Analytic Fits to Parton Distribution Functions

Sohang. C. Gandhi
Supervisor: Maxim Perelstein
LEPP REU
Laboratory for Elementary Particle Physics
Cornell University
Ithaca, NY 14853

Abstract

Analytic functions fitting parton distribution data are obtained. Methods and examples of incorporating uncertainty associated with the data are discussed.

This project included two main portions. The first and larger portion involved gaining a basic understanding of quantum field theory through readings and exercise. I used the textbook by Zee [1] and covered the following topics: Functional Integration/Wick Contraction, Classical Fields, The Kleine Gordon Equation and Field Quantization, Source Terms and ϕ^4 Interaction, Perturbation Theory/Feynman Diagrams for the Klein Gordon Field and QED, Canonical Quantization, Symmetry/Noethers Theorem, Field Theory in Curved Space-Time, The Dirac Equation and Quantization of the Dirac Field, The Lorentz Group, Dirac and Weyl Spinors and 4-Spinor/Clifford Algebra, The Spin-Statistics Connection, Gauge Invariance in QED, Grassman Algebra and Calculus, Grassman Path Integrals, the Origin of the Inverse Square Law/Atraction and Repulsion in QFT through Effect of Coupling on Vacuum Energy/Exchange of Particles, Regularization/Renormalization/Renormalizability, Physical or Clothed Perturbation Theory, Nonrelativistic QFT, The Origin of the Electron Spin Coupling from the Dirac Equation, Anomalous Magnetic Moment, Symmetry Breaking. I have also worked through the following problems in this book: I.2.1-2, I.3.2, I.4.1, I.7.1-4, I.8.2-4, II.1.1-6, II.1.9,II.2.1, II.3.1,II.3.4,II.5.1-2, II.6.2, II.7.1, III.1.1-2.

The second portion of the project was to obtain analytic fits to numerical parton distribution functions (PDFs) available on the web. Below, I will discuss this segment of the project and present its results.

1 Introduction

Complication arises when analyzing scattering experiments involving protons—such as those being conducted at the Fermilab Tevatron or those planed for the immanent CERN LHC—due to the fact that the proton is a composite particle. This complication is somewhat circumvented, however, in the "deep inelastic" régime where the collisions may be treated as single interactions between effectively free partons (quarks or gluons). In this manner of treating the collisions, it is necessary to have knowledge of how the proton's momentum is distributed amongst its constituents. This information is supplied in the form of the probability that the proton contains a certain constituent (parton) with a certain momentum fraction. (Note that transverse momenta of the partons may be neglected in the "deep inelastic" régime.)

More specifically, for each parton of species f, a function, $f_f(\xi, Q^2)$, called the parton distribution function (PDF), is typically supplied, which gives the probability distribution in momentum fraction, ξ ($\xi = 0 \ldots 1$). Here $Q^2 = -q^2$ is the invariant momentum-squared transfer of the collision with q being the momentum transferred through the interaction. In calculation, f_f is typically evaluated at $\xi = x$, with $x \equiv \frac{Q^2}{2p\cdot q}$, where p is the initial momentum of the proton. Hence ξ and x are used interchangeably and f_f is typically considered a function of x and Q^2. The variation of the functions f_f with Q^2, due to an effect known as Bjorken Scaling, is quite slow, and typically only the rough scale of Q needs to be supplied. Hence, the PDF's are frequently represented as functions of x only [2].

The PDF's must be supplied externally through experimental data since they depend on the soft processes which determine the proton's structure as a bound state of quarks and gluons, and thus cannot be calculated using perturbative QCD [2]. Numerical PDF's are obtained by a rather involved extrapolation procedure of experimental data. Such numerical PDF's are produced by a number of groups—some of the major ones being Alekhin, CTEQ, MRST, GRV, and ZEUS—and are available on the web [3]. One may typically find a number of generations of PDF's and corresponding data files from each group.

It may often be more convenient, for the theorist, to have some analytic representation of the PDFs. To meet this need we have endeavored to obtain analytic functions which are in good agreement with the numerical PDFs described above. We have focused our efforts on fitting CTEQ's most recent PDF set—CTEQ6—in conjunction with the table values, CTEQ6m. We found that the x dependence of the PDFs can be fitted well by polynomials in $\ln x$, while the Q dependence can be fitted by polynomials in either $\ln Q$ or $\ln\ln Q$. Once the functional form is fixed, the problem is reduced to finding the coefficients of the polynomial terms that give the best fit to the numerical PDFs. We have found such coefficients using Mathematica. In the sections that follow the analytic fits obtained are presented and described along with methods and examples of incorporating the uncertainty associated with the original, numerical PDFs.

2 Fits to Central Values

Analytic functions which fit the CTEQ6 numerical PDF with acceptable accuracy — defined below — have been obtained for all of the partons excluding the t and $\bar{t}$ quark, and for most of the region of x and Q for which the numerical PDFs are applicable — $0 < x \leq 1$, 1 GeV$\leq Q \leq 10000$ GeV. The reason for the exclusion of t and $\bar{t}$ is that no numerical PDF is available for these quarks since their appearance as fluctuations in the proton is highly suppressed due to their large mass.

For all of the partons, it was necessary to divide the parameter space into six regions, partitioning the range of x into three portions and the range of Q into two portions. Table 1 below lists these divisions along with functional forms used for the d-quark. The divisions and functional forms used for the other partons varied only slightly from what is listed.

Tables of coefficients — the a_{ij} in table 1 — were produced to represent each fit — one table for each of the six regions and each of the eleven partons. An example of such a coefficient table is given in table 2 below which lists the coefficients for the u-quark in the region $0.0001 \leq x \leq 0.5$, 50 GeV$\leq Q \leq 4000$ GeV, to three significant figures.

Associated with the center values for CTEQ6 are uncertainties, available through an online

Forms for $\mathbf{x * f(x, q^2)}$ for the d-quark	
50 GeV $\leq Q \leq$ 4000 GeV	
$0.0001 \leq x \leq 0.5$	$\sum_{i=0}^{12} \sum_{j=0}^{10} a_{ij} (\ln x)^i (\ln Q)^j$
$0.45 \leq x \leq 0.65$	$\sum_{i=0}^{3} \sum_{j=0}^{3} a_{ij} (\ln x)^i (\ln(\ln Q))^j$
$0.6 \leq x \leq 0.94$	$\sum_{i=0}^{18} \sum_{j=0}^{10} a_{ij} (\ln x)^i (\ln Q)^j$
4000 GeV $\leq Q \leq$ 10000 GeV	
$0.0001 \leq x \leq 0.5$	$\sum_{i=0}^{12} \sum_{j=0}^{6} a_{ij} (\ln x)^i (\ln Q)^j$
$0.45 \leq x \leq 0.65$	$\sum_{i=0}^{3} \sum_{j=0}^{3} a_{ij} (\ln x)^i (\ln(\ln Q))^j$
$0.6 \leq x \leq 0.94$	$\sum_{i=0}^{18} \sum_{j=0}^{6} a_{ij} (\ln x)^i (\ln Q)^j$

Table 1: The regions and functional forms used for analytic fit to the PDF in the case of the d-quark.

$\mathbf{i} \diagdown \mathbf{j}$	0	1	2	3	4
0	-2.42	3.33	-1.98	0.694	-0.156
1	-8.11	11.6	-6.95	2.44	-0.547
2	-6.56	10.6	-6.65	2.39	-0.546
3	4.43	-6.73	4.06	-1.42	0.318
4	7.98	-12.4	7.85	-2.83	0.649
5	-2.68	4.17	-2.53	0.888	-0.199
6	-10.7	16.4	-10.3	3.7	-0.846
7	-8.45	12.8	-8.09	2.92	-0.669
8	-3.44	5.19	-3.27	1.18	-0.271
9	-0.818	1.23	-0.773	0.279	-0.064
10	-0.114	0.17	-0.107	0.0387	-0.00886
11	-0.00857	0.0127	-0.00803	0.0029	-0.000664
12	-0.000268	0.000397	-0.00025	0.0000901	-0.0000206
$\mathbf{i} \diagdown \mathbf{j}$	5	6	7	8	9
0	0.0232	-0.00228	0.000144	-5.25E-6	8.47E-8
1	0.0812	-0.00798	0.0005	-0.0000182	2.91E-7
2	0.0825	-0.00822	0.000522	-0.0000192	3.11E-7
3	-0.0472	0.00464	-0.00029	0.0000105	-1.69E-7
4	-0.0982	0.00982	-0.000625	0.000023	-3.73E-7
5	0.0297	-0.00292	0.000184	-6.71E-6	1.08E-7
6	0.128	-0.0127	0.000809	-0.0000298	4.82E-7
7	0.101	-0.0101	0.000641	-0.0000235	3.81E-7
8	0.041	-0.00409	0.000259	-9.51E-6	1.54E-7
9	0.00967	-0.000964	0.0000611	-2.24E-6	3.61E-8
10	0.00134	-0.000133	8.44E-6	-3.08E-7	4.96E-9
11	0.0001	-9.97E-6	6.3E-7	-2.3E-8	3.70E-10
12	3.11E-6	-3.09E-7	1.96E-8	-7.13E-10	1.14E-11

Table 2: Coefficient table for the u-quark in region $0.0001 \leq x \leq 0.5$, 50 GeV$\leq Q \leq$4000 GeV

calculator provided by the Durham website [3]. Acceptable fits should surely lie within the range defined by the upper and lower bounds implied by the center values and these uncertainties. All of our analytic fits conform to this standard. Further, in the regions specified in table 1, the deviation of our fits from CTEQ6 center is negligible compared to the uncertainty reported for the numerical function. This is illustrated in figure 1, which plots the position of our fit for the u-quark, at $Q = 100$ GeV, within the upper and lower bounds for the numerical function. Table 3 lists the range of deviations for each region of the u-quark fit. It is apparent from the plot that the fit lies very close to center. This property is common to all of our fits in the regions specified.

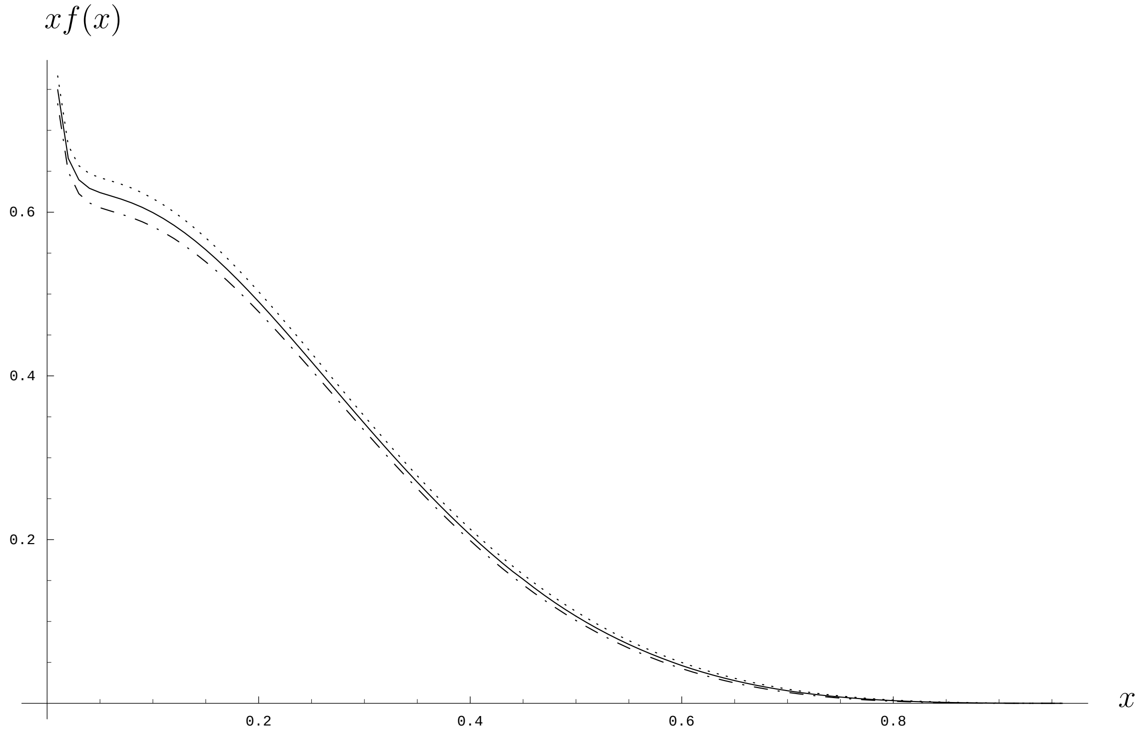

Figure 1: Plot of $xf(x, Q^2)$ vs. x for the u-quark at $q = 100$ GeV. Analytic fit is the solid curve. The upper bound on the numerical value of CTEQ6 is represented by the dotted curve, the lower bound by the dashed curve.

3 The Incorporation of Uncertainty

The values produced by the numerical function for the parton distributions are of course physical numbers with associated error. For our analytic PDF's to be meaningful it is important to develop a means of incorporating this error into our scheme.

Two approaches to doing so are presented here. The first involves reporting the error in the coefficients in the tables described above. Since these coefficient tables are essentially the object of this endeavor, this procedure is the most straightforward. However, as described below, there are some limitations to this representation of error. To address these limitations we have developed a second method of incorporating error using error functions.

Deviations From CTEQ's Center For u-Quark	
$50GeV \le q \le 4000GeV$	
$0 \le x \le 0.5$	2.8E-6% to 0.4%
$0.45 \le x \le 0.65$	3E-5% to 1.2%
$0.6 \le x \le 1$	2E-5% to 2%
$4000GeV \le q \le 10000GeV$	
$0 \le x \le 0.5$	1.5E-8% to 0.45%
$0.45 \le x \le 0.65$	1.4E-6% to 0.6%
$0.6 \le x \le 1$	1E-7% to 3%

Table 3: Deviations for the CTEQ6 center value of our fits for the u-quark.

The uncertainty for CTEQ's PDF's is not currently available in a format which makes practicable the generation of precise values of the uncertainty in the coefficients of two dimensional fits in the case of the first method, and the generation of precise two-dimensional error functions in the second method. Therefore we merely outline the procedure which may be applied to generate complete representations of the error once the uncertainty for CTEQ6 is available in a more convenient form. We also give some examples of coefficient error for some one-dimensional fits in x, and some examples of one-dimensional error functions in x. Finally we give a rough example of a two-dimensional error function which was fitted using only ten sample values in Q, generated manually with the online calculator provided by the Durham site.

3.1 Coefficient Error

Coefficient error may be generated by fitting the upper and lower bounds of the numerical functions to the form appropriate to the region and then taking the difference of the coefficients produced with those corresponding to the central fit:

$$\%\text{Error}_{ij} = 100 * Max\{[(1 + fiterr_{up})a_{ij}^{up} - a_{ij}]/a_{ij},\ [a_{ij} - (1 - fiterr_{down})a_{ij}^{down}]/a_{ij}\}$$

Here, $\%\text{Error}_{ij}$ is the percent-error in the coefficient, a_{ij} and a_{ij}^{up} and a_{ij}^{down} are the coefficients of the fits of the upper and lower bound, respectively. We have also included a factor of $(1 \pm fiterr)$ which takes into account the error associated with the fitting of the upper and lower bounds. $fiterr$ is the estimated maximum error of the fit of the extremum, and thus $(1 \pm fiterr)$ times our fit for the upper and lower bounds, respectively, gives a safe over/under estimate for the bounds. Using these bounds, our coefficient error includes both the uncertainty of the CTEQ numerical data and the error from our fitting process since we are taking the difference of the coefficients with our center fit.

The error in the coefficients of the $(\ln(x))^i$ terms for the u-quark, $0.0001 \le x \le 0.5$, with Q held fixed at 100 GeV are given in table 4. It can be seen that the percent-error is quite reasonable in this example. They are in fact the same size as the maximum numerical error in CTEQ6 on this interval of 5.18%. Table 5 is a similar table for $0.6 \le x \le 0.96$ and Q held at $5000GeV$. This example illustrates a limitation with this approach to representing error. The fact that the percent-error is so large isn't necessarily an issue since the maximum error in CTEQ6 on this interval is 295%, which is roughly the same size. A glance at figure 2 reveals

	Coefficient	% Error
$(\ln(x))^0$	0.101	9.54
$(\ln(x))^1$	0.437	8.91
$(\ln(x))^2$	0.543	3.72
$(\ln(x))^3$	-0.458	4.93
$(\ln(x))^4$	-0.409	5.64
$(\ln(x))^5$	0.312	4.88
$(\ln(x))^6$	0.521	6.12
$(\ln(x))^7$	0.299	6.92
$(\ln(x))^8$	0.0959	7.63
$(\ln(x))^9$	0.0187	8.28
$(\ln(x))^{10}$	0.00222	8.88
$(\ln(x))^{11}$	0.000146	9.42
$(\ln(x))^{12}$	$4.1E-6$	9.88

Table 4: Coefficient error for the u-quark in the region $0.0001 \leq x \leq 0.5$ and $Q = 100$ GeV.

the true defect. This plot shows the percent error in CTEQ6 as a function of x. One may note that the error is reasonable, roughly in the interval up to 0.7. However, there would be no way of deducing this from the coefficient error in table 5; that is coefficient error gives no sense of the local behavior of the error.

3.2 Error Functions

A local measure of error may be provided by error functions. The general procedure involves taking the difference of the lower and upper bounds of CTEQ's numerical data with the data values produced by our analytic fits:

$$\%\text{Error} = 100 * Max\{(\text{upperBound} - \text{fit})/\text{fit},\ (\text{fit} - \text{lowerBound})/\text{fit}\}.$$

Here, "fit" is the value produced by our analytic fit. We then fit this percent-error as a function of x and Q:

$$\text{ErrorFunction}(x, q^2) = (1 + fiterr)\text{ErrorFit}(x, Q).$$

Again, we have included a factor of $(1+fiterr)$, providing a safe overestimate of the error which includes both the error in the original numerical data and the error from our fitting process.

Figure 3 shows a plot of one-dimensional error function in x for the u-quark, $0.0001 \leq x \leq 0.5$, $Q = 100$ GeV. Figure 4 shows the evolution of a rough, two-dimensional error function for the u-quark, fitted using only ten sample Q values, generated manually. It represents the region $0.0001 \leq x \leq 0.5$, 50 Gev$\leq Q \leq 4000$ Gev.

4 Conclusions

We have obtained analytic functions, for each of the partons—excluding t and $\bar{t}$— and for most of the parameter space, which fit the numerical PDF, CTEQ6, to acceptable accuracy. We

	Coefficient	% Error
$(\ln(x))^0$	$8.22E-6$	258.
$(\ln(x))^1$	0.00103	263.
$(\ln(x))^2$	0.0546	271.
$(\ln(x))^3$	1.536	271.
$(\ln(x))^4$	27.0	264.
$(\ln(x))^5$	294.	266.
$(\ln(x))^6$	2210.	262.
$(\ln(x))^7$	11300.	257.
$(\ln(x))^8$	38800.	254.
$(\ln(x))^9$	85100.	251.
$(\ln(x))^{10}$	102000.	249.
$(\ln(x))^{11}$	23500.	248.
$(\ln(x))^{12}$	$-81800.$	248.
$(\ln(x))^{13}$	$-31000.$	252.
$(\ln(x))^{14}$	83400.	249.
$(\ln(x))^{15}$	$-2940.$	175.
$(\ln(x))^{16}$	$-85340.$	252.
$(\ln(x))^{17}$	58800.	248.
$(\ln(x))^{18}$	96200.	255.

Table 5: Coefficient error for the u-quark with $0.6 \leq x \leq 0.96$ and $Q = 5000$ GeV.

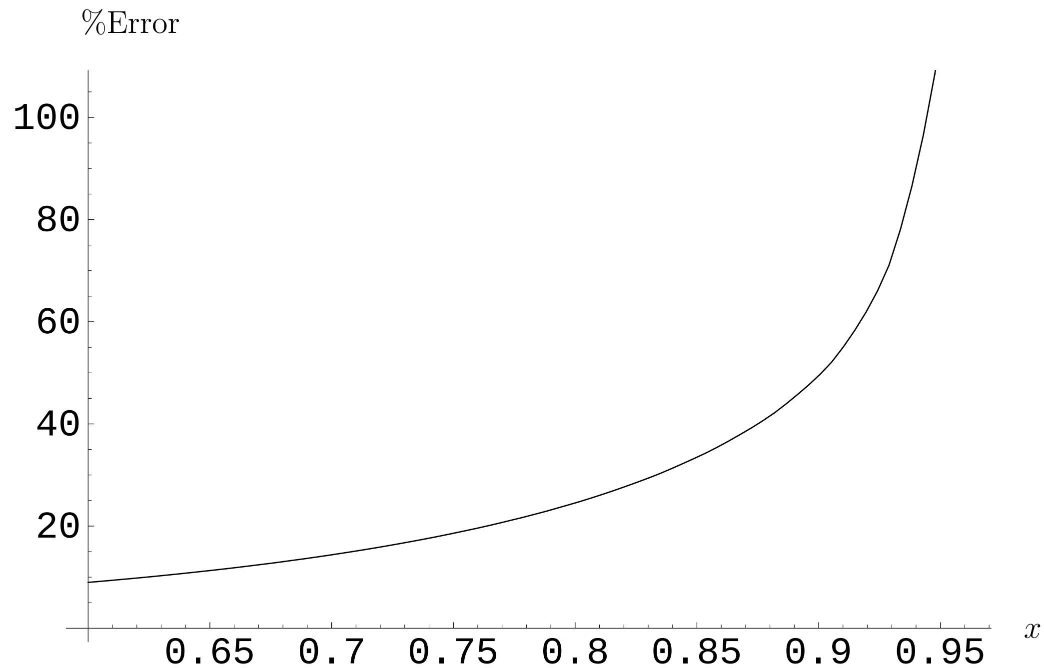

Figure 2: %Error in CTEQ6 versus x for $0.6 \leq x \leq 0.96$, $Q = 5000$ GeV.

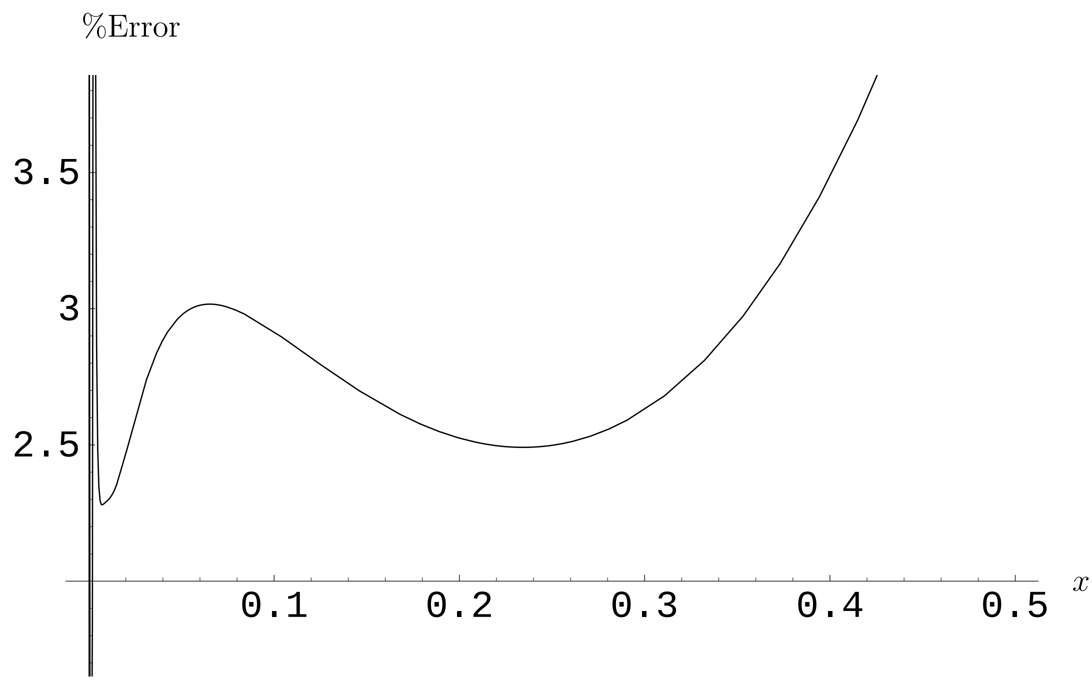

Figure 3: One-dimensional error function for the u-quark with $0.0001 \leq x \leq 0.5$ and $Q =$ 100GeV.

have also developed methods for incorporating uncertainty into our scheme and demonstrated

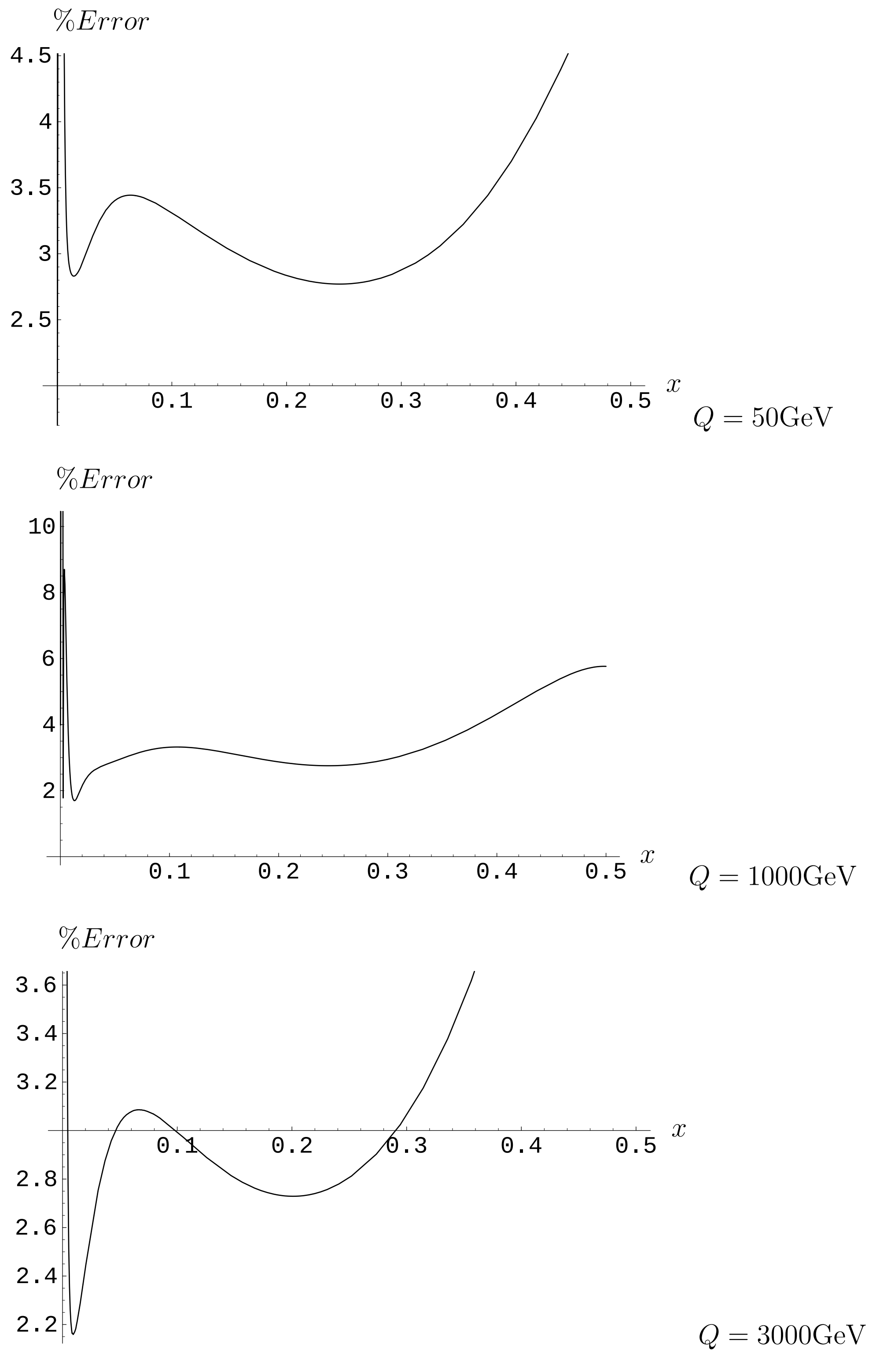

Figure 4: Q evolution for our rough two-dimensional error function for the u-quark. Each plot is for $0.0001 \leq x \leq 0.96$ with Q held fixed.

that they may be applied to our fits. A thorough account of error for all of our fits may be produced once the uncertainty for CTEQ6 is available in a more convenient form.

5 Acknowledgements

This project was funded by the National Science Foundation (grant PHY-0243687), who S.G. would like to thank for its support. S.G. would also like to thank Dr. Maxim Perelstein for his time and guidance during this project and for teaching him so much about particle physics. Finally, S.G. would like to thank Dr. Richard Galik for the tremendous amount of work he put into creating an outstanding REU program for S.G. and his colleagues.

References:

[1] A. Zee, *Quantum Field Theory in a Nutshell*, Princeton 2003
[2] M.E. Peskin and D.V. Schroeder, *An Introduction to Quantum Field Theory*, Westview 1995
[3]Numerical data provided by CTEQ via. Durham website:
http://durpdg.dur.ac.uk/HEPDATA/PDF, maintained by M.R. Whalley
[4]D. Griffiths, *Introduction to Elementary Particles*, Wiley 1987