 and PHY-08555561.

Appendix I: General form of $\chi^2$ near its minimum

This Appendix shows that the qualitative behavior of $\chi^2(y,z)$ hypothesized in Fig. 1 arises under rather general assumptions.

Let us assume as usual that $\chi^2$ can be approximated in the neighborhood of its minimum by Taylor series through second order. To find the uncertainty of a particular variable, we can assume by means of a linear transformation of that variable, that we are interested in the value of a parameter $z$ for which $\chi^2 = z^2 + C$ at $y = 0$, as in Eq. (1). Now let $y$ represent an additional fitting parameter, which was previously held fixed at 0. By Taylor series, the expression for $\chi^2$ expands to become

$$\chi^2 = z^2 + y^2 + 2Azy + 2By + C,$$

(8)

where the coefficient of $y^2$ was chosen to be 1 without loss of generality, by scaling that variable. Eq. (8) implies that the contours of constant $\chi^2$ are ellipses whose major and minor axes make an angle of $\pm 45^\circ$ with respect to the $y$ and $z$ axes, as in the specific examples of Fig. 1. The ratio of minor axis to major axis of the ellipse is $\sqrt{(1 - |A|)/(1 + |A|)}$; and $|A| < 1$ is required, since $\chi^2$ must have a minimum. The minimum of $\chi^2$ occurs at

$$z_0 = AB/(1 - A^2), \quad y_0 = -B/(1 - A^2),$$

(9)

and its value there is

$$\chi^2_0 = C - D,$$

(10)

where $D = B^2/(1 - A^2)$. 

12
Relative to the $y = 0$ situation given by Eq. (1), introducing the additional parameter $y$ thus allows the best-fit $\chi^2$ to be lowered by $D$. At the same time, it demands that the uncertainty range for $z$ be extended at least far enough to include $z_0$, and hence demands $\Delta \chi^2 > z_0^2$ in the $y = 0$ model.

The hypothetical examples shown in Fig. 1 correspond to $A = B/5 = -\sqrt{2/3} \implies (z_0 = 10, D = 50)$, and $A = 2B = -\sqrt{20/21} \implies (z_0 = 10, D = 5)$. Whether or not such large shifts occur in typical PDF fitting is the subject of Sect. 4.

**Appendix II: Details of the Chebyshev method**

This Appendix describes a specific method that uses Chebyshev polynomials to parametrize the PDFs in a manner that allows considerable freedom in the parametrization, while requiring the functions to be smooth in a reasonable sense.

Each flavor is parametrized using the form (2) at scale $\mu_0$, with $p(x)$ given by (3) using $y = 1 - 2\sqrt{x}$. The Chebyshev polynomial coefficients $c_j$ are scaled in a manner such as

$$c_j = d_j/j! \quad (11)$$

to define fitting parameters $d_j$ for which $d_j \lesssim O(1)$ represents an appropriate condition of smoothness. This smoothness condition is enforced in the fitting procedure by adding a penalty $k \sum d_j^2$ to $\chi^2$, with the coefficient $k$ chosen to make the total penalty on the order of 1. These penalties are listed in Table I.

[1] J. Pumplin, J. Huston, H. L. Lai, W. K. Tung and C. P. Yuan, Phys. Rev. D 80, 014019 (2009) [arXiv:0904.2424 [hep-ph]].
[2] A. D. Martin, W. J. Stirling, R. S. Thorne and G. Watt, arXiv:0901.0002 [hep-ph].
[3] J. Pumplin, D. R. Stump and W. K. Tung, Phys. Rev. D 65, 014011 (2001) [arXiv:hep-ph/0008191]; J. Pumplin et al., Phys. Rev. D 65, 014013 (2001) [arXiv:hep-ph/011032].
[4] P. M. Nadolsky et al., Phys. Rev. D 78, 013004 (2008) arXiv:0802.0007 [hep-ph].

[5] A. D. Martin, R. G. Roberts, W. J. Stirling and R. S. Thorne, Phys. Lett. B 604, 61 (2004) arXiv:hep-ph/0410230.

[6] A. M. Cooper-Sarkar [ZEUS Collaboration and H1 Collaboration], arXiv:0808.1854 [hep-ph].

[7] S. Alekhin, Phys. Rev. D 68, 014002 (2003) arXiv:hep-ph/0211096; S. Alekhin, JETP Lett. 82, 628 (2005) [Pisma Zh. Eksp. Teor. Fiz. 82, 710 (2005)] arXiv:hep-ph/0508248.

[8] C. Amsler et al. [Particle Data Group], Phys. Lett. B 667, 1 (2008).

[9] J. Pumplin, Phys. Rev. D 80, 034002 (2009) arXiv:0904.2425 [hep-ph].

[10] J. Pumplin, arXiv:0909.0268 [hep-ph].

[11] V. Radescu, A. Glazov, and S. Moch talk at Low-x meeting http://www-d0.fnal.gov/~royon/lowx_italy/.

[12] F. Aaron et al. [H1 Collaboration and ZEUS Collaboration], JHEP 1001, 109 (2010) arXiv:0911.0884 [hep-ex].

[13] R. D. Ball et al. [NNPDF Collaboration], Nucl. Phys. B 809, 1 (2009) arXiv:0808.1231 [hep-ph]; A. Guffanti, J. Rojo and M. Ubiali, “The NNPDF1.2 parton set: implications for the LHC,” arXiv:0907.4614 [hep-ph]; R. D. Ball, L. Del Debbio, S. Forte, A. Guffanti, J. I. Latorre, J. Rojo and M. Ubiali, arXiv:1002.4407 [hep-ph].

[14] Private communication from A. Cooper-Sarkar.