On the calculation of parton distributions from Lattice QCD

Joseph Matheson Karpie

Fairport, New York

Master of Science, College of William & Mary, 2016
Bachelor of Science, Rensselaer Polytechnic Institute, 2014

A Dissertation presented to the Graduate Faculty
of The College of William & Mary in Candidacy for the Degree of
Doctor of Philosophy

Department of Physics

College of William & Mary
August 2019

©2019
Joseph Matheson Karpie
All rights reserved.

APPROVAL PAGE

This Dissertation is submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy

Joseph Matheson Karpie

Approved by the Committee July 2019

Committee Chair
Professor Konstantinos Orginos, Physics
College of William & Mary

Class of 1962 Professor Carl Carlson, Physics
College of William & Mary

Governor's Distinguished CEBAF Professor Jianwei Qiu, Physics
College of William & Mary

Professor Marc Sher, Physics
College of William & Mary

Professor & Eminent Scholar Anatoly Radyushkin, Physics
Old Dominion University

ABSTRACT

A new method for calculating parton distribution functions from lattice QCD is implemented and studied. Lattice QCD calculable matrix elements with space-like separated fields have an analogous operator product expansion to experimental scattering cross sections and thus these matrix elements are known as “Good Lattice Cross Sections”. Using the collinear factorization approach, a Good Lattice Cross Section can be factorized to the short distance matching kernels that are computed in perturbation theory and the non-perturbative parton distribution functions. As a result, using the perturbative matching kernels and the non-perturbatively computed matrix elements, one can obtain the parton distribution functions. The nucleon and pion matrix elements are determined on a set of 2+1 flavors of clover improved quarks with heavier than physical pion mass. The determination of the parton distributions from Good Lattice Cross Sections constitutes an ill-posed inverse problem. Methods for accurate determination of parton distributions from the Good Lattice Cross Sections are studied. With the calculation of several Good Lattice Cross Sections, the determination of the parton distributions can be improved with a simultaneous analysis, similar to the global parton distribution fits to experimental cross sections.

TABLE OF CONTENTS

Acknowledgments	iv
Dedication	v
List of Tables	vi
List of Figures	vii
CHAPTER	
1 Introduction	2
2 Light cone parton distributions	8
2.1 DIS and Hadronic Tensor	8
2.2 Parton model	11
2.3 Operator definitions	17
2.3.1 DGLAP evolution	20
2.4 Operator Product Expansion	21
3 Lattice QCD	24
3.1 The Path Integral and Statistical Mechanics	25
3.2 Wilson's Actions	27
3.2.1 Symanzik Improvement	29
3.3 Determination of Matrix Elements	31
3.3.1 Momentum smearing	32
3.3.2 Ratio method	35
3.3.3 Feynman-Hellmann summation method	36

4	Lattice Cross Sections	41
4.1	LaMET approach	43
4.2	Lattice Cross Sections and Factorization	46
4.3	OPE without OPE	50
4.4	Wilson line operator and pseudo-PDFs	51
4.4.1	Pseudo-distributions	51
4.5	Two current operators and pseudo-Structure Functions	56
4.5.1	Axial Vector - Vector combination	62
5	Calculation of pseudo-ITD	65
5.1	Method of Calculation	66
5.2	Bare matrix element calculation	67
5.3	Renormalization of power divergence	70
5.4	Reduced pseudo-ITD results	72
5.5	Evolution and Matching	73
5.6	Presence of anti-quark distributions	75
6	Calculation of Ioffe time pseudo-SF	89
6.1	Method of Calculation	89
6.2	Extraction of Matrix elements	93
7	Extraction of PDF	98
7.1	Issue of Fourier Transformation	99
7.2	Calculation of pseudo-moments	103
7.3	Fits of Model Ansatz	105
7.3.1	Nucleon Reduced pseudo-ITD	107
7.3.2	Axial vector-vector current pseudo-SF	109
7.4	Mock Data Tests	113

7.4.1	Backus-Gilbert method	115
7.4.2	Neural Network parameterization	122
7.4.3	Bayesian Reconstruction	129
7.4.4	Restricted χ^2 sampling	139
7.5	Backus Gilbert	141
7.6	Comparison of PDF extractions	144
8	Summary and Outlook	187
	Bibliography	190

ACKNOWLEDGMENTS

There are a great many people who have helped me through my graduate studies. My parents and siblings have encouraged me throughout this process. I owe a deal to all my friends who have supported me celebrated with me in my highs during this time at William and Mary. I also want to thank my “physics big brothers”, Arjun Gambhir, Raza Sufian, and Savvas Zafeiropoulos, for helping me with my eternal struggle against computers and answering the stupid questions I dare not ask anyone else. Without your help, I do not think I would have been able to convince anyone that I know what I am talking about.

I have had a number of mentors all of whom I will try to emulate some aspect of as I develop.

One cannot spend much time in the theory center at Jefferson Lab without hearing of, or more likely just hearing, Anatoly Radyushkin. He is a man of both great intelligence and integrity. He will argue vehemently for the ideas he believes are right, but is quick to accept and adapt when he is proven wrong. I will keep much of his advice with me, especially his correct answer to all hard questions, “I do not know”.

At times the scientific community can seem full of rivalries and grudges, but amongst all this there is David Richards, who appears to have the respect of everyone I have met. He is a kind man who is willing to entertain any notion without prejudice. For as long as I have known him, he has encouraged and supported me. To me, he has been a teacher, a colleague, a fine traveling companion, and a friend.

Finally, I am extremely indebted to Kostas Orginos. I would not be on the path I am if it were not for his guidance. The opportunities he has given me have put me in a position to become a scientist and the wisdom he has imparted has given me the means to do it. More than anyone I have met, he truly loves what he does. It is hard to imagine anyone as excited as Kostas can get when discussing a new idea or studying new data. I hope to be able to mimic this passion for whatever challenges I have ahead.

This work is dedicated to my parents, Sarah and Marty, and to their parents, Betty Ray, Bill, Anne, and Joe, who all struggled to get me where I am today.

LIST OF TABLES

5.1	Parameters for the lattices generated by the JLab/W&M collaboration [1] using 2+1 flavors of clover Wilson fermions and a tree-level tadpole-improved Symanzik gauge action. The lattice spacings, a , are estimated using the Wilson flow scale w_0 . Stout smearing implemented in the fermion action makes the tadpole corrected tree-level clover coefficient c_{SW} used to be very close to the value when determined non-pertubatively with the Schrödinger functional method. The $a127m415L$ contains 10 independent streams of 256 configurations each. The other ensembles only contain a single stream each.	66
6.1	The fit parameters of the V-A matrix elements with $z = (0, 0, a)$ for momenta along the z -direction $p = 0.610$ GeV and $p = 1.525$ GeV, respectively.	95
7.1	The parameterized used in the preconditioning function, with the functional form in Eq. (2.25), for the BG on each ensemble. The results of the fit to this functional form were modified such that the preconditioned PDF will vanish at $x = 0$ and $x = 1$	142

LIST OF FIGURES

2.1	The two topologies of Wick contractions for the hadronic tensor. On the left is the “handbag” diagram where the same quark line is struck by the incoming currents. On the right is the “cat ears” diagram where different quarks are struck by the currents.	13
2.2	The CJ15 PDF fit results [2]. This analysis included 4500 data points from several types of experiments. DIS cross sections, the F_2 structure function, pp and pd Drell-Yan cross sections, jet cross sections, γ +jet cross sections, and W and Z boson asymmetries were all used.	17
3.1	The signal-to-noise ratio for a nucleon two point function at various Euclidean times. The plots on the left and right are from ensembles $a094m390$ and $a094m390L$ respectively which are described in Tab. 5.1. The different choices of momentum smearing parameter, $\zeta = 0$ in blue, $\zeta = 1.5$ in red, and $\zeta = 3.0$ in green, have different ranges of momenta where the signal-to-noise ratio is improved. As designed, the signal-to-noise ratio has a Gaussian shape in momentum space whose peak is shifted by changing the momentum smearing parameter ζ	39
3.2	The effective mass for a nucleon (left) and pion (right) two point functions at various momenta from ensemble $a094m390L$. The signal-to-noise ratio for the effective mass, particularly for the pion, decays without momentum smearing. Even without tuning ζ for each momentum, the momentum smearing procedure significantly improves the signal and allows for precision determination of high momentum effective masses which were not attainable with Gaussian smearing alone.	40
4.1	PDFs, from quasi-PDFs, calculated at physical pion mass. The ETM collaboration [3] (left) and LP ³ collaboration [4] (right) calculated the quasi-PDF with maximum momentum of 1.38 GeV and 3.0 GeV respectively.	45
4.2	The convolutions for evolution and matching	54
4.3	The lowest order Feynman diagrams contributing to the 2 current LCSs in Eq. (4.31) for an asymptotic on-shell quark state of momentum $k = xp$	58

5.1	Typical fits of the reduced Ioffe time pseudo-distribution from the ensemble $a127m415L$. The left and right plots correspond to real and imaginary component of the bare matrix element respectively. The ranges of T shown is the range of data used in the fit. The points and bands over the same color correspond to a correlation fitted to Eq (5.3) and the resulting fit respectively. The grey band corresponds to the bare matrix element result.	77
5.2	Typical fits of the reduced Ioffe time pseudo-distribution from the ensemble $a127m415$. The left and right plots correspond to real and imaginary component of the bare matrix element respectively. The ranges of T shown is the range of data used in the fit. The points and bands over the same color correspond to a correlation fitted to Eq (5.3) and the resulting fit respectively. The grey band corresponds to the bare matrix element result.	78
5.3	Typical fits of the reduced Ioffe time pseudo-distribution from the ensemble $a094m390$. The left and right plots correspond to real and imaginary component of the bare matrix element respectively. The ranges of T shown is the range of data used in the fit. The points and bands over the same color correspond to a correlation fitted to Eq (5.3) and the resulting fit respectively. The grey band corresponds to the bare matrix element result.	79
5.4	The rest frame distribution $\mathcal{M}^0(z \cdot p, z^2) _{p=0}$. The top left plot is from the ensemble $a127m415$, the bottom left plot is from the ensemble $a127m415L$, and the top right plot is from the ensemble $a094m390$. The cusp as $z \rightarrow 0$ is the signal for the power divergences which occur in perturbation theory. The large z limit reveals the exponential nature of the Wilson line renormalization constant.	80
5.5	The reduced Ioffe time pseudo-distribution as a function of p . Here the momenta and displacement are labeled in lattice units, $p_3 = 2\pi p/L$ and $z_3 = za$. The left plots are the real component and the right are the imaginary component. The top plots are from the ensemble $a127m415$, the middle plots are from the ensemble $a127m415L$, and the bottom plots are from the ensemble $a094m390$.	81
5.6	The reduced Ioffe time pseudo-distribution as a function of z . Here the momenta and displacement are labeled in lattice units, $p_3 = 2\pi p/L$ and $z_3 = za$. The left plots are the real component and the right are the imaginary component. The top plots are from the ensemble $a127m415$, the middle plots are from the ensemble $a127m415L$, and the bottom plots are from the ensemble $a094m390$.	82
5.7	The real and imaginary components of the reduced Ioffe time pseudo-distribution on the ensemble $a127m415$ as a function of $\nu = z \cdot p$.	83
5.8	The real and imaginary components of the reduced Ioffe time pseudo-distribution on the ensemble $a127m415L$ as a function of $\nu = z \cdot p$.	83
5.9	The real and imaginary components of the reduced Ioffe time pseudo-distribution on the ensemble $a094m390$ as a function of $\nu = z \cdot p$.	83

5.10	The convolutions required for the evolution and matching of the reduced pseudo-ITD to the $\overline{\text{MS}}$ ITD on the ensemble $a127m415$. The reduced pseudo-ITD was interpolated by fitting a polynomial.	84
5.11	The convolutions required for the evolution and matching of the reduced pseudo-ITD to the $\overline{\text{MS}}$ ITD on the ensemble $a127m415L$. The reduced pseudo-ITD was interpolated by fitting a polynomial.	85
5.12	The convolutions required for the evolution and matching of the reduced pseudo-ITD to the $\overline{\text{MS}}$ ITD on the ensemble $a094m390$. The reduced pseudo-ITD was interpolated by fitting a polynomial.	86
5.13	The evolution and matching of the pseudo-ITD to the ITD. The left and right plots show the real and imaginary components respectively. The top plots are from ensemble $a127m415$ and the lower plots are from ensemble $a127m415L$	87
5.14	Estimating the size of the antiquark distribution. The left and right columns are from ensemble $a127m415$ and $a127m415L$ respectively. The top plots are the real component, middle plots are the imaginary component, and the bottoms plots are the sine transform of the antiquark distribution. The trend for ensemble $a127m415$ uses $a = -0.2$ and $b = 2.5$. The trend for ensemble $a127m415L$ uses $a = -0.1$ and $b = 2.75$	88
6.1	The four point-function is constructed by combining a heavy quark propagator, represented by the green line, and a modified sequential source, represented by the black lines. An advantage of this setup is that any pair of currents with any separation may be constructed without additional costly propagator inversions.	93
6.2	Jackknife ensemble ratio data of the 4pt to 2pt correlation functions used in the extraction of antisymmetric V-A current matrix elements. Figs 6.2a and 6.2b show fits to the matrix elements for $p = 0.610$ GeV with spatial separation between the currents $z = 1a$ and $z = 4a$, respectively. The blue data points are obtained from the lattice QCD calculation, the green band shows the two-state fit to the data, and the red band shows the extracted value of the matrix element in the asymptotically large source-sink separation limit.	96
6.3	Jackknife ensemble ratio data of the 4pt to 2pt correlation functions used in the extraction of antisymmetric V-A current matrix elements. Figs 6.3a and 6.3b show fits to the matrix elements for $p = 1.525$ GeV with spatial separation between the currents $z = 1a$ and $z = 4a$, respectively. The green data points are obtained from the lattice QCD calculation, the gray band shows the two-state fit to the data, and the red band shows the extracted value of the matrix element in the asymptotically large source-sink separation limit.	96

6.4	The pseudo-ITSF of the axial vector-vector current combination.	97
7.1	The eigenvalues of the coefficient matrix from direct inversion (right) and derivative method inversion (left). In both cases, the ill-posed nature of the problem suppresses the eigenvalues exponentially except for the final range which is a genuine discretized Fourier transform.	101
7.2	The moments of the pseudo-PDF compared to phenomenologically determined PDF moments from the NLO global fit CJ15nlo [2], and the NNLO global fits MSTW2008nnlo68cl_nf4 [5] and NNPDF31_nnlo_pch_as_0118_mc_164 [6] all evolved to 4 GeV. The top, middle, and bottom plots are from the ensembles $a127m415$, $a127m415L$ and $a094m390$ respectively. The left and right columns show the first and second moments respectively. Only the lowest two moments have signal for most z . The higher moments only have signal for the largest z where the maximum Ioffe time used allows for identifying more than the leading behavior in the Taylor expansion.	146
7.3	The nucleon valence distribution obtained from the ensemble $a127m415$ using the fit form in Eq. (2.25). The upper and lower plots are from including all the data in the fit or only using data originating with $z < 4a$ respectively. The $\chi^2/\text{d.o.f.}$ for the fit with all the data is 2.1(7). The $\chi^2/\text{d.o.f.}$ for the fit with points with $z < 4a$ is 5(1). The uncertainty band is obtained from the fits to the Jackknife samples of the data.	147
7.4	The nucleon valence distribution obtained from the ensemble $a127m415L$ using the fit form in Eq. (2.25). The upper and lower plots are from including all the data in the fit or only using data originating with $z < 4a$ respectively. The $\chi^2/\text{d.o.f.}$ for the fit with all the data is 1.5(5). The $\chi^2/\text{d.o.f.}$ for the fit with points with $z < 4a$ is 1.8(8). The uncertainty band is obtained from the fits to the Jackknife samples of the data.	148
7.5	The nucleon valence distribution obtained from the ensemble $a094m390$ using the fit form in Eq. (2.25). The upper and lower plots are from including all the data in the fit or only using data originating with $z < 4a$ respectively. The $\chi^2/\text{d.o.f.}$ for the fit with all the data is 1.8(5). The $\chi^2/\text{d.o.f.}$ for the fit with points with $z < 4a$ is 3.0(9). The uncertainty band is obtained from the fits to the Jackknife samples of the data.	149
7.6	The nucleon valence distribution obtained from the ensemble $a127m415$ fit to the form used by the JAM collaboration in Eq. (7.10). The upper and lower plots are from including all the data in the fit or only using data originating with $z < 4a$ respectively. The $\chi^2/\text{d.o.f.}$ for the fit with all the data is 2.1(8). The $\chi^2/\text{d.o.f.}$ for the fit with points with $z < 4a$ is 6(1). The uncertainty band is obtained from the fits to the Jackknife samples of the data.	150

7.7	The nucleon valence distribution obtained from the ensemble <i>a127m415L</i> fit to the form used by the JAM collaboration in Eq. (7.10). The upper and lower plots are from including all the data in the fit or only using data originating with $z < 4a$ respectively. The $\chi^2/\text{d.o.f.}$ for the fit with all the data is 1.6(5). The $\chi^2/\text{d.o.f.}$ for the fit with points with $z < 4a$ is 2.2(9). The uncertainty band is obtained from the fits to the Jackknife samples of the data.	151
7.8	The nucleon valence distribution obtained from the ensemble <i>a094m390</i> fit to the form used by the JAM collaboration in Eq. (7.10). The upper and lower plots are from including all the data in the fit or only using data originating with $z < 4a$ respectively. The $\chi^2/\text{d.o.f.}$ for the fit with all the data is 1.9(5). The $\chi^2/\text{d.o.f.}$ for the fit with points with $z < 4a$ is 3.4(1.2). The uncertainty band is obtained from the fits to the Jackknife samples of the data.	152
7.9	The nucleon valence distribution obtained from the ensemble <i>a127m415</i> fit to the form using the polynomial in Eq. (7.14). The upper and lower plots are from including all the data in the fit or only using data originating with $z < 4a$ respectively. The $\chi^2/\text{d.o.f.}$ for the fit with all the data is 2.3(9). The $\chi^2/\text{d.o.f.}$ for the fit with points with $z < 4a$ is 7(2). The uncertainty band is obtained from the fits to the Jackknife samples of the data.	153
7.10	The nucleon valence distribution obtained from the ensemble <i>a127m415L</i> fit to the form using the polynomial in Eq. (7.14). The upper and lower plots are from including all the data in the fit or only using data originating with $z < 4a$ respectively. The $\chi^2/\text{d.o.f.}$ for the fit with all the data is 1.7(6). The $\chi^2/\text{d.o.f.}$ for the fit with points with $z < 4a$ is 2.7(9). The uncertainty band is obtained from the fits to the Jackknife samples of the data.	154
7.11	The nucleon valence distribution obtained from the ensemble <i>a094m390</i> fit to the form using the polynomial in Eq. (7.14). The upper and lower plots are from including all the data in the fit or only using data originating with $z < 4a$ respectively. The $\chi^2/\text{d.o.f.}$ for the fit with all the data is 2.0(1.2). The $\chi^2/\text{d.o.f.}$ for the fit with points with $z < 4a$ is 4.2(1.3). The uncertainty band is obtained from the fits to the Jackknife samples of the data.	155
7.12	The nucleon valence distribution obtained from the ensemble <i>a127m415</i> fit to the form using the polynomial in Eq. (7.16). The upper and lower plots are from including all the data in the fit or only using data originating with $z < 4a$ respectively. The $\chi^2/\text{d.o.f.}$ for the fit with all the data is 2.3(9). The $\chi^2/\text{d.o.f.}$ for the fit with points with $z < 4a$ is 7(2). The uncertainty band is obtained from the fits to the Jackknife samples of the data.	156

7.13	The nucleon valence distribution obtained from the ensemble <i>a127m415L</i> fit to the form using the polynomial in Eq. (7.16). The upper and lower plots are from including all the data in the fit or only using data originating with $z < 4a$ respectively. The $\chi^2/\text{d.o.f.}$ for the fit with all the data is 1.4(5). The $\chi^2/\text{d.o.f.}$ for the fit with points with $z < 4a$ is 3(1). The uncertainty band is obtained from the fits to the Jackknife samples of the data.	157
7.14	The nucleon valence distribution obtained from the ensemble <i>a094m390</i> fit to the form using the polynomial in Eq. (7.16). The upper and lower plots are from including all the data in the fit or only using data originating with $z < 4a$ respectively. The $\chi^2/\text{d.o.f.}$ for the fit with all the data is 4.9(8). The $\chi^2/\text{d.o.f.}$ for the fit with points with $z < 4a$ is 7.5(7.0). The uncertainty band is obtained from the fits to the Jackknife samples of the data.	158
7.15	The pseudo-ITSF of the axial vector-vector current combination fit to the PDF functional form in Eq. (7.18) using the leading order kernel in Eq. (4.48). The uncertainty band is obtained from the fits to the Jackknife samples of the data.	158
7.16	The pion valence distribution obtained using the LO perturbative kernel in Eq. (4.48) derived in Sec. 4.5.1 and the functional form of the PDF in Eq. (7.18). Figure (7.16a) shows the pion valence distribution $q_v^\pi(x)$ and Figure (7.16b) shows the $xq_v^\pi(x)$ -distribution. The uncertainty band is obtained from the fits to the Jackknife samples of the data.	159
7.17	Comparison of pion $xq_v^\pi(x)$ -distribution with the leading-order (LO) extraction from Drell-Yan data [7] (gray data points with uncertainties), next-to-leading order (NLO) fits [8, 9, 10] (orange band, magenta curve, and red band), and model calculations [11, 12] (black and blue lines). This lattice QCD calculation of $q_v^\pi(x)$ is evolved from an initial scale $\mu_0^2 = 1 \text{ GeV}^2$ at LO. All the results are at evolved to an evolution scale of $\mu^2 = 27 \text{ GeV}^2$	159
7.18	The two mock datasets used for testing the methods. The PDF on the left. The Ioffe time PDF on the right. Scenario A (red band) will test the case of a PDF which diverges at small x and scenario B (blue band) will test the case of a PDF which converges at small x	160
7.19	Best fit PDF (red solid line) from (left) realistic PDF data [mock scenario A] and (right) a modified scenario with the PDF vanishing at the origin [mock scenario B]. The actual mock PDF in the former cases is given as gray solid line. To determine the dependence of our results on the choice of default model, three further choices for m are plotted, two arising from varying the best fit parameters by factors of 2, one being the constant default model $m=1$	160

7.20	x -space PDF's reconstructed using the Backus-Gilbert (BG) method from $N_\nu = 12$ Ioffe-time data points on the interval $\nu = [0, 10]$ (top) as well as the input data (gray crosses) compared to the data arising from the reconstructed PDF (red crosses) in the bottom panels. The plots in the left column show the results for mock data based on a phenomenological PDF, while the right column from the modified scenario where the PDF vanishes at the origin. In both scenarios no preconditioning has been employed. . . .	161
7.21	x -space PDF's reconstructed using the Backus-Gilbert (BG) method from $N_\nu = 12$ Ioffe-time data points on the interval $\nu = [0, 10]$ (top) as well as the input data (gray crosses) compared to the data arising from the reconstructed PDF (red crosses) in the bottom panels. The plots in the left column show the results for mock data based on a phenomenological PDF, while the right column from the modified scenario where the PDF vanishes at the origin. The reconstruction was performed with preconditioning exponents $a = -0.35$ and $b = 2$ for scenario A and $a = 0.3$ and $b = 2$ for scenario B.	162
7.22	x -space PDF's reconstructed using the Backus-Gilbert (BG) method from $N_\nu = 23$ Ioffe-time data points on the interval $\nu = [0, 20]$ (top) and from $N_\nu = 112$ Ioffe-time data points on the interval $\nu = [0, 100]$ (bottom). The plots in the left column show the results for mock data based on a phenomenological PDF, while the right column from the modified scenario where the PDF vanishes at the origin. The reconstruction was performed with preconditioning exponents $a = -0.35$ and $b = 2$ for scenario A and $a = 0.3$ and $b = 2$ for scenario B. It is noteworthy that by having lattice data up to $\nu_{\max} = 20$ one can achieve a very good reconstruction with the BG method.	163
7.23	A neural network can be used as a general parametrization of an unknown function from $\mathbb{R}^{N_1} \rightarrow \mathbb{R}^{N_L}$. For the case of a PDF, a single valued function of a single argument, the input and output layers have only one neuron. . .	164
7.24	The genetically trained neural nets. The blue band is the original data. The curves are the responses for the final population of genetically trained neural networks. The left column is with NNPDF data. The right column is with modified data. The first row has a network geometry of 1-3-1. The second row has a geometry of 1-4-1. The third row has a geometry of 1-2-2-1.	165
7.25	The genetically trained neural nets. The blue band is the original data. The red band is the reconstructed PDF with statistical and systematic errors. The left column is with NNPDF data. The right column is with modified data. The first row has a network geometry of 1-3-1. The second row has a geometry of 1-4-1. The third row has a geometry of 1-2-2-1.	166

7.26	The genetically trained neural nets. The grey points are the original data while the red points are the ones reconstructed by the NN. The left column is with NNPDF data. The right column is with modified data. The first row has a network geometry of 1-3-1. The second row has a geometry of 1-4-1. The third row has a geometry of 1-2-2-1.	167
7.27	The PDF reconstructions from preconditioned Neural Networks (top). The grey points are the original data while the red points are the ones reconstructed by the NN (bottom). The left is scenario A and the right is scenario B.	168
7.28	x -space PDF's reconstructed using the Maximum Entropy method (MEM) from $N_\nu = 12$ Ioffe-time data points on the interval $\nu = [0, 10]$ (top) as well as the input data (gray crosses) compared to the data arising from the reconstructed PDF (red crosses) in the bottom panels. The plots in the left column denote the results for mock data based on a phenomenological PDF, while the right column from the modified scenario where the PDF vanishes at the origin. (top) The original mock PDF is given as gray dashed line and the result based on the BR method with the best fit default model as solid red line. Statistical uncertainty is visualized by a red, systematic uncertainty from the choice of default model by a blue errorband respectively. (bottom) The red solid line corresponds to the relative deviation of the reconstruction from the input mock PDF, the red errorband encodes the combined uncertainty of the BR reconstruction, the gray errorband the uncertainty of the original mock PDF.	169
7.29	x -space PDF's reconstructed using the Bayesian reconstruction (BR) method from $N_\nu = 12$ Ioffe-time data points on the interval $\nu = [0, 10]$ (top) as well as the input data (gray crosses) compared to the data arising from the reconstructed PDF (red crosses) in the bottom panels.	170
7.30	x -space PDF's reconstructed using the Maximum Entropy Method (top) as well as the Bayesian reconstruction (BR) method (bottom) from $N_\nu = 12$ Ioffe-time data points on the interval $\nu = [0, 20]$	171
7.31	x -space PDF's reconstructed using a quadratic prior Bayesian (QDR) method from $N_\nu = 12$ Ioffe-time data points on the interval $\nu = [0, 10]$ (top) as well as the input data (gray crosses) compared to the data arising from the reconstructed PDF (red crosses) in the bottom panels.	172
7.32	x -space PDF's reconstructed using the generalized Bayesian reconstruction (BRg) method from $N_\nu = 12$ Ioffe-time data points on the interval $\nu = [0, 10]$ (top) as well as the input data (gray crosses) compared to the data arising from the reconstructed PDF (red crosses) in the bottom panels. . .	173

7.33	Reconstructed PDF (red dashed) from restricted sampling of the χ^2 functional based on $N_\nu = 12$ Ioffe-time data points on the interval $\nu = [0, 10]$. Mock scenario A is shown on the left, scenario B on the right. While positivity alone already provides a powerful regularization, for this realistic scenario we do observe deviations from the correct result, especially for scenario B.	174
7.34	Reconstructed PDF (red dashed) from restricted sampling of the χ^2 functional based on $N_\nu = 100$ Ioffe-time data points on the interval $\nu = [0, 100]$. Mock scenario A is shown on the left, scenario B on the right. In this almost ideal setting, positivity alone provides a regularization that is sufficient for a reasonable determination of the PDF of scenario A, while it is unable to suppress ringing artifacts in scenario B.	174
7.35	Reconstructed PDF from covariance regulated Backus Gilbert for ensemble <i>a127m415</i>	175
7.36	Reconstructed PDF from covariance regulated Backus Gilbert for ensemble <i>a127m415</i>	176
7.37	Reconstructed PDF from covariance regulated Backus Gilbert for ensemble <i>a127m415L</i>	177
7.38	Reconstructed PDF from covariance regulated Backus Gilbert for ensemble <i>a127m415L</i>	178
7.39	Reconstructed PDF from covariance regulated Backus Gilbert for ensemble <i>a094m390</i>	179
7.40	Reconstructed PDF from covariance regulated Backus Gilbert for ensemble <i>a094m390</i>	180
7.41	Reconstructed PDF from Tikhonov regulated Backus Gilbert for ensemble <i>a127m415</i>	181
7.42	Reconstructed PDF from Tikhonov regulated Backus Gilbert for ensemble <i>a127m415L</i>	182
7.43	Reconstructed PDF from Tikhonov regulated Backus Gilbert for ensemble <i>a094m390</i>	183
7.44	Widths σ from covariance regulated Backus Gilbert inversions. The top left and top right plots are from ensembles <i>a127m415</i> and <i>a127m415L</i> respectively and the lower plot is from ensemble <i>a094m390</i>	184
7.45	Widths σ from Tikhonov regulated Backus Gilbert inversions. The top left and top right plots are from ensembles <i>a127m415</i> and <i>a127m415L</i> respectively and the lower plot is from ensemble <i>a094m390</i>	185
7.46	Comparisons of the PDF extractions. The top, middle and bottom plots are from ensembles <i>a127m415</i> , <i>a127m415L</i> , and <i>a094m390</i> respectively. The plots on the right column have been normalized by taking the central value from the fit to the functional form in Eq. (7.10).	186

ON THE CALCULATION OF PARTON DISTRIBUTIONS FROM LATTICE QCD

CHAPTER 1

Introduction

Almost all observed phenomena in physics can, in principle, be explained by the Standard Model (SM) of particle physics. In the SM, matter is described by fermions which interact through the exchanges of bosons. The SM is defined by the groups describing the local gauge symmetries in the model and what representation of those groups are used for each of the quantum fields. There are three gauge symmetries, $U(1)$, $SU(2)_L$, and $SU(3)$, which govern the Electromagnetic, Weak, and Strong interactions respectively. There exist a few phenomena, such as dark matter and gravitation, and a few aesthetic problems, such as the Weak hierarchy problem and the Strong CP problem, do not have an experimentally verified quantum descriptions. Current explanations of these phenomena go by the name of Beyond the Standard Model (BSM) physics.

The theory describing the Electromagnetic force, Quantum ElectroDynamics (QED), governed by a $U(1)$ local gauge symmetry is used as a framework for the development of the other theories. The Lagrangian of QED is given by

$$\mathcal{L}_{\text{QED}}(x) = \sum_f \bar{\psi}_f(x)(i\not{D} - m_f)\psi_f(x) - \frac{1}{4}F^{\mu\nu}(x)F_{\mu\nu}(x), \quad (1.1)$$

where f labels the various charged fermions, $D_\mu = \partial_\mu - igA_\mu$ is the gauge covariant derivative, and $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$ is the field strength tensor. Unless otherwise stated, it is assumed repeated indices are summed over. In QED, the coupling g is small and the Feynman path integral can be approximated by a truncated perturbative expansion in this coupling. Calculations can be systematically improved by including more terms in this expansion, at the cost of requiring more complicated integrations. This procedure of steadily improving calculations has been very successful in describes the interactions of light and electrons.

The Weak force, responsible for flavor changing interactions, such as nuclear beta decay, is governed by a $SU(2)_L$ symmetry. The force is mediated by the exchange of massive gauge bosons, $W^\pm$ and Z^0 . Due to the heavy mass of the gauge bosons, there exists a long range effective theory which can describe weak interactions. The Feynman path integral is expanded in powers of $\frac{p^2}{M^2}$. Just as in QED, the calculations can be systematically improved by including more terms in the expansion.

Quantum ChromoDynamics (QCD) describes the Strong interaction in terms of the matter fermions, called quarks, and the massless force bosons, called gluons. The QCD Lagrangian is given by

$$\mathcal{L}_{\text{QCD}}(x) = \sum_q \bar{\psi}_q^a(x)(i\not{D}^{ab} - m_q\delta^{ab})\psi_f^b(x) - \frac{1}{4} [G_a^{\mu\nu}(x)G_{\mu\nu}^a(x)] , \quad (1.2)$$

where q labels the quarks, $D_\mu^{ab} = \partial_\mu\delta^{ab} - igt_{ab}^c A_\mu^c$ is the $SU(3)$ gauge covariant derivative, $G_a^{\mu\nu}(x) = \partial^\mu A_a^\nu - \partial^\nu A_a^\mu - gf^{abc}A_b^\mu A_c^\nu$ is the $SU(3)$ field strength tensor, and t^c are the generating matrices of $SU(3)$. This Lagrangian is very similar to the QED Lagrangian, except the interactions with gluon fields, A_μ^a , are described by $SU(3)$ matrices instead of $U(1)$ complex numbers. This slight modification creates drastic differences in the dynamics of the theory. For example, gluons, unlike photons, can directly interact with each other.

QCD, unlike the ElectroWeak interactions, does not have a satisfactory perturbative expansion for all energy ranges of interest to a particle physicist. Only at high energy is QCD “asymptotically free” and the perturbative truncation techniques developed for QED can be applied. For interactions with low energies, the strong coupling is far too large for a perturbative expansion to converge and can even become greater than 1. In other words, low energy events with large number of interactions between the quantum fields are favored over events with only a few interactions. The interactions of quarks and gluons is an example of a strongly correlated system, which are notoriously difficult to describe analytically, but also come with a wealth of phenomena to study. In condensed matter physics, strongly correlated systems can have unusual and interesting properties, such as high temperature superconductivity.

The evaluation of Feynman path integrals for the SM is complicated by a plethora of infinities which must be removed before the theories can be used for precise predictions. In a perturbative expansion, the leading order (LO) contribution will typically come from a tree level diagram, in which all momenta are fixed by the conservation of momentum, and its evaluation results in a finite number. The next-to-leading order (NLO) contribution will have diagrams which contain a closed loop. The fields in the loop have unspecified momenta whose value will be integrated over. In theories like the SM, these integrals will result in infinity. In order to reach a finite answer, the theory must be redefined in such a way, called regularization, that the necessary integrals will be finite up to the limit which reproduces the original problem. One of the more popular choices is dimensional regularization where the number of space-time dimensions is varied and the infinities appear when the number of dimensions goes to four. Another choice, which is particularly useful for QCD, is lattice regularization [13] where the spacetime locations of fields are restricted to be on the sites or links between sites of a four dimensional lattice. With lattice regularization, the infinities appear in the continuum limit, in which the spacing

between sites goes to zero. The divergences in these integrals are then subtracted from the theory by redefining its parameters, such as the masses and charges, to include terms which exactly cancel them, a process called renormalization. The different choices of how to perform renormalization are called renormalization schemes.

The strong interaction provides a unique problem for SM quantum field theories. Unlike fields with only Electromagnetic or Weak interactions, quarks and gluons are not experimentally observable, a feature called “confinement”. Instead, experiments observe hadrons, which are composite states of quarks and gluons. Confinement can complicate the use of perturbative QCD techniques for analyzing experimental results. QCD with a lattice regularization, called Lattice QCD, can be performed in a perturbative approach to describe quark and gluon interactions for comparison with other methods or it can be performed numerically in order to calculate hadronic properties. The numerical implementation of Lattice QCD is the only known technique capable of calculating hadron observables directly from QCD with controllable systematic errors.

Understanding how quarks and gluons, collectively referred to as partons, combine to create hadrons is an active field of study called hadron structure. Lattice QCD has been very successful in calculation of bulk hadronic properties such as masses and charges. Determining the position and momentum distributions of partons inside of a hadron remains an important theoretical goal in hadron structure. The nucleon unpolarized Parton Distribution Function (PDF) is the most precisely known partonic distribution. The PDF describes the parton’s momentum distribution in the direction of a hadron’s motion. There also exist more complicated functions which describe the distribution of spin asymmetries, transverse momentum, and impact parameter. The Wigner distributions [14] are the most general of such objects.

Until recently, these partonic distributions were only accessible through analyzing experimental results. Their definitions are in terms of hadronic matrix elements, which

could not be calculated with perturbative techniques. Parton distributions also could not be calculated in numerical Lattice QCD, because they are defined by matrix elements with light-like separated fields, which is impossible to have on the required Euclidean lattice. Instead, experimental cross sections are interpreted as a convolution of a perturbatively calculable, shorter distance, process dependent Wilson Coefficient and a non-perturbative, long distance, universal parton distribution. The separation of the cross section into these two distance scales is referred to as factorization. The cross sections of many different types of experiments are analyzed simultaneously to fit the nucleon PDFs. This procedure has allowed for very precise phenomenological determination of the unpolarized PDF by a number of groups.

In recent years, a number of proposals have been made of using Lattice QCD calculations in order to measure related matrix elements, from which quark and gluon distributions can be calculated, but no direct calculation is currently possible. In analogy with the phenomenological extraction, certain hadronic matrix elements, or their Fourier transforms, can be interpreted as a perturbatively calculable, matrix element dependent matching kernel and the universal PDF.

Almost every experiment attempting to discover BSM physics or testing precision SM results require interactions with hadrons, but every BSM theory describes interactions with quarks. A precise understanding of how the partons are distributed within hadrons is required to explain the hadronic events in terms of the quark interactions. In certain cases, the theoretical errors associated with these quark and gluon distributions are a significant of error for interpreting these experiments in terms of SM and BSM theories.

The remainder of this is organized as following. In Ch. 2, the definitions of parton distributions will be given in terms of light-like separated matrix elements. In Ch. 3, the method of Lattice QCD will be highlighted. In Ch. 4, a class of Lattice QCD calculable matrix elements will be related to the PDF. In Ch. 5 and Ch. 6, two of these types of

matrix elements are calculated. In Ch. 7, the methods of determining the PDF from these matrix elements are studied.

CHAPTER 2

Light cone parton distributions

The modern theoretical understanding structure of hadron relies on studying the available experimental cross sections. Deep inelastic scattering (DIS) of leptons was crucial in the development of parton theory and the discovery of QCD. It was proposed by Bjorken [15] that the cross sections of electron-nucleon DIS would exhibit a scaling behavior, where it is independent of the momentum transferred between the electron and nucleon. This behavior was in contradiction to the expectations of the cross section for elastic scattering off a finite sized particle, which falls off rapidly as a function of the momentum transfer. The Nobel prize winning SLAC-MIT experiment discovered an approximate scaling behavior in the DIS cross section. Theories which are asymptotically free would have the approximate scaling behaviors in the limit of large momentum transfer, which is what began the interest in and development of QCD.

2.1 DIS and Hadronic Tensor

In a DIS experiment, the lepton and hadron enter with known momenta l^μ and p^μ . After the interaction, the lepton leaves with a measured momentum l'^μ . The remnant of

the hadron is not observed, and one considers all possible outgoing hadronic states X . At leading order in QED, the lepton emits a single virtual photon with momentum $q^\mu = l^\mu - l'^\mu$. For the remainder of this chapter, all Electromagnetic and Weak interactions will be taken to be at this leading order. There exist two Lorentz invariants important for the hadronic part of the system, the virtuality of the photon $Q^2 = -q^2$ and the Bjorken scaling variable $x = \frac{Q^2}{2p \cdot q}$. The full system has the typical Mandelstam variables, $s = (p + l)^2$, $t = q^2$, and $u = (l' - p)^2$. The DIS limit is where both Q^2 and the mass of the outgoing hadronic states $m_X^2 = (p + q)^2 = m_h^2 + \frac{Q^2}{x}(1 - x)$, where m_h is the mass of the target hadron, are both very large compared to typical hadronic scales, characterized by Λ_{QCD} .

The inclusive DIS cross section can be separated into leptonic and hadronic matrix elements as

$$E' \frac{d\sigma}{d^4l'} \approx \frac{\pi e^4}{2s} \sum_X \delta^{(4)}(p_X - p - q) |\langle l' | j^\mu | l \rangle \frac{1}{q^2} \langle X | j_\mu | p \rangle|^2 = \frac{2\alpha^2}{sQ^4} L_{\mu\nu} H^{\mu\nu} \quad (2.1)$$

where $|l\rangle$ and $|l'\rangle$ are the incoming and outgoing electron states respectively and $|p\rangle$ is the target hadron state. The electron and the hadron masses are neglected in the limit $\frac{m^2}{s} \ll 1$. The sum over X represents the appropriate sums and integrals of the Lorentz invariant hadronic states $|X\rangle$. For unpolarized leptons, the leptonic tensor $L_{\mu\nu}$ is given by

$$L_{\mu\nu}(l, l') = \frac{1}{2} \text{Tr} \left[\not{l} \gamma_\mu \not{l}' \gamma_\nu \right]. \quad (2.2)$$

The hadronic tensor is given by

$$H^{\mu\nu}(p, q) = \frac{1}{4\pi} \int d^4z e^{iq \cdot z} \langle p, S | j^\mu(z) j^\nu(0) | p, S \rangle, \quad (2.3)$$

where $|p, S\rangle$ denotes a hadronic state with momentum p and spin S . To get to Eq. (2.3)

from Eq. (2.1), one must translate one of the currents from the origin to a location z and the completeness relation of hadronic final states is used to remove their sum. This last step, $\sum_X |X\rangle\langle X| = 1$ explains why the full inclusion of all hadronic final states is reasonable for the theoretical analysis.

The hadronic tensor can also be defined as a forward matrix element of the commutator of the two currents,

$$\begin{aligned} H'^{\mu\nu}(p, q) &= \frac{1}{4\pi} \int d^4z e^{iq \cdot z} \langle p, S | [j^\mu(z), j^\nu(0)] | p, S \rangle, \\ &= \frac{1}{4\pi} \int d^4z \left(e^{iq \cdot z} \langle p, S | j^\mu(z) j^\nu(0) | p, S \rangle + e^{-iq \cdot z} \langle p, S | j^\nu(z) j^\mu(0) | p, S \rangle \right). \end{aligned} \quad (2.4)$$

One can identify the second term in the commutator matrix element as $-H^{\nu\mu}(-q, p)$ by using translational symmetry. The terms in the commutator matrix element will create either an outgoing final state X with momentum $p + q$ or $p - q$. With $|p, S\rangle$ being a stable, on-shell, single particle state, only $p + q$ or $p - q$, but not both, can possibly be a physical state for X while the other would not contribute. This makes the two definitions equivalent.

The hadronic tensor, and the functions in its Lorentz decomposition, will be important objects for understanding hadronic structure. The hadronic tensor has some properties to reduce the number of available Lorentz structures. First, QCD is parity invariant. The hadronic tensor is Hermitian. The conservation of the electromagnetic current forces a Ward-Takahashi identity, $q_\mu H^{\mu\nu} = 0 = q_\nu H^{\mu\nu}$. Finally, the hadronic tensor must be linear in the spin axial vector S^μ . With all of these properties in mind, the Hadronic tensor can

be Lorentz decomposed into

$$\begin{aligned}
H^{\mu\nu}(p, q) = & \left(-g^{\mu\nu} + \frac{q^\mu q^\nu}{q^2} \right) F_1(x, Q^2) + \frac{(p^\mu - q^\mu \frac{p \cdot q}{q^2})(p^\nu - q^\nu \frac{p \cdot q}{q^2})}{p \cdot q} F_2(x, Q^2) \\
& + i\epsilon^{\mu\nu\alpha\beta} \frac{q_\alpha S_\beta}{p \cdot q} g_1(x, Q^2) + i\epsilon^{\mu\nu\alpha\beta} \frac{q_\alpha (S_\beta - p_\beta \frac{S \cdot q}{p \cdot q})}{p \cdot q} g_2(x, Q^2), \quad (2.5)
\end{aligned}$$

where the Lorentz invariant functions $F_i(x, Q^2)$ and $g_i(x, Q^2)$ are known as the structure functions (SF). It is these Lorentz invariant functions which are measured from experiments and analyzed to extract parton distributions.

2.2 Parton model

In order to understand the relationship between the structure functions and the parton distributions, it is necessary to describe what is actually meant by an electromagnetic current interacting with the hadron. The parton model, as originally proposed by Feynman [16], describes the photon interacting with a single constituent of the hadron, called a parton. Partons are not exclusively the Strongly interacting particles, quarks, antiquarks, and gluons, though these have the largest distributions. The distributions of these Strongly interacting partons at lowest orders are proportional to $O(1)$ or $O(\alpha_s)$. Electrons and photons, amongst other fields, can be considered as partons and have their own parton distributions. These distributions, coming from ElectroWeak interactions and therefore being $O(G_F)$ or $O(\alpha_{\text{EM}})$, are much smaller than the PDFs for the Strongly interacting fields. These parton distributions are frequently neglected, but there exist efforts, such as [17], to measure them.

The goal of DIS experiments is to study the internal structure of the proton, which requires very short distance probes, or equivalently large momentum transfer Q^2 . The asymptotic freedom inherent in QCD allows for partons interacting with probes at high

energy scales, compared to hadronic scales such as Λ_{QCD} , to interact individually with the probes. The other partons are separated by a typical distance scales of the order of $\Lambda_{\text{QCD}}^{-1}$, which is assumed large compared to Q^{-1} , and their interactions are neglected. This separation of length scales leads to the concept of factorization of the spin averaged Hadronic tensor into these two regions

$$H^{\mu\nu}(p, q) = \sum_i \int_x^1 dy \, C_i^{\mu\nu}(y, p, q, \frac{Q^2}{\mu^2}) f_i(\frac{x}{y}, \mu^2) + O(\frac{\Lambda_{\text{QCD}}^2}{Q^2}) + O(\frac{m^2}{Q^2}), \quad (2.6)$$

where i labels the different types of partons, C_i is a perturbatively calculable, short distance, hard scattering coefficient, f_i is the non-perturbative long distance parton distribution, and $O(\frac{\Lambda_{\text{QCD}}^2}{Q^2})$ and $O(\frac{m^2}{Q^2})$ represent neglected correction terms. As a motivation for the procedure which will be used later for relating two current matrix elements to PDFs, the leading order $C_i^{\mu\nu}$ will be determined using the parton model. The parton model can be seen as an approximation of QCD with non-interacting quarks. In renormalizable theories, such as QCD, the scale dependence only occurs from the terms at higher order in α_s . It is important to note that without interactions here will be no scale dependence of the result. The separation of distance scales will only occur in the QCD corrections to the parton model result.

For the following, it will be useful to consider the Breit frame. The positive z direction will be defined by the direction of the hadron's motion and the photon is moving in the negative z direction. For the parton model in this frame, the struck parton's 3-momentum will exactly reverse after the interaction, which is why the Breit frame is also known as the brickwall frame. The frame is set by the hadron and photon momenta, defined as

$$p = (p^+, \frac{m^2}{2p^+}, 0_T) \quad q = (-xp^+, \frac{Q^2}{2xp^+}, 0_T). \quad (2.7)$$

The parton has a momentum k^μ whose $+$ direction is a fraction of the hadron's momentum $k^+ = \xi p^+$. The momentum fraction ξ is limited to the range $[-1, 1]$ because no parton can have more energy than the constituent hadron. In the parton model, the difference between the Lorentz invariant quantity x and the momentum fraction ξ given by $x - \xi \approx xm^2/Q^2$ and will be neglected.

In the DIS regime, the leading order in momentum Q contributing to the Hadronic tensor is where the two currents interact with the same parton with incoming and outgoing momentum k and k' defined by

$$k = (\xi p^+, k^-, k_T) \quad k' = k + q = ((\xi - x)p^+, k^- + \frac{Q^2}{2xp^+}, k_T). \quad (2.8)$$

This situation is schematically represented in the “handbag” diagram shown in Fig. 2.1a. The alternative, called the “cat ears” diagram, shown in Fig. 2.1b, is subleading in powers of $\frac{m}{Q}$ or $\frac{\Lambda_{\text{QCD}}}{Q}$. In both of these figures, the bubbles represent all possible graphs with lines with momenta colinear to the quark lines.

FIG. 2.1: The two topologies of Wick contractions for the hadronic tensor. On the left is the “handbag” diagram where the same quark line is struck by the incoming currents. On the right is the “cat ears” diagram where different quarks are struck by the currents.

At leading order in QED, the handbag diagram's contribution to the Hadronic tensor

can be written as the integrals over the parton's momentum

$$H^{\mu\nu} = \sum_i \frac{e_i^2}{4\pi} \int \frac{d^4k}{(2\pi)^4} \text{Tr} [\gamma^\mu U_i(k') \gamma^\nu L_i(k, P)] . \quad (2.9)$$

where i labels the different parton types and U_i and L_i represent the upper and lower bubbles which have suppressed color and spin indices. In order to isolate the leading power in m/Q and Λ_{QCD}/Q , the smallest components of the parton's momenta k^- , k_T , and $k'^+ = (\xi - x)p^+$ are neglected when possible. This approximation is done for the U_i bubble by considering it in a frame where k' has no transverse component and then neglecting the k^- contributions

$$k' \rightarrow (k^+ + q^+ - \frac{k_T^2}{2(q^- + k^-)}, q^- + k^-, 0_T) \approx (k^+ + q^+ - \frac{k_T^2}{2q^-}, q^-, 0_T) . \quad (2.10)$$

For the L_i bubble, the component k^+ is approximated to be fixed at xp^+ , which allows the integral over the parton's momentum to be separated into

$$H^{\mu\nu} = \sum_i \frac{e_i^2}{4\pi} \text{Tr} \left[\gamma^\mu \left(\int \frac{dl^+}{2\pi} U_i((l^+, q^-, 0_T)) \right) \gamma^\nu \left(\int \frac{dk^- d^2k_T}{(2\pi)^3} L_i((xp^+, k^-, k_T), p) \right) \right] . \quad (2.11)$$

where $l^+ = k^+ + q^+ - \frac{k_T^2}{2q^-}$. L_i can be decomposed into the basis of Dirac structures

$$L = \mathcal{L}_S + \mathcal{L}_P \gamma_5 + \mathcal{L}_V^\mu \gamma_\mu + \mathcal{L}_A^\mu \gamma_\mu \gamma_5 + \mathcal{L}_T^{\mu\nu} \sigma_{\mu\nu} . \quad (2.12)$$

In the Breit frame, $+$ components are grown exponentially while $-$ components are shrunk by the same factor. L is approximated to be dominated by $\mathcal{L}_V^+$, $\mathcal{L}_A^+$, and $\mathcal{L}_T^{+i}$, where i is one of the transverse directions. By choosing appropriate projectors, $\mathcal{P}_A = \frac{1}{2} \gamma^- \gamma^+$ and

$\mathcal{P}_B = \frac{1}{2}\gamma^+\gamma^-$, these components can be isolated by modifying the integral to

$$H^{\mu\nu} = \sum_i \frac{e_i^2}{4\pi} \text{Tr} \left[\gamma^\mu \left(\int \frac{dl^+}{2\pi} \mathcal{P}_B U_i((l^+, q^-, 0_T)) \mathcal{P}_A \right) \gamma^\nu \left(\int \frac{dk^- d^2 k_T}{(2\pi)^3} \mathcal{P}_A L_i((xp^+, k^-, k_T), p) \mathcal{P}_B \right) \right]. \quad (2.13)$$

The unpolarized and polarized PDFs can be defined by the integrals

$$\begin{aligned} f_i(x) &= \int \frac{dk^- d^2 k_T}{(2\pi)^4} \text{Tr} \frac{\gamma^+}{2} L_i(k, p), \\ \Delta f_i(x) &= \frac{1}{\lambda} \int \frac{dk^- d^2 k_T}{(2\pi)^4} \text{Tr} \frac{\gamma^+ \gamma_5}{2} L_i(k, p), \\ \delta_T^j f_i(x) &= \frac{1}{b_T^j} \int \frac{dk^- d^2 k_T}{(2\pi)^4} \text{Tr} \frac{\gamma^+ \gamma^j \gamma_5}{2} L_i(k, p), \end{aligned} \quad (2.14)$$

where λ and b_T^j are the helicity and the polarization in the transverse j direction of the target hadron. Inserting these definitions into Eq. (2.13), one arrives at

$$H^{\mu\nu} = \sum_i \frac{e_i^2}{4\pi} \int \frac{dl^+}{xp^+} \text{Tr} \left[\gamma^\mu \mathcal{P}_B U_i((l^+, q^-, 0_T)) \mathcal{P}_A \gamma^\nu \frac{\hat{k}}{2} (f_i(x) - \gamma_5 \lambda \Delta f_i(x) - \gamma_5 \gamma^j b_T^j \delta_T^j f_i(x)) \right], \quad (2.15)$$

where $\hat{k} = (xp^+, 0, 0_T)$ is an approximation to the parton momentum.

In the parton model, which is assumed to be super renormalizable, the upper bubble can be approximated by the leading order cut massless propagator $\mathcal{P}_B U_i \mathcal{P}_A = q^- \gamma^+ (2\pi) \delta(2l^+ q^-)$. Inserting this approximation into Eq. (2.15), one obtains

$$\begin{aligned} H^{\mu\nu} &= \frac{2\pi}{Q^2} \text{Tr} \left[\gamma^\mu (\hat{k} + \not{q}) \gamma^\nu \frac{\hat{k}}{2} (f_i(x) - \gamma_5 \lambda \Delta f_i(x) - \gamma_5 \gamma^j b_T^j \delta_T^j f_i(x)) \right], \\ &= \sum_i \frac{e_i^2}{4\pi} \left[\frac{1}{2} (-g^{\mu\nu} + \frac{q^\mu q^\nu}{q^2}) f_i(x) + \frac{(\hat{k}^\mu - q^\mu \frac{\hat{k} \cdot q}{q^2})(\hat{k}^\nu - q^\nu \frac{\hat{k} \cdot q}{q^2})}{\hat{k} \cdot q} f_i(x) + \frac{1}{2} i \epsilon^{\mu\nu\alpha\beta} \frac{q_\alpha \hat{k}_\beta}{\hat{k} \cdot q} \lambda \Delta f_i(x) \right] \end{aligned} \quad (2.16)$$

In the parton model, as will be in leading order QCD, the matching coefficients $C_i^{\mu\nu}$ in Eq. (2.6) will contain some, but not all, of the same Lorentz structures as the Hadronic

tensor's Lorentz decomposition and delta functions of $1 - y$. By comparing Eq. (2.16) to the definitions of the SFs in Eq. (2.5), the SFs can be related to sums of the PDFs

$$F_1(x, Q^2) = \frac{1}{2} \sum_i e_i^2 f_i(x, Q^2), \quad F_2(x, Q^2) = x \sum_i e_i^2 f_i(x, Q^2),$$

$$g_1 = \frac{1}{2} \sum_i e_i^2 \Delta f_i(x), \quad g_2 = 0. \quad (2.17)$$

The parton model is a useful tool for demonstrating the decomposition of a matrix element such as hadronic tensor into a sum of PDFs. This procedure did require a number of assumptions, not all of which, such as super renormalizability, are valid for QCD. Even so, it does make predictions, such as Bjorken scaling, which do approximately hold in certain regions of x and Q .

The relationship between the SFs and the PDFs will motivate the method of calculating the PDFs from lattice calculable matrix elements with two current operators which is shown in Sec. 4.5. As with the case of the hadronic tensor, the matrix elements can be decomposed into Lorentz structures with Lorentz invariant functions. For a particular matrix element some, but not all, of these Lorentz invariant functions will be factorizable into the PDF.

It is from experimental cross sections, such as the one in Eq. (2.1), from which the PDF can be fit. For example, the CJ collaboration [2] used DIS cross sections, the F_2 structure function, pp and pd Drell-Yan cross sections, jet cross sections, γ +jet cross sections, and W and Z boson asymmetries in the CJ15 analysis. This data set included over 4500 independent measurements. The results of the fit are shown in Fig 2.2.

FIG. 2.2: The CJ15 PDF fit results [2]. This analysis included 4500 data points from several types of experiments. DIS cross sections, the F_2 structure function, pp and pd Drell-Yan cross sections, jet cross sections, γ +jet cross sections, and W and Z boson asymmetries were all used.

2.3 Operator definitions

A useful feature of the factorization is that the long distance component is independent of the experimental process. The PDF itself is a universal property of the hadron and not specific on whether it was used for DIS or DY or any other type of experiment. Unlike the above definition of the PDF through the parton model and factorization, this universal nature motivates a simple hadronic matrix element definition. The quark PDF can be

defined by

$$f_q(x, \mu^2) = \int \frac{d\xi^-}{2\pi} e^{-i\xi^- p^+ x} \langle p | \bar{\psi}_q(0, \xi^-, 0_T) \frac{\gamma^+}{2} W((0, \xi^-, 0_T); 0) \psi_q(0) | p \rangle_\mu \quad (2.18)$$

where $W((0, \xi^-, 0_T); 0)$ is a straight Wilson line running between the two quark fields. The antiquark PDF has a similar definition

$$f_{\bar{q}}(x, \mu^2) = \int \frac{d\xi^-}{2\pi} e^{-i\xi^- p^+ x} \langle p | \psi_q(0, \xi^-, 0_T) \frac{\gamma^+}{2} W((0, \xi^-, 0_T); 0) \bar{\psi}_q(0) | p \rangle_\mu \quad (2.19)$$

and with a change of variables, one can see the crossing symmetry that $f_q(x) = -f_{\bar{q}}(-x)$. The polarized PDFs have analogous definitions with γ^+ replaced by $\gamma^+ \gamma_5$ and $\gamma^+ \gamma^j \gamma_5$, with $j = x, y$, for the helicity and transversity PDFs respectively.

It has been proposed [18], instead to study the Fourier transform of the PDF, called the Ioffe time distribution function (ITD)

$$Q(\nu, \mu^2) = \int_{-1}^1 dx e^{i\nu x} f(x, \mu^2), \quad (2.20)$$

where $\nu = p \cdot \xi$ is called the Ioffe time. The Ioffe time measures the typical time scale in which the probes interact with the parton, in units of the inverse mass of the hadron. The ITD, being directly proportional to the non-local matrix element, not its Fourier transform, will be useful to consider for future lattice QCD applications.

The ITD, unlike the PDF, is a complex valued function. By analyzing the real and imaginary components individually, different aspects of the quark PDFs can be revealed. Consider the quark and antiquark PDFs within the limits $x \in [0, 1]$ which are defined by

$$q(x) = f(x) \quad , \quad \bar{q}(x) = -f(-x) \quad \forall x \in [0, 1]. \quad (2.21)$$

The ITD can be written as

$$Q(\nu, \mu^2) = \int_0^1 dx e^{i\nu x} f(x, \mu^2) + \int_{-1}^0 dx e^{i\nu x} f(x, \mu^2) = \int_0^1 dx (e^{i\nu x} q(x, \mu^2) - e^{-i\nu x} \bar{q}(x, \mu^2)) . \quad (2.22)$$

The real and imaginary components of the ITD are given by

$$\begin{aligned} \text{Re } Q(\nu, \mu^2) &= \int_0^1 \cos(\nu x) (q(x) - \bar{q}(x)) = \int_0^1 \cos(\nu x) q_v(x) , \\ \text{Im } Q(\nu, \mu^2) &= \int_0^1 \sin(\nu x) (q(x) + \bar{q}(x)) = \int_0^1 \sin(\nu x) (q_v(x) + 2\bar{q}(x)) , \end{aligned} \quad (2.23)$$

where the valence quark distribution $q_v(x) = q(x) - \bar{q}(x)$ and it is assumed that the “sea” distribution is equal to the antiquark distribution. The valence quark distribution satisfies a special property that its integral is fixed by the quark flavor quantum numbers of the hadron

$$n_q = \int_0^1 dx q_v(x) . \quad (2.24)$$

where n_q is the number of times the quark of flavor q appears in the quark model wave-function for that hadron. In a nucleon, the normalizations for the up and down quarks are $n_u = 2$ and $n_d = 1$ respectively. This feature is called the quark sum rule.

The PDF can be modeled by its low and high x asymptotic behaviors. A simple Ansatz for the quark valence PDF

$$q_v(x) = n_q \frac{x^a (1-x)^b}{B(a+1, b+1)} , \quad (2.25)$$

where $B(a, b)$ is the Beta function and n_q is the normalization from the q flavor valence

quark sum rule. The ITD for this model is given by

$$Q(\nu) = \frac{n_q \pi}{2^{a+b+1}} \Gamma(a+1) \Gamma(b+1) {}_2\tilde{F}_3 \left(\frac{a+1}{2}, \frac{a+2}{2}; \frac{1}{2}, \frac{1}{2}(a+b+2), \frac{1}{2}(a+b+3); -\frac{\nu^2}{4} \right), \quad (2.26)$$

where ${}_n\tilde{F}_m$ is the generalized hypergeometric function.

2.3.1 DGLAP evolution

So far, the second argument of the PDF, the scale μ^2 , has been largely neglected. In the parton model, unlike QCD, the SFs are completely independent of the scale Q^2 , a feature called Bjorken scaling. In a renormalizable theory, like QCD, the matrix element in Eq. (2.18) will require a proper treatment of divergences, typically in the $\overline{\text{MS}}$ renormalization scheme. The PDF is renormalized by

$$f_i(x, \mu^2) = \sum_{i'} \int_x^1 \frac{dz}{z} Z_{ii'}(z, \alpha_s, \epsilon, \mu^2) f_{i'}^{(0)}\left(\frac{x}{z}\right). \quad (2.27)$$

An interesting feature of the renormalization is that it can mix PDFs for different types of partons.

The scale dependence of the PDF is governed by the Renormalization Group equations. The PDF's Renormalization Group equation, given by

$$\mu^2 \frac{d}{d\mu^2} f_i(x, \mu^2) = 2 \sum_{i'} \int_x^1 \frac{dz}{z} P_{ii'}(z, \alpha_s) f_{i'}\left(\frac{x}{z}, \mu^2\right), \quad (2.28)$$

is called the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) equations, named for its many discoverers [19, 20, 21]. The DGLAP kernel $P_{ii'}$ can be calculated in perturbation theory. This equation can be used to evolve PDFs along different μ for simultaneously analyzing cross sections which originated with different Q^2 .

2.4 Operator Product Expansion

The PDF and ITD are defined in terms of a nonlocal hadronic matrix element. This matrix element is unsuitable for calculation in Euclidean spacetime, a requirement of numerical Lattice QCD applications, due to the light-like separation involved. Unlike in Minkowski space, in Euclidean space there exists no spacetime point which is separated from the origin such that $\xi^2 = 0$. Using the Operator Product Expansion (OPE), local matrix elements, which are calculable in Lattice QCD, can be related to properties of the PDF. These local matrix elements can be simpler objects to deal with in perturbation theory than the nonlocal matrix elements. For example, high order evolution equations are derived for the local matrix elements first and subsequently used to construct the PDF evolution equations [22, 23].

Consider the OPE for the general version of the Wilson line operator defining the ITD which is given by

$$\bar{\psi}_q(\xi) \frac{\gamma^\alpha}{2} W(\xi; 0) \psi_q(0) = \sum_{n=0} \frac{1}{n!} \left[\prod_{i=1}^n \xi_{\mu_i} \right] \bar{\psi}_q \frac{\gamma^\alpha}{2} \left[\prod_{i=1}^n D^{\mu_i} \right] \psi_q(0) \quad (2.29)$$

where D_μ is the gauge covariant derivative and the products of ξ_{μ_i} and D_{μ_i} are 1 for the $n = 0$ term. Matrix elements of these local derivative operators will be proportional to factors of momentum vectors and metric tensors multiplying a reduced matrix element

$$\langle p | \bar{\psi}_q \frac{\gamma^\alpha}{2} \left[\prod_{i=1}^n D^{\mu_i} \right] \psi_q(0) | p \rangle = \frac{p^\alpha}{2} \left[\prod_{i=1}^n p^{\mu_i} \right] m_n(\mu^2) + \text{traces}, \quad (2.30)$$

where the traces term represents the terms proportional to factors of the metric tensor. The terms with metric tensors will contract two separation vectors which vanish for separations on the light cone with $\xi^2 = 0$. After these higher twist terms have been removed, the ITD

is given by the expansion

$$\begin{aligned}
p^\alpha Q(\nu, \mu^2) &= \langle p | \bar{\psi}_q(\xi) \frac{\gamma^\alpha}{2} W(\xi; 0) \psi_q(0) | p \rangle_\mu = p^\alpha \sum_{n=0} \frac{1}{n!} \left[\prod_{j=1}^n i \xi_{\mu_j} p^{\mu_j} \right] m_n(\mu^2) \\
&= p^\alpha \sum_n \frac{i^n \nu^n}{n!} m_n(\mu^2), \quad (2.31)
\end{aligned}$$

where $m_n(\mu^2)$ are the reduced matrix elements of the leading twist local operator. Compare this expression to the Taylor series of the ITD with respect to ν about $\nu = 0$

$$Q(\nu, \mu^2) = \int_{-1}^1 dx \left(1 + i\nu x + \frac{1}{2}(i\nu x)^2 + \dots \right) f(x, \mu^2). \quad (2.32)$$

Therefore, the Mellin moments of the PDF are the reduced matrix elements of the leading twist local operators

$$m_n(\mu^2) = \int_{-1}^1 dx x^n f(x, \mu^2). \quad (2.33)$$

This relationship provides an alternative definition of the PDF. The PDF is the function whose n^{th} Mellin moment is the reduced matrix element of the operator

$$O^{\{\mu_0 \mu_1 \dots \mu_n\}}(\mu^2) = \bar{\psi} \gamma^{\{\mu_0} i D^{\mu_1} \dots i D^{\mu_n\}} \psi - \text{traces} \quad (2.34)$$

where the indices in the brackets are symmetrized and the traces are lower dimension operators times factors of the metric tensor. This definition, unlike the matrix element definition in Eq. (2.18), is completely Lorentz covariant. It explicitly highlights a source of errors which will appear in the lattice calculable alternatives to the PDF, the higher twist terms proportional to ξ^2 which were dropped.

In some cases, the moments of the PDF are simpler objects to work with than the PDF itself. This typically manifests itself in the equations which involve convolutions of the PDF, such as the DGLAP evolution formula in Eq. (2.28). The evolution of moments

is multiplicative, governed by the Mellin moments of the DGLAP kernel. Consider the generic convolution for the PDF's evolution.

$$f(x, \mu'^2) = \int_x^1 \frac{dy}{y} P\left(y, \frac{\mu'^2}{\mu^2}, \alpha_s\right) f\left(\frac{x}{y}, \mu^2\right) \quad (2.35)$$

Multiplying by x^n and integrating over x , one gets

$$m_n(\mu'^2) = \int_0^1 dx x^n \int_x^1 \frac{dy}{y} P\left(y, \frac{\mu'^2}{\mu^2}, \alpha_s\right) f\left(\frac{x}{y}, \mu^2\right). \quad (2.36)$$

With a change of variables $z = \frac{x}{y}$, a multiplicative evolution for the moments is revealed

$$m_n(\mu'^2) = \int_0^1 dy y^n P\left(y, \frac{\mu'^2}{\mu^2}, \alpha_s\right) \int_0^1 dz z^n f(z, \mu^2) = P_n\left(\frac{\mu'^2}{\mu^2}, \alpha_s\right) m_n(\mu^2). \quad (2.37)$$

where $P_n(\frac{\mu'^2}{\mu^2}, \alpha_s)$ is the Mellin moment of the PDF's evolution kernel $P(y, \frac{\mu'^2}{\mu^2}, \alpha_s)$

CHAPTER 3

Lattice QCD

Lattice regularization is a method of making the Feynman path integral finite by imposing that all fields exist only in a finite volume, discretized space-time lattice. The fields exist on either the sites or links between the sites. Lattice regularization, just like other regulators, can be approximated with a perturbative small coupling expansion calculations, but one must pay the price of a large number of trigonometric functions and a proliferation of diagrams. Even more interestingly, Lattice Field Theories (LFTs) also allow for approximations with a strong coupling expansion. Most useful for QCD, LFTs can be solved numerically by Wick rotating the Feynman path integral into a Euclidean metric. Numerical Lattice QCD is the only method of dealing with Strong dynamics in a systematically controllable way. In principle, numerical LFT can be used to evaluate Feynman path integrals in any regime of coupling to any desired precision, though in practice certain calculations may require more computational time than is currently available.

In a LFT, the lattice spacing a and the lattice volume L are UV and IR regulators providing an upper momentum cutoff at $\frac{2\pi}{a}$ and lower momentum cutoff at $\frac{2\pi}{L}$ respectively. For a proper comparison with experiments, a continuum limit, $a \rightarrow 0$, and an infinite volume

limit, $L \rightarrow \infty$, are necessary. Since in practice, lacking an infinitely sized computer, one cannot actually perform a calculation in these extreme limits and instead extrapolations are made. The error from this extrapolation can be systematically improved by paying the computational cost and lowering the value of a and raising the value of L . Unlike other methods for calculating QCD observables, lattice QCD is the only known method with this sort of systematically controllable errors, making it a vital tool for studying Strong interactions.

3.1 The Path Integral and Statistical Mechanics

The numerical evaluation of the Feynman path integral uses the same algorithms as for evaluating thermodynamic quantities. The Wick rotated path integral is recast as an expectation value with a Gibb's probability measure

$$\langle T\{O(\phi)\} \rangle = \mathcal{Z}^{-1} \int D\phi O(\phi) e^{-S(\phi)} \rightarrow \langle O \rangle = Z^{-1} \sum_n O_n e^{-\beta E_n}, \quad (3.1)$$

where $\mathcal{Z}$ is the generating functional of the LFT and Z is the partition function of the thermal system. To highlight the importance of the interplay between the LFT and Statistical Mechanics pictures, consider the uses of LFT analogs of the inverse temperature β . In the case of Yang Mills theory, β can be identified with the inverse of the bare coupling. A high temperature, low β , expansion in the statistical mechanics language would be a strong coupling expansion in LFT. Using the strong coupling expansion, a $SU(N)$ Yang-Mills gauge theory can be shown to be confining in the limit $\beta \rightarrow 0$. In asymptotically free theories, such as QCD, there exists a second order phase transition at zero coupling or infinite β . As one approaches a critical point, the systems correlation length ξ grows infinite. The correlation length is inversely proportional to the lightest particle in the LFT's

spectrum in this case the pion, $\xi^{-1} = m_\pi a$. With the pion mass held finite and fixed, it is the limit of $\xi^{-1} \propto a \rightarrow 0$ or equivalently $\beta \rightarrow \infty$ which constitutes the continuum limit. Unfortunately, due to the large correlation length, calculations near critical points are increasing computationally demanding.

Monte Carlo integration, a stochastic method of numerical integration, is performed by randomly sampling the variables and evaluating the integrand. It is well suited for high dimensional systems. A modest Lattice QCD calculation can require evaluating a 10 million dimensional integral. The integrand of the path integral and many practical thermal systems are widely varying and sharply peaked functions with complicated dependence on the specific configuration of fields.

Due to these sharp peaks in the Gibb's probability measure, in the LFT case given by

$$P_G = e^{-S}/\mathcal{Z}, \quad (3.2)$$

most regions of field space do not contribute to the integral significantly. One cannot randomly sample this space with a uniform distribution and expect a precise result in a calculation of reasonable length. Instead one must perform importance sampling where only the regions which contribute the most to the integral are considered. The integrand is split into two parts, the sharply peaked and rapid varying which will be the Gibb's probability measure, P_G , and a slow smooth function which will be the observables, $O(\phi)$. The observable is averaged over a set of field configurations generated with a distribution described by the Gibb's probability measure and

$$\langle T\{O(\phi)\} \rangle = \int D\phi O(\phi) P_G(\phi) \approx \frac{1}{N} \sum_{i=1}^N O(\phi_i) \quad (3.3)$$

where the set $\{\phi_i\}$ was generated according to the probability distribution $P_G(\phi)$.

In order to generate the field configurations to a specific probability distribution, one uses a Monte Carlo Markov chain. A Monte Carlo Markov chain is a stochastic process for creating a sequence of possible configurations ϕ_i . A key feature of a Markov chain is that the selection of the next point only depends on the present point and not on the history of previous points. One requirement of this procedure is ergodicity, that all configurations are possible to occur. This feature means that the entire phase space will eventually be sampled and by increasing the number of configurations the approximation in Eq. (3.3) will systematically improve.

3.2 Wilson's Actions

In a Lattice Gauge Theory (LGT), matter fields exist on the lattice sites and gauge fields exist on the links between the lattice sites. The lattice gauge fields are related to the continuum gauge fields by the path ordered integral

$$U_\mu(x) = P \exp \left[i \int_x^{x+a\hat{\mu}} dy A_\mu(y) \right], \quad (3.4)$$

where $U_\mu(x)$ is represented by the link between lattice sites x and $x + a\hat{\mu}$. With a SU(3), or any non-Abelian group, gauge symmetry, these links are directional. It is useful to note the relation between cross the links in the forward $+\hat{\mu}$ and backward $-\hat{\mu}$ directions, $U_\mu^\dagger(x) = U_{-\mu}(x + a\hat{\mu})$. The difference between the lattice matter fields and the continuum matter fields is only the restriction to lattice sites.

The gauge transformation for these gauge links is given by $U_\mu(x) \rightarrow \Omega(x)U_\mu(x)\Omega^\dagger(x + a\hat{\mu})$. The quark fields gauge transformations are given by $\psi(x) \rightarrow \Omega(x)\psi(x)$ and $\bar{\psi}(x) \rightarrow \bar{\psi}(x)\Omega^\dagger(x)$. Any closed path of gauge links or any pairs of quark and antiquark fields connected by gauge links will create a gauge invariant quantity. In a LGT, all observables

must be made up of gauge invariant quantities. The path integral of a gauge variant quantity will be explicitly zero.

The action must be written in terms of gauge invariant closed loops of gauge links. The smallest non-trivial loop possible, called the plaquette, is the basis for the Wilson gauge action [24]

$$P_{\mu\nu}(x) = U_\mu(x)U_\nu(x + a\hat{\mu})U_\mu^\dagger(x + a\hat{\nu})U_\nu^\dagger(x). \quad (3.5)$$

The Wilson gauge action is given by a sum over the plaquettes

$$S_g = \frac{2}{g^2} \sum_x \sum_{\mu < \nu} \text{ReTr} [1 - P_{\mu\nu}(x)] = \frac{a^4}{2g^2} \sum_x \sum_{\mu, \nu} \text{Tr} [F_{\mu\nu}(x)F^{\mu\nu}(x)] + O(a^2). \quad (3.6)$$

This action reproduces the continuum gauge action up to $O(a^2)$.

In the quark action, the gauge covariant derivative is replaced by a finite difference derivative containing a gauge link. The naïve lattice quark action is given by

$$S_q = a^4 \sum_x \bar{\psi}(x) \sum_\mu \gamma_\mu \frac{1}{2a} (U_\mu(x)\psi(x + a\hat{\mu}) - U_{-\mu}(x)\psi(x - a\hat{\mu})) + m\bar{\psi}(x)\psi(x). \quad (3.7)$$

This action is referred to as naïve, because if one examines the massless, free quark propagator in momentum space, given by

$$G(p) = \frac{-ia^{-1} \sum_\mu \gamma_\mu \sin(p_\mu a)}{a^{-2} \sum_\mu \sin(p_\mu a)^2}, \quad (3.8)$$

there exist extra unphysical poles besides $p^2 = 0$. These poles, referred to as doublers, would spoil any calculation performed with this action and must be removed. The Wilson

fermion action [25] given by

$$S_q = a^4 \sum_x \bar{\psi}(x) \sum_{\mu} (1 - \gamma_{\mu}) \frac{1}{2a} (U_{\mu}(x) \psi(x + a\hat{\mu}) - U_{-\mu}(x) \psi(x - a\hat{\mu})) + \left(m + \frac{4}{a}\right) \bar{\psi}(x) \psi(x). \quad (3.9)$$

It removes the doublers by adding a term which gives them a mass which diverges in the continuum limit. This action reproduces the continuum action up to $O(a)$. It is common to write the discretized actions generically in terms of the Dirac matrix $D(x, y)$

$$S_q = a^8 \sum_{x, y} \bar{\psi}(x) D(x, y) \psi(y). \quad (3.10)$$

Quark propagators, $G(x, y)$, are calculated by solving the system of equations

$$a^4 \sum_x D(z, x) G(x, y) = \delta^{(4)}(z - y). \quad (3.11)$$

These propagators are referred to as point-to-all since they have a fixed source point, but go to all possible sink points. The size of the Dirac matrix, which can be a 25 million dimension square matrix for a modest lattice volume, makes this calculation rather costly and highly optimized inversion codes are used.

3.2.1 Symanzik Improvement

In theory, these simple Wilson actions are all that is needed to systematically calculate any observable with Lattice QCD. In practice, calculations at smaller lattice spacing have a demanding computational cost. In order to improve the lattice's approximation of the continuum without requiring finer lattice spacings, these actions can be modified by terms which vanish in the continuum limit. These terms can be specifically tuned to cancel the discretization effects which appear in the discretized actions. This procedure is known as

Symanzik improvement [26, 27]. The Symanzik improvement program can be applied as well to operators for matrix elements in order to remove their discretization effects.

For improving the quark action, there exist five possible operators which can be used to remove the $O(a)$ errors,

$$\begin{aligned}
\mathcal{O}_1 &= a\bar{\psi}(x)\sigma_{\mu\nu}F_{\mu\nu}(x)\psi(x), \\
\mathcal{O}_2 &= a\bar{\psi}(x)\vec{D}_\mu(x)\vec{D}_\mu(x)\psi(x) + a\bar{\psi}(x)\overleftarrow{D}_\mu(x)\overleftarrow{D}_\mu(x)\psi(x), \\
\mathcal{O}_3 &= am\text{Tr}[F_{\mu\nu}(x)F^{\mu\nu}(x)], \\
\mathcal{O}_4 &= am\left(\bar{\psi}(x)\vec{D}(x)\psi(x) - \bar{\psi}(x)\overleftarrow{D}(x)\psi(x)\right), \\
\mathcal{O}_5 &= am^2\bar{\psi}(x)\psi(x).
\end{aligned} \tag{3.12}$$

Of these operators, two of them, $\mathcal{O}_2$ and $\mathcal{O}_4$, can be eliminated by the equations of motion. Two already exist in the action, $\mathcal{O}_3$ and $\mathcal{O}_5$, now carrying an additional factor of the quark mass, and can be removed by a redefinition of the original parameters. The remaining operator, $\mathcal{O}_1$, called the clover term, is added to the quark action

$$S_{\text{imp}} = S_q + c_{\text{SW}}a^5\frac{1}{2}\sum_x\sum_{\mu<\nu}\bar{\psi}(x)\sigma_{\mu\nu}F_{\mu\nu}(x)\psi(x), \tag{3.13}$$

where c_{SW} is the Sheikholeslami-Wohlert coefficient [28]. This term is called the clover term, because the typical discretization of $F^{\mu\nu}(x)$, given by the sum of plaquettes

$$F_{\mu\nu}(x) \approx \frac{-i}{8a^2}(C_{\mu\nu}(x) - C_{\nu\mu}(x)) \quad C_{\mu\nu} = P_{\mu,\nu}(x) + P_{-\mu,\nu}(x) + P_{\mu,-\nu}(x) + P_{-\mu,-\nu}(x) \tag{3.14}$$

has the shape of a four leaf clover's pedals. The real coefficient c_{SW} can be determined perturbatively or non-perturbatively. Without the non-perturbative determination, the discretization effects will not exactly canceled, but hopefully will be made milder. On

the other hand, the non-perturbative value must be recalculated when the lattice setup is changed.

3.3 Determination of Matrix Elements

Though one may desire the hadronic matrix elements of some operator, in numerical LFT, one can only directly estimate vacuum expectation values, or correlation functions, of quark and gluon operators. Consider the correlation function of two Euclidean time separated operators O_h with the quantum numbers of some hadron h , called interpolating operators,

$$C_2(T) = \langle O_h(T) \bar{O}_h(0) \rangle = \mathcal{Z}^{-1} \int D\bar{\psi} D\psi D U_\mu O_h(T) \bar{O}_h(0) e^{-S(\bar{\psi}, \psi, U_\mu)}. \quad (3.15)$$

This correlation function is known as the two point function. The operators at Euclidean time 0 and T are referred to as the source and sink interpolating operators respectively. Time evolving the sink operator and inserting a complete set of energy eigenstates transforms the two point correlation function into the form

$$C_2(T) = \sum_n |Z_n|^2 e^{-E_n T} \quad Z_n = \langle \Omega | O_h(0) | n \rangle, \quad (3.16)$$

where n label all the energy eigenstates with the quantum numbers of h and the overlap factor Z_n measures how well the state $|n\rangle$ is represented by the interpolating operator. Keeping the operators well separated in time T , only the lower lying energy eigenstates will contribute to the sum. It is by analyzing the exponentials at large T that the spectrum of hadrons can be determined in lattice QCD.

Particularly for high energy states, the numerical calculation of the two point function

is frequently limited by the available machine precision. The signal-to-noise ratio of a two point correlation function of a hadron h boosted to momentum p is given by

$$R_{\text{S/N}} = \frac{C_2^h(T)}{\text{std}[C_2^h(T)]} = \frac{C_2^h(T)}{|C_2^h(T)|^2} = \frac{C_2^h(T)}{[C_2^\pi(T)]^{\frac{n_q}{2}}} = \frac{|Z_h|^2 e^{-E(p)T}}{(|Z_\pi|^2 e^{-m_\pi T})^{\frac{n_q}{2}}} \propto e^{-(E_h(p) - \frac{n_q}{2} m_\pi)T} \quad (3.17)$$

where $C_2^\pi(T)$ is the pion correlation function and n_q is the number of quark fields in the interpolating operator. If the overlap factor is too small for the state of interest, the exponentially decaying two point function will become very small reaching the level of machine precision. Two point function calculations are improved by choosing the interpolator such that the overlap with the desired state is as large as possible. A good choice increases the maximum Euclidean time separations which are free of numerical machine errors, and also so that the overlap with all other states are as small as possible, decreasing the minimum Euclidean time which have large higher state contamination. A large number of smearing procedures have been developed for providing a Gaussian shaped width to the quarks in the interpolating operators which have been successful in improving hadronic rest frame calculations.

3.3.1 Momentum smearing

As will be explained in Ch. 4, parton distribution calculations in lattice QCD will require results with as many momenta as are available and large momentum results are particularly desirable. Lattice QCD calculations of moving hadrons are plagued by poor statistical errors. There are two primary reasons for the signal-to-noise issue. The exponentially decreasing signal-to-noise ratio as Euclidean time increases is an unavoidable consequence of high momentum calculations. The lighter the hadron the faster the relative signal-to-noise ratio diminishes with momentum which makes high momentum pion

calculations exceptionally difficult.

The second reason is that the overlap of a smeared interpolating field decreases with momentum. As an example, consider Gaussian quark smearing which modifies the quark fields by iteratively applying the transformation

$$\tilde{q}(x) = q(x) + \rho \sum_k \left(2q(x) - U_k(x)q(x+k) - U_k(x-k)q(x-k) \right). \quad (3.18)$$

With sufficient iterations, this process will approximate $\tilde{q}(x) = \exp[-\rho \nabla]q(x)$, where ∇ is the gauge covariant Laplacian. This gauge covariant procedure is intended to approximate a Gaussian profile in position space for the interpolating field. The momentum space profile, the Fourier transform of this position space profile, will also be a Gaussian centered around momentum 0. This procedure improves the overlap of the smeared operator and the rest state, but the overlap of the moving hadron state, made up of moving quarks, will decrease exponentially with the momentum. The low overlap means that fewer time separations will be available before the inevitable numerical machine errors begin to dominate. With fewer reliable time separations, extraction of masses and matrix elements will be systematically more difficult at high momentum.

It has been suggested [29], that smearing processes should be modified to shift the quark's profile in momentum space, effectively boosting the quark field. For each desired momentum, this shift could be tuned so that the overlap is the same as the rest state. For Gaussian smearing, the momentum smearing procedure is to iteratively apply

$$\tilde{q}(x) = q(x) + \rho \sum_k \left(2q(x) - e^{i(\frac{2\pi}{L})\vec{\zeta} \cdot \hat{k}} U_k(x)q(x+k) - e^{i(\frac{2\pi}{L})\vec{\zeta} \cdot \hat{k}} U_k(x-k)q(x-k) \right). \quad (3.19)$$

After momentum smearing, the quark's momentum space profile will be a Gaussian cen-

tered at $\vec{\zeta}$. In the free field approximation, a proton with momentum $\vec{p}$ would require a momentum smearing parameter $(\frac{2\pi}{L}) \vec{\zeta} = \vec{p}/3$ and a pion would require $(\frac{2\pi}{L}) \vec{\zeta} = \vec{p}/2$. For a realistic case, the optimal momentum smearing parameter was found to be $(\frac{2\pi}{L}) \vec{\zeta} = 0.45 \vec{p}$ for a nucleon and $(\frac{2\pi}{L}) \vec{\zeta} = 0.8 \vec{p}$ for a pion [29].

The antiquark fields in the pion interpolating field are smeared for boosting in the opposite direction as the quark fields. If this change was not done and the same ζ were naïvely used, for an iterative smearing procedure such as the Gaussian procedure above, the phase factors will approximately cancel each other and the interpolating field will be unboosted for any initial momentum smearing parameter. The effectiveness of the momentum smearing procedure for improving the overlap can be shown by comparing signal-to-noise ratios at a fixed Euclidean time, provided the time is large enough to neglect excited states. Fig. 3.1 shows the signal-to-noise ratio for the nucleon at multiple momentum smearing parameters. As expected, the Gaussian shape of the overlap is shifted as momentum smearing is applied and the signal-to-noise is improved at larger momenta.

The advantage of momentum smearing can also be seen in effective mass calculations. The effective mass, defined by

$$m_{\text{eff}} = \lim_{T \rightarrow \infty} \ln \frac{C_2(T)}{C_2(T+1)} \quad (3.20)$$

is an estimator for the energy of a hadron from its two point function. Fig. 3.2 shows the effective mass of the nucleon and pion at three different momentum smearings. The relative error drops significantly for the large momentum effective masses. In some cases, energy levels of high momentum states previously unavailable are able to be resolved with momentum smearing.

To improve the signal-to-noise ratio for matrix elements of moving hadron states, it is not necessary to use precisely tuned values of the momentum smearing parameter, ζ .

As long as the overlap of the interpolating fields is improved, then the signal-to-noise ratio will be improved. Depending on the method of matrix element calculation, it may be possible to tune the parameters for each momentum individually without additional costs, for example in the use of sequential sources. It is also possible, as will be the case for using sequential operators, that a few momentum smearing parameters should be chosen to create overlapping Gaussians in momentum space. As long as these Gaussians span the desired region of momenta, the resulting correlation functions can be fit simultaneously to determine the matrix elements. Both of these motifs will be used in the following calculations.

3.3.2 Ratio method

In order to calculate matrix elements, additional operators must be inserted into the vacuum expectation value. The spectral decomposition can be performed with any number of time separated operators. The insertion of $O_{\text{op}}(t)$ at a time between the interpolating operators $0 < t < T$,

$$C_3(T, t) = \langle O_h(T) O_{\text{op}}(t) \bar{O}_h(0) \rangle = \sum_{n,m} Z_n Z_m^* e^{-E_n t} e^{-E_m (T-t)} \langle m | O_{\text{op}} | n \rangle, \quad (3.21)$$

is called the three point function. The three point function, in the limits of $t, T, (T-t) \rightarrow \infty$, can be used to extract a ground state matrix element of O_{op} . There exist a number of variations on how to calculate the two and three point functions and the determination of the matrix element. The details of the calculation of the correlation functions used in this work will be given in Ch. 5 and Ch. 6.

The standard method of calculating hadron matrix elements uses the Euclidean time separation dependence of the ratio of two point and three point correlation functions. In

the large Euclidean time limit,

$$R(T, t) = \frac{C_3(T, t)}{C_2(T)} = M^0 (1 + Ae^{-\Delta(T-t)} + Be^{-\Delta t}) \quad (3.22)$$

where $\Delta = E_1 - E_0$ is the energy gap between the ground state and lowest excited state, A and B are ratios of the ground and excited state overlap factors and transition matrix elements. In practice, this ratio is calculated for all t between the source and sink interpolating fields and a few T .

3.3.3 Feynman-Hellmann summation method

An alternative to the above approach is to consider the ratio in Eq. (3.22) with the operator insertion time t summed over. Neglecting higher state corrections, the resulting function of the source-sink separation will be a line whose slope is the desired matrix element. The operator would ideally be summed over only the insertion times between the interpolating fields, but the contributions from improper time ordering only contribute to higher state effects [30]. This method will also have significantly reduced excited state contamination at large times, T . We refer the reader to [30, 31, 32] for more details on the implementation of this method. The key points are highlighted below.

The Feynman-Hellmann theorem relates matrix elements of energy eigenstates, $|n\rangle_\lambda$ to derivatives of the corresponding energy eigenvalue, E_n with respect to a parameter of the theory, λ ,

$$\frac{dE_n}{d\lambda} = {}_\lambda\langle n | \frac{dH_\lambda}{d\lambda} | n \rangle_\lambda, \quad (3.23)$$

In order to calculate matrix elements of arbitrary operators, following the procedure in

[30], consider a change to the action,

$$S_\lambda = \lambda \int d^4x j(x) \quad (3.24)$$

where $j(x)$ is some bilinear current. The vacuum state is labeled by $|\lambda\rangle$. The 2-pt correlation functions are defined as

$$C_\lambda(t) = \langle \lambda | O(t) O^\dagger(0) | \lambda \rangle = \frac{1}{Z_\lambda} \int D\Phi e^{-(S+S_\lambda)} O(t) O^\dagger(0) \quad (3.25)$$

where $O(t)$ is an interpolating field for a desired hadron and Z_λ is the partition function defined by

$$Z_\lambda = \int D\Phi e^{-(S+S_\lambda)} O(t) O^\dagger(0). \quad (3.26)$$

These three objects, $|\lambda\rangle$, C_λ , and Z_λ , become the true QCD vacuum, correlation function, and partition function respectively in the limit that $\lambda \rightarrow 0$. In this formulation, the hadron matrix element of the current $j(x)$ can be found through the Feynman-Hellmann theorem,

$$\left. \frac{\partial m_\lambda^{h_n}}{\partial \lambda} \right|_{\lambda=0} = \int dt' \langle h_n | J(t') | h_n \rangle, \quad (3.27)$$

where $J(t) = \int d^3x j(\vec{x}, t)$ and $|h_n\rangle$ is the n^{th} hadron state with the quantum numbers of $O(t)$ with mass m^{h_n} . In the large Euclidean time limit, the ground state mass can be approximated by the effective mass

$$m_\lambda^{eff}(t, \tau) = \frac{1}{\tau} \ln \left(\frac{C_\lambda(t)}{C_\lambda(t+\tau)} \right) \xrightarrow{t \rightarrow \infty} \frac{1}{\tau} \ln(e^{m_\lambda^0 \tau}). \quad (3.28)$$

The derivative of the effective mass can be shown to be

$$\left. \frac{\partial m_\lambda^{eff}(t, \tau)}{\partial \lambda} \right|_{\lambda=0} = \frac{1}{\tau} (R(t+\tau) - R(t)), \quad (3.29)$$

where

$$R(t) = \frac{\int dt' \langle \Omega | T \{ O(t) J(t') O^\dagger(0) \} | \Omega \rangle}{C(t)}. \quad (3.30)$$

The contributions to the time ordered matrix element in the numerator of $R(t)$ from the regions where the current is not inserted between the hadron states where shown in [30] to be suppressed at large times in the difference in Eq. (3.29). At large Euclidean time, t , Eq. (3.29) isolates the matrix element of interest up to exponential corrections

$$\left. \frac{\partial m_\lambda^{eff}(t, \tau)}{\partial \lambda} \right|_{\lambda=0} = \frac{\langle h^0 | J(t) | h^0 \rangle}{2m^{h^0}} (1 + Ae^{-\Delta t} + Bte^{-\Delta t} + O(e^{-\Delta' t})). \quad (3.31)$$

where A and B fit parameters related to matrix elements of the lowest excited state with mass gap Δ . $O(e^{-\Delta' t})$ represents the neglected higher state corrections.

This method of matrix element extraction has a number of advantages to the more common technique in Sec. 3.3.2. First, the summation over the operator insertion time eliminates one of the independent variables of the correlation function. This reduction allows for clear identification of the excited state contamination. Without this summation, points with several source/sink separations and operator insertion time are required to visually demonstrate the presence of a plateau. A second advantage, demonstrated in [32], is that matrix element calculation can begin at much earlier times compared to sequential source calculations. This advantage is particularly important for physical mass calculations where the excited state effects require longer and longer time extents, T .

FIG. 3.1: The signal-to-noise ratio for a nucleon two point function at various Euclidean times. The plots on the left and right are from ensembles *a094m390* and *a094m390L* respectively which are described in Tab. 5.1. The different choices of momentum smearing parameter, $\zeta = 0$ in blue, $\zeta = 1.5$ in red, and $\zeta = 3.0$ in green, have different ranges of momenta where the signal-to-noise ratio is improved. As designed, the signal-to-noise ratio has a Gaussian shape in momentum space whose peak is shifted by changing the momentum smearing parameter ζ .

FIG. 3.2: The effective mass for a nucleon (left) and pion (right) two point functions at various momenta from ensemble $a094m390L$. The signal-to-noise ratio for the effective mass, particularly for the pion, decays without momentum smearing. Even without tuning ζ for each momentum, the momentum smearing procedure significantly improves the signal and allows for precision determination of high momentum effective masses which were not attainable with Gaussian smearing alone.

CHAPTER 4

Lattice Cross Sections

Since matrix elements with light-like separations cannot be directly calculated on a Euclidean lattice, alternatives to calculating the PDF needed to be constructed. An alternative approach has been the calculation of PDF moments. The OPE applied to the ITD expands the matrix element into certain local twist-2 matrix elements that can be related to the PDF moments through Eq. (2.31). The calculation of moments is hindered by the reduced rotational symmetry on the lattice. Since the lattice is only symmetric under certain rotations about the lattice axes, specifically the hypercubic rotational group, spin is no longer a quantum number. Instead there is a finite number of irreducible representations, which can be mapped to various continuum spin irreducible representations. For the operators relating to the higher PDF moments, there exists a power divergent mixing between operators with different mass dimension which renders the computation impractical [33, 34]. In principle, non-perturbative subtraction methods exist which can remove the power divergent mixing, but these methods are computationally expensive for the $a \rightarrow 0$ and $n \rightarrow \infty$ limits. There exist proposals of how to calculate these local operators by replacing the divergence in lattice spacing with a divergence in a smearing

length, which is kept finite in the continuum limit [35]. Without reliable calculations of sufficiently high moments, the full x dependence of the PDF cannot be reconstructed in a model independent way from local matrix element calculations [36].

Though there had existed earlier ideas, the seminal proposal for calculating the x dependence of the PDF is the quasi-PDF [37]. This proposal, described in more detail in Sec 4.1, involves the calculation of non-local matrix elements whose Fourier transform, the quasi-PDF, can be related to the PDF. This non-local matrix element involves fields with space-like separations and is therefore calculable on a Euclidean lattice. Importantly, it is guaranteed that if the matrix element only contains space-like separations, it is the same if calculated in Euclidean or Minkowski space [38]. The matching relationship between the quasi-PDF and PDF is calculated using Large Momentum Effective Theory (LaMET). LaMET is not an Effective Field Theory, but a large momentum expansion technique where one neglects terms subleading in powers of the hadron’s momentum. In the infinite momentum limit, the quasi-PDF can be related to the PDF through a factorization procedure. At finite momentum, the quasi-PDF is expanded in inverse powers of the hadron’s 3-momentum. The matching relationships have been derived to one-loop accuracy for various distributions [39, 40, 41, 42, 43]. The higher order terms in the p_3^{-1} expansion come from the neglected higher twist effects and some non-perturbative sources of these corrections have been studied in [44, 45, 46].

Afterwards many proposals were suggested on how to calculate PDFs from Lattice QCD. A number of them revolve around calculating what is called a “Good Lattice Cross Section” (LCS) [47, 48], or possibly its Fourier transform, such as the quasi-PDF method in [37] or the “OPE without OPE” proposal in [49]. A LCS is a matrix element which is calculable in Lattice QCD and has a factorization relationship with a parton distribution. These include the matrix element used to form the quasi-PDF and matrix elements of pairs of currents. The intrinsic scale of these matrix elements is the separation, z^2 , between the

fields. The $z^2 \rightarrow 0$ limit is singular. These matrix elements have the same divergence in this scale as the PDF has in its renormalization scale, μ^2 , but also contain contamination from higher twist effects proportional to powers of this z^2 scale. Utilizing this shared divergences, there exist factorization relationships between LCSs and the PDF. Two LCSs have been calculated, the pseudo-ITD [50], defined in Sec 4.4, and the antisymmetric combination of an axial-vector current and a vector current [51].

To be more specific, particular hadronic matrix elements qualify as LCSs if they have the following properties:

- They are a Lorentz covariant single-hadron matrix element computable on a Euclidean lattice.
- They have a well-defined continuum limit when the lattice spacing $a \rightarrow 0$.
- They or the functions in their Lorentz decompositions are factorizable to PDFs convoluted with infrared (IR)-safe hard coefficients plus small and controllable power corrections.

There exist a few methods which do not classify as LCSs [52, 49]. Many of these methods involve transforms of a LCS and approach the PDF through a different application of factorization. To illustrate the procedure of extracting a PDF through a factorization relationship, it is instructive to begin with the quasi-PDF and LaMET and then show factorization of a LCS with the Lorentz invariant space-like separation scale.

4.1 LaMET approach

The proposal of the quasi-PDF [37] begins by studying the operators in the OPE of the PDF in Eq. (2.29). For the particular light cone separation for the Wilson line,

the relevant operator is $\mathcal{O}^{+\cdots+}(\mu)$ whose matrix elements have no contribution from the trace terms, which have operators of lower mass dimension, and fewer Lorentz indices, multiplying factors of the metric tensor $g^{\mu_n\mu_m}$. For a lattice calculation, time independent operators are preferred. Instead consider the operator

$$\mathcal{O}^{\alpha 3 \dots 3} = \bar{\psi} \gamma^\alpha i D^3 \dots i D^3 \psi - \text{traces}, \quad (4.1)$$

whose matrix elements will include factors of the momentum p_3 and the trace terms proportional to the metric tensor,

$$\langle p | \bar{\psi} \gamma^\alpha i D^3 \dots i D^3 \psi | p \rangle_\mu = 2m_n(\mu) p^\alpha (p^3)^{n-1} + O\left(\frac{\Lambda_{\text{QCD}}^2}{p_3^2}, \frac{m^2}{p_3^2}\right). \quad (4.2)$$

The relationship between the Wilson line operator and the local matrix elements can be performed for this operator up to the higher twist approximations

$$\langle p | \mathcal{O}_{\text{WL}}^\alpha(z) | p \rangle_\mu = \sum_n C_n(\mu^2) z_\alpha (z_3)^{n-1} \langle p | \mathcal{O}^{\alpha 3 \dots 3} | p \rangle + O\left(\frac{\Lambda_{\text{QCD}}^2}{p_3^2}, \frac{m^2}{p_3^2}, \Lambda_{\text{QCD}}^2 z^2\right). \quad (4.3)$$

The Wilson line operator $\mathcal{O}_{\text{WL}}^\alpha(z)$ is given by

$$\mathcal{O}_{\text{WL}}^\alpha(z) = \bar{\psi}(z) \gamma^\alpha U(z; 0) \psi(0), \quad (4.4)$$

where γ^α is a choice of Dirac gamma matrix, τ^3 is flavor Pauli iso-spin matrix, and $U(z; 0)$ is a straight Wilson line between the space-like separated quark fields.

The Fourier transform of this operator, called the quasi-PDF is an off-light cone generalization of the PDF given by

$$\tilde{q}(y, \mu, p^3) = \int \frac{dz_3}{4\pi} e^{iyp^3 z_3} \langle p | \mathcal{O}_{\text{WL}}^\alpha(z) | p \rangle_\mu + O\left(\frac{\Lambda_{\text{QCD}}^2}{p_3^2}, \frac{m^2}{p_3^2}, \Lambda_{\text{QCD}}^2 z^2\right). \quad (4.5)$$

In the large p^3 limit, the quasi-PDF is factorizable to the PDF through

$$\tilde{q}(y, \mu, p_3) = \int_{-1}^1 \frac{dx}{|x|} C\left(\frac{x}{y}, \frac{\mu}{|x|p_3}\right) q(x, \mu) + O\left(\frac{\Lambda_{\text{QCD}}^2}{p_3^2}, \frac{m^2}{p_3^2}\right) + \int \frac{dz_3}{4\pi} e^{iyp^3 z_3} O(\Lambda_{\text{QCD}}^2 z^2). \quad (4.6)$$

The matching coefficient C , calculated using LaMET, depends on the particular choice of scheme used to calculate the quasi-PDF. In practice, the quasi-PDF has been calculated with the RI-MOM scheme and subsequently matched to $\overline{\text{MS}}$. It should be noted that the Wilson line matrix element in Eq. (4.5) can be replaced with a Renormalization Group Invariant (RGI) quantity, as will be done for the LCS analog. With this change, the matching relationship will be only dependent on the PDF's choice of scheme, typically $\overline{\text{MS}}$.

Fig. 4.1 show the current state-of-the-art calculations of the PDF from the quasi-PDF performed by the ETM collaboration [3] and the LP³ collaboration [4]. These calculations have been performed at physical pion mass and with maximum momenta of 1.38 GeV and 3.0 GeV respectively.

FIG. 4.1: PDFs, from quasi-PDFs, calculated at physical pion mass. The ETM collaboration [3] (left) and LP³ collaboration [4] (right) calculated the quasi-PDF with maximum momentum of 1.38 GeV and 3.0 GeV respectively.

4.2 Lattice Cross Sections and Factorization

Many proposals for calculating parton distributions from lattice QCD revolve around calculating matrix elements with space-like separations. In the quasi-PDF approach, Lorentz invariance is broken by the z_3 Fourier transform and the resulting quasi-PDF is subsequently factorized into the PDF. In the LCS approach, the matrix element itself, if it is Lorentz invariant, or a particular Lorentz projection, if it is Lorentz covariant, is what is factorized into the PDF. Unlike the LaMET analysis, the Lorentz invariant LCSs utilize the separation scale, z^2 , as the analog of the PDF factorization scale.

Considering at first only scalar matrix elements, the factorization of the matrix elements in question, $\mathcal{M}^i$, is given by

$$\langle p | \mathcal{O}^i(z) | p \rangle = \mathcal{M}^i(\nu, z^2) = \int_{-1}^1 dx \sum_a K_a^i(x\nu, \mu^2 z^2, \alpha_s) f_a(x, \mu^2) + O(m^2 z^2) + O(\Lambda_{\text{QCD}}^2 z^2). \quad (4.7)$$

where f_a is the PDF for a j parton and K_a^i are perturbatively calculable matching coefficients between the a parton PDF to the i matrix element. Two types of operators, which here are generically called $\mathcal{O}^i(z)$, for these matrix elements are the Wilson line operator from Eq. (4.4) and pairs of space-like separated currents. The proof of the factorization equation in Eq. (4.7) for the different matrix elements are the same.

Following the arguments in [35, 48], the OPE for matrix elements of the non-local $\mathcal{O}^i(z)$ for small but non vanishing z^2 is given by

$$\mathcal{M}^i(\nu, z^2) = \sum_{J=0} \sum_a C_{(J,a)}^i(z^2 \mu^2, \alpha_s) z_{\nu_1} \dots z_{\nu_J} \langle p | \mathcal{O}_{(J,a)}^{\nu_1 \dots \nu_J}(\mu^2) | p \rangle + O(z^2), \quad (4.8)$$

where $\mathcal{O}_{(J,a)}^{\nu_1 \dots \nu_J}(\mu^2)$ is a local traceless symmetric operator of spin J with renormalization scale μ and a is the label for the different operators of the same spin. These local matrix

elements have the following Lorentz decomposition

$$\langle p | \mathcal{O}_{(J,a)}^{\nu_1 \dots \nu_J}(\mu^2) | p \rangle = 2m_{(J,a)}(\mu^2) (p^{\nu_1} \dots p^{\nu_J} - \text{traces}) . \quad (4.9)$$

The reduced matrix elements $m_{(J,a)}(\mu^2)$ for a given spin J will have different mass dimensions. At a given spin J , the terms with the higher mass dimensional operators, which come from the trace terms, will be proportional to $(\Lambda_{\text{QCD}} z^2)^{n/2}$ when compared to operators which have n mass dimension fewer. This feature can be seen from the contraction of the indices, with Λ_{QCD} is a place holder for the mass of the hadron. The first approximation is to neglect these “power suppressed” higher dimensional operators which are subleading in the OPE. What remains are the local twist-2 operators which are the moments of the PDF

$$m_{(J,a)}(\mu^2) = \frac{1}{S_a} \int_{-1}^1 dx x^{J-1} f_a(x, \mu^2) , \quad (4.10)$$

where S_a is a symmetry factor which is 1 for quarks and 2 for gluons. It should be noted that there are no twist-2 scalar operators, so the spin is restricted to $J \geq 1$. Now, inserting Eq (4.9) into Eq (4.8), one finds

$$\mathcal{M}^i(\nu, z^2) = 2 \sum_{J=1} \sum_a C_{(J,a)}^i(z^2, \mu^2, \alpha_s) m_{(J,a)}(\mu^2) \sum_{i=0}^{\lfloor J/2 \rfloor} \binom{J-i}{i} \left(-\frac{m^2 z^2}{4} \right)^i \nu^{J-2i} , \quad (4.11)$$

where $\lfloor n \rfloor$ is the floor function, $\binom{n}{k}$ is the binomial coefficient, and m is the mass of the hadron. Once again, the terms subleading in the OPE, those proportional to powers of $m^2 z^2$, will be neglected. What remains is a relationship between the Ioffe time behavior of the LCSs and the moments of the PDF. One can identify the LCS matching kernel as

$$K_a^i(x\nu, \mu^2 z^2) = \sum_{J=1} \frac{2}{S_a} C_{(J,a)}^i(z^2 \mu^2, \alpha_s) (x\nu)^J , \quad (4.12)$$

which is the inverse Mellin transform of the Wilson coefficients from the OPE.

The OPE and the subsequent proof of factorization only work in a regime where the components of z all go to zero uniformly with other variables kept fixed. This region corresponds to $|\nu| \ll 1$, $|\Lambda_{\text{QCD}}^2 z^2| \ll 1$, and $|m^2 z^2| \ll 1$. The validity of the OPE requires the non-local matrix element, $\mathcal{M}(\nu, z^2)$, to be an analytic function of ν in the neighborhood of $\nu = 0$ with a Taylor expansion described by Eq. (4.11). With fixed and small z , ν can be modified by adjusting the hadron momentum p . In perturbation theory, changing the external momentum cannot introduce new divergences except at infinite momentum, or $\nu = \infty$. Therefore the non-local matrix element is an analytic function and the factorization formula in Eq. (4.11) is valid for all finite ν , up to the neglected power corrections $O(m^2 z^2)$ and $O(\Lambda_{\text{QCD}}^2 z^2)$.

It is crucial to mention that a large p alone does not guarantee the applicability of the operator product expansion of the matrix element and contributions from large z can invalidate the perturbative factorization, whether for the case of quark-antiquark fields linked by a Wilson line, or for the case of spatially-separated currents. The validity of perturbative factorization requires that the separation's scale, z^2 , be much smaller than the inverse square of typical hadronic scale, Λ_{QCD} , namely $\xi^2 \Lambda_{\text{QCD}}^2 \ll 1$. As can be seen in Eq. (4.6), the z_3 Fourier transform of the non-local matrix element includes the Fourier transform of the higher twist contaminations. The Fourier transform in z_3 defining the quasi-PDF mixes matrix elements with many different z^2 scales, both in perturbative and non-perturbative regimes. The large momentum limit, $p_z \rightarrow \infty$, does not necessarily force these terms to be small.

The coordinate space approach provides distinct theoretical advantages. One can

write the Fourier transform of $\mathcal{M}^i(\nu, z^2)$:

$$\widetilde{\mathcal{M}}^i(\widetilde{\omega}, q^2) \equiv \int \frac{d^4 z}{z^4} e^{iq \cdot z} \mathcal{M}^i(\nu, z^2), \quad (4.13)$$

where corresponding $\mathcal{O}^i$ can be any operator defined in Eq. (4.7), and $\widetilde{\omega} \equiv \frac{2p \cdot q}{-q^2} = \frac{1}{x_B}$ with x_B the Bjorken variable for the lepton-hadron DIS. However, though q is related to z through the Fourier transform above, it is not a one-to-one relation. $\widetilde{\mathcal{M}}^i$ with small or large values of q will receive contribution from $\mathcal{M}^i$ with all values of z , small and large. In particular, it can involve contributions from values of $z \gg \frac{1}{\Lambda_{\text{QCD}}}$, thereby violating factorization. This is why, in the LCSs approach, z is a very well-defined quantity, analogous to a hard probe of hadron structure in a DIS experiment. It has also been demonstrated in Ref. [48] that, the non-analytic cut of $\widetilde{\mathcal{M}}^i$ comes from the integration region of large z . That is, even if we demand $|q^2| \gg \Lambda_{\text{QCD}}^2$, $\widetilde{\mathcal{M}}^i$ in momentum space can always receive contribution from large z region so long as $\widetilde{\omega}^2 > 1$. On the other hand, in coordinate space, if we fix z to be short-distance, we do not have contribution from the large z region and thus $\mathcal{M}^i$ has a good analytic behavior. It should be noted that not all Fourier transforms of LCSs suffer from this potential pitfall. The ν Fourier transform, $P(x, z^2)$, defined by

$$\mathcal{M}(\nu, z^2) = \int_{-\infty}^{\infty} dx e^{i\nu x} P(x, z^2), \quad (4.14)$$

does not have the issues of z Fourier transforms. The mixing of scales is completely avoided by considering this object instead.

4.3 OPE without OPE

Though the ultimate goal is to determine the full x dependence of the PDF, as will be shown in Chapter 7, large ranges of Ioffe time may be required an accurate calculation, particularly in the low x region. On the other hand, the low Ioffe time behavior of a LCS gives access to the PDF moments by using Eq. (4.11). The calculation of a set of local matrix elements through a non-local matrix element is not a new concept. It had been previously proposed for the calculation of hadronic Weak interactions [53], where it was coined “OPE without OPE”, and for extracting PDF moments from a hadronic tensor [49].

The Taylor expansion with respect to ν of the LCS, $\mathcal{M}(\nu, z^2)$ is expressed as

$$\mathcal{M}(\nu, z^2) = 1 + c_1(z^2)\nu + \frac{c_2(z^2)}{2}\nu^2 + \dots = \int_{-\infty}^{\infty} dy (1 + iy\nu - \frac{y}{2}\nu^2 + \dots) P(y, z^2). \quad (4.15)$$

The Taylor coefficients of the LCS, $c_n(z^2)$, are, up to the factors of $n!$ and powers of i , the moments of the LCS’s inverse Fourier transform, $b_n(z^2) = \int dy y^n P(y, z^2)$, which will be called the pseudo-moments. The pseudo-moments are given as i weighted derivatives of the LCS at $\nu = 0$ which when applied to Eq. (4.11) relates them to the PDF moments,

$$b_n(z^2) = (-i)^n \frac{\partial^n \mathcal{M}(\nu, z^2)}{\partial \nu^n} \Big|_{\nu=0} = \sum_a C_{(n,a)}^i(z^2 \mu^2) m_{(n,a)}(\mu^2) + O(z^2), \quad (4.16)$$

where all terms in the sum over i with $i \geq 1$ are collected into the higher twist corrections $O(z^2)$. After matching the pseudo-moments to the PDF moments with the leading twist formula, the residual z^2 dependence can be used as a signal of the size of higher twist effects in the LCS.

Note that this extraction of the low moments needs a precise computation of $\mathfrak{M}(\nu, z^2)$ in the small ν region at fixed z^2 . This is the region in which lattice computations can easily achieve high precision. In principle, this procedure will allow for the calculation of any

PDF moment. Unfortunately, the extraction of higher PDF moments require sufficiently precise data in a region of Ioffe time where the higher order coefficients in the Taylor expansion are resolvable. As will be shown, the limitations of a lattice calculation will restrict the signal for most PDF moments.

4.4 Wilson line operator and pseudo-PDFs

The first LCS considered [50] was a matrix element of the Wilson line operator in Eq. (4.4). The original goal was to calculate the Parton pseudo-Distribution Function (pseudo-PDF), which is the Fourier transform of this LCS as defined in Eq. (4.14). The pseudo-PDF was then to be extrapolated to the PDF in the $z^2 \rightarrow 0$ limit. This original procedure is modified in this presence of the scale dependence of the PDF. The $z^2 \rightarrow 0$ limit contains a logarithmic divergence generated by the DGLAP evolution of the PDF. Instead of a simple extrapolation to $z \rightarrow 0$, a factorization formula is calculated [54] and the procedure for LCSs is used.

4.4.1 Pseudo-distributions

The unpolarized ITD is described in terms of a special case of the helicity averaged, renormalized, forward, non-local matrix element,

$$M^\alpha(p, z) = \langle p | \bar{\psi}(z) \gamma^\alpha U(z; 0) \psi(0) | p \rangle_\mu, \quad (4.17)$$

for $p = (p^+, \frac{m^2}{2p^+}, 0_T)$, $z = (0, z_-, 0_T)$, and $\alpha = +$. Given arbitrary choices of p , z , and α , the Lorentz decomposition of the matrix element in Eq. (4.17)

$$M^\alpha(z, p) = 2p^\alpha \mathcal{M}(\nu, z^2) + 2z^\alpha \mathcal{N}(\nu, z^2). \quad (4.18)$$

For the choice of parameters which corresponds to the ITD, only $\mathcal{M}$ contributes to the matrix element. The term in the Lorentz decomposition containing the most powers of p is typically the leading twist contribution. For arbitrary z^2 , the $\mathcal{M}$ function, called the Ioffe time pseudo-distribution function or pseudo-ITD, can be thought as a generalization of the ITD to separations other than light-like. The pseudo-ITD contains the leading twist contributions, but also contains higher twist contributions at $O(z^2\Lambda_{\text{QCD}}^2)$. The removal of these contributions, either through cancellation or small z^2 , is necessary for accurately determining the ITD from the small z^2 limit of the pseudo-ITD.

It can be shown [44] that the Fourier transform of the pseudo-ITD with respect to ν will be a function of $x \in [-1, 1]$. This feature defines the pseudo-PDF

$$\mathcal{M}(\nu, z^2) = \int_{-1}^1 dx e^{i\nu x} P(x, z^2). \quad (4.19)$$

In order to improve the calculation of the ITD from the pseudo-ITD, it has been suggested [50] to remove $O(z^2\Lambda_{\text{QCD}}^2)$ by considering the reduced pseudo-ITD

$$\mathfrak{M}(\nu, z^2) = \frac{\mathcal{M}(\nu, z^2)}{D(z^2)}. \quad (4.20)$$

Ideally, $D(z^2)$ will contain all the non-trivial z^2 dependence from higher twist effects. The chief requirement is that in the $z^2 \rightarrow 0$ limit $D(z^2)$ approaches a non-zero finite constant. This feature ensures that the OPE for this reduced pseudo-ITD will be the same as the original pseudo-ITD and factorization will lead to the same PDF moments. For future purposes in a lattice calculation, the choice of $D(z^2) = \mathcal{M}(0, z^2)$ is particularly useful and in the following this choice will be assumed.

In order to take a continuum limit, the operator O_{WL} must first be renormalized. In QCD, the renormalization constant of a Wilson line is proportional to $e^{-gm_a^z + c(z)}$ which

when expanded in g gives a power divergence in perturbation theory and $c(z)$ is a term which depends on the number of cusps in the Wilson line and the angles they form [55, 56]. The full Wilson line operator O_{WL} is multiplicatively renormalizable [57]. This convenient feature allows the ratio in Eq. (4.20), with the choice $D(z^2) = \mathcal{M}(0, z^2)$, to have a finite continuum limit. The renormalization constants in the numerator and denominator, which only depend on the length and shape of the Wilson line, not the external momentum, cancel exactly.

z^2 evolution and $\overline{\text{MS}}$ matching

All phenomenological PDF fits are performed in the $\overline{\text{MS}}$ scheme either with NLO or NNLO truncation in the perturbative series of the Wilson Coefficients. Since the reduced pseudo-ITD is RGI, its z^2 dependence is independent of any particular scheme. The perturbative z^2 evolution equation is given by [45]

$$z^2 \frac{d\mathfrak{M}}{dz^2}(\nu, z^2) = -\frac{\alpha_s}{2\pi} C_F \int_0^1 du B(u) \mathfrak{M}(\nu, z^2), \quad (4.21)$$

where $B(u) = \left[\frac{1+u^2}{1-u} \right]_+$ is the Altarelli-Parisi kernel. This equation is an analog of the PDF's DGLAP evolution equation. Solving Eq. (4.21) in the leading log approximation, as used in [31], the reduced pseudo-ITD at various scales z^2 can be evolved to the common scale z_0^2 by applying

$$\mathfrak{M}(\nu, z_0^2) = \mathfrak{M}(\nu, z^2) + \ln\left(\frac{z^2}{z_0^2}\right) \frac{\alpha_s C_f}{2\pi} B \otimes \mathfrak{M}(\nu, z^2), \quad (4.22)$$

where

$$B \otimes \mathfrak{M}(\nu, z^2) = \int_0^1 du \left[\frac{1+u^2}{1-u} \right]_+ \mathfrak{M}(u\nu, z^2). \quad (4.23)$$

As was done in [31], one can estimate the effects of evolution and matching by per-

forming the convolution on a model reduced pseudo-ITD. Consider the pseudo-ITD for the pseudo-PDF $P(x) = x^a(1-x)^b/B(a+1, b+1)$ for $a = 0.5$ and $b = 3$. The convolutions for the DGLAP kernel are shown in Fig. 4.2. For the real component, perturbative z^2 evolution will always increase the value as z^2 increases. For the imaginary component, when z^2 increases the value will increase if $\nu \lesssim 5.5$ and it will decrease otherwise. After the reduced pseudo-ITD has been evolved to a single z^2 with Eq. (4.22), then the $\overline{\text{MS}}$ matching relationship can be applied.

FIG. 4.2: The convolutions for evolution and matching

This kernel $\mathcal{K}$ for matching the pseudo-ITD to the $\overline{\text{MS}}$ ITD has been calculated at

NLO [58, 59, 42]

$$\mathcal{K}(u, z^2 \mu^2, \alpha_s) = \delta(1 - u) + \frac{\alpha_s C_F}{2\pi} \left[\ln(z^2 \mu^2 \frac{e^{2\gamma_E+1}}{4}) B(u) + L(u) \right], \quad (4.24)$$

where

$$L(u) = \left[4 \frac{\ln(1-u)}{1-u} - 2(1-u) \right]_+ \quad (4.25)$$

is a scale independent piece related to the specific choice of the $\overline{\text{MS}}$ scheme. This gives the full NLO matching relationship

$$\mathfrak{M}(\nu, z^2) = Q(\nu, \mu^2) + \frac{\alpha_s C_F}{2\pi} \int_0^1 du \left[\ln(z^2 \mu^2 \frac{e^{2\gamma_E+1}}{4}) B(u) + L(u) \right] Q(u\nu, \mu^2). \quad (4.26)$$

The convolution of $L(u)$ is for the pseudo-PDF model is shown in Fig. 4.2. The scale dependent part of the kernel can be identified as the previously mentioned evolution process to a scale $z_0^2 = 4e^{-2\gamma_E-1}\mu^{-2}$. The reduced pseudo-ITD at many scales could be directly matched to the $\overline{\text{MS}}$ ITD in a single step. The separation of the evolution and matching procedure can allow for the two steps to be performed at different orders in α_s or to take into account the higher twist contamination in different ways.

The Mellin moments of this kernel also allow the matching between the pseudo-moments and the $\overline{\text{MS}}$ PDF moments. As given in [60], the Wilson coefficients in Eq. (4.8) are given by

$$C_n(z^2 \mu^2, \alpha_s) = 1 - \frac{\alpha_s}{2\pi} C_F \left[\gamma_n \ln \left(z^2 \mu^2 \frac{e^{2\gamma_E+1}}{4} \right) + l_n \right], \quad (4.27)$$

where

$$\gamma_n = \int_0^1 du B(u) u^n = \frac{3}{2} - \frac{1}{1+n} - \frac{1}{2+n} - 2 \sum_{k=1}^n \frac{1}{k}, \quad (4.28)$$

which are the well known moments of the Altarelli-Parisi kernel, and

$$l_n = \int_0^1 du L(u) u^n = 2 \left[\left(\sum_{k=1}^n \frac{1}{k} \right)^2 + \frac{2\pi^2 + n(n+3)(3+\pi^2)}{6(n+1)(n+2)} - \psi^{(1)}(n+1) \right]. \quad (4.29)$$

Here $\psi^{(1)}(z)$ is the polygamma function defined as $\psi^{(1)}(z) = d^2 \ln \Gamma(z) / dz^2$ with $\Gamma(z)$ being the Γ -function. With the Wilson coefficients computed we can now obtain the $\overline{\text{MS}}$ moments up to $\mathcal{O}(\alpha_s^2, z^2)$ directly from the reduced function $\mathfrak{M}(\nu, z^2)$ as

$$m_n(\mu^2) = (-i)^n \frac{1}{C_n(z^2 \mu^2, \alpha_s)} \left. \frac{\partial^n \mathfrak{M}(\nu, z^2)}{\partial \nu^n} \right|_{\nu=0} + O(z^2 \Lambda_{\text{QCD}}^2, \alpha_s^2). \quad (4.30)$$

4.5 Two current operators and pseudo-Structure Functions

The LCS formalism attempts to create a similar advantage by considering not just the Wilson line quark bilinear, but also a plethora of current-current correlations. Each of these pairs of currents have different perturbatively calculable kernels. A combined PDF analysis of many current-current correlations, and possibly including the reduced pseudo-ITD, can be performed just as the global PDF fits combine many experimental results.

The renormalization of the current-current correlation matrix elements is theoretically less complicated than the pseudo-ITD. The renormalization constants are simply those for the local current constants, which may even have been previously calculated for the desired ensemble of lattices. There are no potential complications with multiple scales, a power divergence, or mixings with other matrix elements which can occur for the Wilson line operators. This advantage has previously been noted in [61] where matrix elements of two

local vector currents, which already have a finite continuum limit, are calculated in order to determine the hadronic tensor through Eq. (2.3). Their method differs from the LCS formalism by allowing the currents to be separated in the Euclidean time direction as well as spatial directions. In Euclidean space, the integral over the temporal separation is a Laplace transform, instead of the Fourier transform in Minkowski space.

In [48], a number of scalar matrix elements were proposed and the leading order tree-level kernels were calculated. These matrix elements are of the form

$$\mathcal{M}_n(\nu, z^2) = \langle p | \mathcal{O}_n(z) | p \rangle = \langle p | J_i(z) J_j(0) | p \rangle \quad (4.31)$$

where J_i and J_j are renormalized scalar currents or have their Lorentz indices contracted. For example, possible current pairs are

$$\mathcal{O}_S(z) = z^4 Z_S^2 [\bar{q}q](z) [\bar{q}q](0), \quad (4.32a)$$

$$\mathcal{O}_V(z) = z^2 Z_V^2 [\bar{q}\not{z}q](z) [\bar{q}\not{z}q](0), \quad (4.32b)$$

$$\mathcal{O}_{\tilde{V}}(z) = -\frac{z^4}{2} Z_V^2 [\bar{q}\gamma_\alpha q](z) [\bar{q}\gamma^\alpha q](0), \quad (4.32c)$$

$$\mathcal{O}_{V'}(z) = z^2 Z_{V'}^2 ([\bar{q}\not{z}Q](z) [\bar{Q}\not{z}q](0) + [\bar{Q}\not{z}q](z) [\bar{q}\not{z}Q](0)), \quad (4.32d)$$

where Z_i are the renormalization constants for the local currents. The powers of the scale z^2 are to make the operators dimensionless. Each of these currents follows the factorization formula in Eq. (4.7). It should be noted that currents can also be flavor changing, such as $\mathcal{O}_{V'}(z)$. The flavor changing has a cost of the matrix element overlapping with the PDFs of both quarks. If one of the quarks is such that its PDF is small, such as being a heavy quark in a light hadron, then this cost is not a significant issue. For the remainder of this work, if an auxiliary quark is used, it will be assumed that its mass is larger than the light quark masses, but its inverse is smaller than the typical separation scales, $m_{u,d} < m_Q \ll z^{-1}$.

Even though the matching kernels $\mathcal{K}$ are dependent on the particular operator in the LCS, they do not depend on the external states. The case is the opposite for the PDFs which depend on the particular hadron, but not on the operator in the LCS. For experimental cross sections, this situation is the same for the process-dependent Wilson coefficients and process-independent PDFs. In order to determine the leading order matching kernel, one considers the matrix element for asymptotic on-shell quark states instead of a hadron state. At leading order in perturbation theory, the “quark inside a quark PDF” is given by the delta function, $f_q^{(0)}(x, \mu^2) = \delta(1 - x)$ and there is no scale dependence at zeroth order in α_s . Inserting this PDF into Eq. (4.7) one finds

$$\mathcal{M}_q^{n(0)}(\nu, z^2) = \mathcal{K}(\nu, \mu^2 z^2) + O(z^2). \quad (4.33)$$

At the leading twist and lowest order in perturbation theory, this matrix element requires calculating the Feynman diagrams in Fig. 4.3. The two terms in $\mathcal{O}_{V'}(z)$ are such that the flavor changing set of currents have the same Feynman diagrams as the flavor conserving sets of currents.

FIG. 4.3: The lowest order Feynman diagrams contributing to the 2 current LCSs in Eq. (4.31) for an asymptotic on-shell quark state of momentum $k = xp$.

Up to now only scalar matrix elements have been taken into account, but the leading order matching kernels for all possible currents could be of interest. Consider the spin averaged quark matrix element of two generic local currents, ignoring the coefficients and

powers of z^2 ,

$$\mathcal{M}_{ij}^{(0)} = \frac{1}{2} \sum_s \langle q(k, s) | [\bar{q}\Gamma_i q](z) [\bar{q}\Gamma_j q](0) | q(k, s) \rangle = \mathcal{M}_{ij}^{(a)} + \mathcal{M}_{ij}^{(b)} \quad (4.34)$$

where i and j label the $\Gamma_{i,j}$ which can be any element of the Dirac algebra, $|q(k, s)\rangle$ is a quark state with momentum k and spin s , and $\mathcal{M}_{ij}^{(a)}$ and $\mathcal{M}_{ij}^{(b)}$ represent the two Wick contractions shown in Figs. 4.3a and 4.3b respectively. Note that i and j are labels for the Γ structure and not Lorentz indices. The first Wick contraction in Fig. 4.3a is given by

$$\begin{aligned} \mathcal{M}_{ij}^{(a)} &= \frac{1}{2} \sum_s \bar{u}^s(k) e^{ik \cdot z} \Gamma_i G(z; 0) \Gamma_j u^s(k) \\ &= \frac{1}{2} e^{ik \cdot z} \text{Tr} \left[\not{k} \Gamma_i \int \frac{d^4 q}{(2\pi)^4} \frac{i \not{q}}{q^2} e^{-iq \cdot z} \Gamma_j \right] \\ &= \frac{i}{4\pi^2} \frac{k^\alpha z^\beta}{z^4} T_{\alpha\beta}^{ij} e^{ik \cdot z} \end{aligned} \quad (4.35)$$

where $G(z; 0)$ is the position space quark propagator, the dependence on the choice of current is collected into the trace $T_{\alpha\beta}^{ij} = \text{Tr}[\gamma_\alpha \Gamma_i \gamma_\beta \Gamma_j]$, and $k = xp$ was used. Depending on the particular choice of Γ_i and Γ_j , the trace factor, $T_{\alpha\beta}^{ij}$ could have 0 to 4 additional Lorentz indices. The second Wick contraction in Fig. 4.3b is given by

$$\begin{aligned} \mathcal{M}_{ij}^{(b)} &= \frac{1}{2} \sum_s \bar{u}^s(k) \Gamma_j G(0; z) \Gamma_i e^{-ik \cdot z} u^s(k) \\ &= \frac{1}{2} e^{-ik \cdot z} \text{Tr} \left[\not{k} \Gamma_j \int \frac{d^4 q}{(2\pi)^4} \frac{i \not{q}}{q^2} e^{iq \cdot z} \Gamma_i \right] \\ &= -\frac{i}{4\pi^2} \frac{k^\alpha z^\beta}{z^4} T_{\alpha\beta}^{ji} e^{-ik \cdot z}. \end{aligned} \quad (4.36)$$

Combining Eqs. (4.35) and (4.36) and writing the quark momentum as $k = xp$, one gets the leading order, which is denoted by the superscripts (0), quark matrix element and

matching kernel

$$\mathcal{M}_{ij}^{(0)}(p, z) = \frac{ix}{4\pi^2} \frac{p^\alpha z^\beta}{z^4} [T_{\alpha\beta}^{ij} e^{i\nu x} - T_{\alpha\beta}^{ji} e^{-i\nu x}] . \quad (4.37)$$

The traces for the operators given in Eq. (4.32a), where the two scalar currents are the same, are such that $T^{ij} = T^{ji}$. Inserting the appropriate Γ and contractions for these operators, one can see that they all have the same leading order kernel

$$\mathcal{K}^{(0)}(x\nu) = \frac{i\nu x}{\pi^2} (e^{i\nu x} - e^{-i\nu x}) . \quad (4.38)$$

for the operators $\mathcal{O}_{S,V,\tilde{V},V'}$. The factor of $-\frac{1}{2}$ in the definition $\mathcal{O}_{\tilde{V}}$ was included so that this leading order kernel would be the same as the kernels for the other operators. This leading order kernel is also the same when all of the vector currents are replaced with axial-vector currents to define $\mathcal{O}_{A,\tilde{A},A'}$. Through the traces $T^{SS} = -T^{PP}$, one can see that for a pair of pseudoscalar currents the leading order matching kernel is $\mathcal{K}_P^{(0)} = -\mathcal{K}_S^{(0)}$. By performing the same analysis as in Eq. (2.22), one can see that these current combinations will give access to the combination $q(x) + \bar{q}(x)$, but does not allow for direct calculation of the valence distribution.

So far, the only two-current LCSs that have been considered are scalar operators. When the Γ structures have uncontracted Lorentz indices, one must consider the Lorentz decomposition of the matrix element, just as when defining the pseudo-ITD. The Lorentz invariant functions, which will be called the Ioffe time pseudo-Structure Functions (pseudo-ITSF), are the LCSs which will be factorized into the PDF. As an example, take the matrix element of two separated vector currents, given by $\Gamma_i = \gamma^\mu$ and $\Gamma_j = \gamma^\nu$, which is the

position space analog of the hadronic tensor, given by

$$\begin{aligned}\mathcal{M}^{\mu\nu} = \langle p | J_V^\mu(z) J_V^\nu(0) | p \rangle &= \frac{p^\mu z^\nu + p^\nu z^\mu}{\nu} \mathcal{M}_{pz}(\nu, z^2) + g^{\mu\nu} \mathcal{M}_g(\nu, z^2) \\ &+ \frac{p^\mu p^\nu}{p^2} \mathcal{M}_{pp}(\nu, z^2) + \frac{z^\mu z^\nu}{z^2} \mathcal{M}_{zz}(\nu, z^2)\end{aligned}\quad (4.39)$$

where $J_V^\mu(x) = [\bar{q}\gamma^\mu q](x)$ and $\mathcal{M}_j$ are the pseudo-ITSF. This particular Lorentz decomposition actually contains a redundancy. In momentum space, the hadronic tensor satisfies a Ward identity, $q_\mu H^{\mu\nu} = 0 = q_\nu H^{\mu\nu}$, which allows for the elimination of some of the SF. Unfortunately the Ward identity in position space, $\partial_\mu \mathcal{M}^{\mu\nu} = 0 = \partial_\nu \mathcal{M}^{\mu\nu}$, is more complicated to apply since one generally does not know the z^μ derivative of the matrix element. Regardless, one can still continue with the extra pseudo-ITSF.

For this case, the leading order matrix element in Eq. (4.37) is given by

$$\mathcal{M}_{\mu\nu}^{(0)} = \frac{i\nu x}{\pi^2} (e^{i\nu x} - e^{-i\nu x}) \left(\frac{1}{\nu} (p_\mu z_\nu + p_\nu z_\mu) - g_{\mu\nu} \right) \quad (4.40)$$

which gives the leading order matching kernels for the 4 pseudo-ITSFs, T_i , as

$$\begin{aligned}\mathcal{K}_{pz}^{(0)}(x\nu) &= \frac{ix\nu}{\pi^2} (e^{ix\nu} - e^{-ix\nu}) \\ \mathcal{K}_g^{(0)}(x\nu) &= -\frac{ix\nu}{\pi^2} (e^{ix\nu} - e^{-ix\nu}) \\ \mathcal{K}_{pp}^{(0)}(x\nu) &= 0 \\ \mathcal{K}_{zz}^{(0)}(x\nu) &= 0.\end{aligned}$$

Just as for the isolation of the pseudo-ITD from the Lorentz decomposition of the Wilson line operator matrix element, a particular choice of p , z , μ and ν can be made in order to isolate one of the pseudo-ITSFs with a non-zero leading order kernel and subsequently calculate the PDF from it.

So far, all of the leading order two-current matching kernels mentioned have all been real functions, unlike the pseudo-ITD matching kernel. Specifically, they have been proportional to $\sin(\nu x)$. Following the analysis in Eq. (2.22), this sine kernel will allow these LCSs to give access to the CP odd combination $q_+(x) = q_v(x) + 2\bar{q}(x)$. This feature is unfortunate, because with any of the previously mentioned pairs of currents one cannot isolate the valence quark PDF from the anti-quark PDF as one can with the real and imaginary components of the pseudo-ITD.

4.5.1 Axial Vector - Vector combination

The hadronic tensor for Weak currents instead of the typical Electromagnetic currents, which is observable in unpolarized neutrino DIS, is decomposed with an additional SF

$$H^{\mu\nu}(p, q) = \left(-g^{\mu\nu} + \frac{q^\mu q^\nu}{q^2}\right) F_1(x, Q^2) + \frac{(p^\mu - q^\mu \frac{p \cdot q}{q^2})(p^\nu - q^\nu \frac{p \cdot q}{q^2})}{p \cdot q} F_2(x, Q^2) + i\epsilon^{\mu\nu\alpha\beta} \frac{q_\alpha p_\beta}{2p \cdot q} F_3(x, Q^2) \quad (4.41)$$

Just like for F_1 and F_2 , this new structure function, in the parton model, is a linear combination of the quark PDFs. For W^+ scattering, the SF is given by

$$F_3^{W^+} = 2[d(x) - \bar{u}(x) + s(x) - \bar{c}(x) + \dots] \quad (4.42)$$

and for W^- scattering it is given by

$$F_3^{W^-} = 2[u(x) - \bar{d}(x) + c(x) - \bar{s}(x) + \dots] . \quad (4.43)$$

The sum of F_3 for W^+ and W^- scattering will be proportional to the sum of the valence quark PDFs. This parity violating SF arises from the matrix elements of mixed vector and

axial vector currents. As will be shown below, the Lorentz decomposition of the mixed space-like separated vector and axial vector current matrix elements will be a LCS for the valence quark PDF.

Since the Strong interaction is invariant under parity $\mathcal{P}$ and time reversal $\mathcal{T}$ transformations, the matrix elements defining a LCS should have the property

$$\mathcal{M}_n = \langle p | [O_n(z)]^\dagger | p \rangle = \langle p | (\mathcal{PT}) [O_n(z)]^\dagger (\mathcal{PT})^{-1} | p \rangle. \quad (4.44)$$

Consider the case of two local quark bilinear currents, given in Eq. (4.31). The currents $J_i^\mu(x)$ and $J_j^\nu(x)$ have the transformation property $(\mathcal{PT})J_i^\mu(x)(\mathcal{PT})^{-1} = \eta_i J_i^\mu(-x)$ and $\eta_i = \pm 1$. Using translational symmetry, the property in Eq. (4.44) can be written

$$\mathcal{M}_{ij}^{\mu\nu} = \eta_i \eta_j \mathcal{M}_{ji}^{\nu\mu}. \quad (4.45)$$

For the case of the vector and axial-vector currents, $\eta_V = 1$ and $\eta_A = -1$. One can see that the combination $\tilde{\mathcal{M}}^{\mu\nu} = \frac{1}{2}(\mathcal{M}_{AV}^{\mu\nu} + \mathcal{M}_{VA}^{\mu\nu})$ will be antisymmetric in the indices μ and ν . There exist only two antisymmetric Lorentz structures which can be made out of the two available vectors p^μ and z^μ . This combination of current pairs has the Lorentz decomposition

$$\tilde{\mathcal{M}}^{\mu\nu}(p, z) = \frac{\epsilon^{\mu\nu\alpha\beta} p_\alpha z_\beta}{\nu} \mathcal{T}_1(\nu, z^2) + \frac{p^\mu z^\nu - p^\nu z^\mu}{\nu} \mathcal{T}_2(\nu, z^2). \quad (4.46)$$

Once again, in order to find the leading order kernel, the gamma structures are inserted into the traces

$$T_{\alpha\beta}^{VA} = \text{Tr}[\gamma_\alpha \gamma_\mu \gamma_\beta \gamma_\nu \gamma_5] = 4i\epsilon_{\mu\nu\alpha\beta} = -T_{\alpha\beta}^{AV} = -\text{Tr}[\gamma_\alpha \gamma_\mu \gamma_5 \gamma_\beta \gamma_\nu]. \quad (4.47)$$

Inserting the traces into Eq.(4.37), one finds the leading order matching coefficients

$$\begin{aligned}\tilde{K}_1(x\nu) &= \frac{x\nu}{\pi^2} (e^{ix\nu} + e^{-ix\nu}) \\ \tilde{K}_2(x\nu) &= 0.\end{aligned}\tag{4.48}$$

An important feature of this combination, unlike the cases where the two currents are the same, is that the traces from the two contractions have different signs. Following the analysis in Eq. (2.22) this feature will isolate the valence quark PDF, but does not give access to the antiquark distribution.

The cases of the axial vector-vector and the vector-vector current combinations demonstrate the advantage, and even necessity, of dealing with multiple LCSs. By considering only one of these two LCSs, there only exists sensitivity to a single combination of the PDF, either $q + \bar{q}$ or $q - \bar{q}$. Only by taking into account both LCSs, just as with the real and imaginary components of the pseudo-ITD, one can extract individually $q_v(x)$ and $\bar{q}(x)$.

CHAPTER 5

Calculation of pseudo-ITD

The numerical calculation of the reduced pseudo-PDF is performed on three different ensembles of lattice QCD configurations. The ensembles that are used in this article were generated by the JLab/W&M collaboration [1] employing 2+1 flavors of clover Wilson fermions and a tree-level tadpole-improved Symanzik gauge action. The ensembles were generated using a rational Hybrid Monte Carlo (RHMC) update algorithm. [62] In the fermionic action, one iteration of stout smearing was used with the weight $\rho = 0.125$ for the staples. A direct consequence of this smearing is that the tadpole corrected tree-level clover coefficient c_{SW} used is very close to the non-perturbative value determined, a posteriori, employing the Schrödinger functional method [1]. The parameters used for the three ensembles are listed in Table 5.1. The lattice spacings, a , are estimated using the Wilson flow scale w_0 described in Ref [63].

As mentioned in Sec. 4.4.1, a judicious choice of α , p , and z will isolate the pseudo-ITD, otherwise leading to a higher twist contamination from $\mathcal{N}$. Another potential difficulty to avoid the mixing of renormalization constants of the Wilson line quark bilinear and a higher twist scalar operator [64, 65]. If the separation has a non zero z^α component, the

ID	$a(\text{fm})$	$M_\pi(\text{MeV})$	β	c_{SW}	am_l	am_s	$L^3 \times T$	N_{cfg}
$a127m415$	0.127(2)	415(23)	6.1	1.24930971	-0.2800	-0.2450	$24^3 \times 64$	2147
$a127m415L$	0.127(2)	415(23)	6.1	1.24930971	-0.2800	-0.2450	$32^3 \times 96$	2560
$a094m390$	0.094(1)	390(71)	6.3	1.20536588	-0.2350	-0.2050	$32^3 \times 64$	417

TABLE 5.1: Parameters for the lattices generated by the JLab/W&M collaboration [1] using 2+1 flavors of clover Wilson fermions and a tree-level tadpole-improved Symanzik gauge action. The lattice spacings, a , are estimated using the Wilson flow scale w_0 . Stout smearing implemented in the fermion action makes the tadpole corrected tree-level clover coefficient c_{SW} used to be very close to the value when determined non-pertubatively with the Schrödinger functional method. The $a127m415L$ contains 10 independent streams of 256 configurations each. The other ensembles only contain a single stream each.

renormalization of the matrix element is no longer multiplicative. Considering Eq. (4.18), the choices of $\alpha = 4$, $p = (0, 0, p_3, E)$, and $z = (0, 0, z_3, 0)$ are taken for the following calculation, in order to avoid both the higher twist contamination from the term with $\mathcal{N}$ in Eq. (4.18) and the mixing of matrix elements under renormalization.

5.1 Method of Calculation

The matrix elements for the reduced pseudo-ITD is calculated using the Feynman Hellmann summation method described in Sec. 3.3.3. The three point correlation function is calculated using the sequential operator technique. In this method, the sequential operator is a generalized propagator, created from quark propagators, with the operator of interest inserted at all time slices. To construct the three point function, the sequential operator replaces a single regular propagator in what would otherwise be a regular two point function calculation. An advantage of this technique is, for a given operator, the three point function can be calculated for all possible interpolating fields and at all possible sink times without costly additional Dirac matrix inversions. On the other hand, a disadvantage of this technique is that each new operator requires another Dirac matrix inversion.

The sequential operator, H , is defined as the solution to the system of equations, with suppressed spin and color indices

$$\sum_{x,t} D(y, s; x, t) H(O_{\text{op}}, \vec{q}; x, t; x_0, t_0) = e^{i\vec{q} \cdot \vec{y}} O_{\text{op}}(y, s) G(y, s; x_0, t_0), \quad (5.1)$$

where $D(y, s; x, t)$ is the Dirac matrix, $\vec{q}$ is a transferred momentum, and $G(y, s; x_0, t_0)$ is the point-to-all propagator from a randomly chosen source point (x_0, t_0) . The three point function is calculated by replacing a single propagator in a two point function calculation with the sequential operator. The sequential operator contains a sum over $\vec{y}$ and s which project the operator onto a definite momentum transfer $\vec{q}$ and inserts the operator at all time slices respectively. For the case of the reduced pseudo-ITD, which is defined by a forward matrix element, there will be no momentum transferred, but a future calculation of a GPD will require the phase factor in Eq. (5.1).

In the calculation of the two point and three point functions for a given momentum smearing parameter ζ , there will be two choice of smearing procedure. The quarks in the source interpolating fields will always be smeared, but the quarks in the sink interpolating fields will either be smeared, called smeared-smeared (SS), or left as points, called smeared-point (SP). Typically the SS correlation functions will have lower excited state contaminations. On the other hand, the SP correlation functions will typically have less statistical noise.

5.2 Bare matrix element calculation

The simplest extraction of the reduced pseudo-ITD from the effective matrix element is to analyze the large time limit with a mean value in a region where the effective reduced matrix element has a plateau. This data has plateau regions, especially at low momentum,

where this procedure could be performed. A more sophisticated method can be used to take into account the effects from the lowest excited states which would provide a more justified account of the systematic effects. The effective bare matrix element has excited state contamination of the form

$$M^{eff}(\nu, z^2, T) = M(\nu, z^2)(1 + A_p(z^2)e^{-\Delta_p T} + B_p(z^2)Te^{-\Delta_p T}) \quad (5.2)$$

where $A_p(z^2)$ and $B_p(z^2)$ are the contribution from the matrix elements containing the first excited state and Δ_p is the effective energy gap between the ground state and first excited state. The excited state term coefficients will also be correlation function dependent, i.e. different for each set of smearing parameters.

Further simplifications on this functional form are possible to reduce the number of fit parameters. One could adopt the lattice energy dispersion relation to fix Δ_p to Δ_0 removing another parameter. The rest frame energy could be extracted from two point functions prior to matrix element or left as a fit parameter common to all matrix elements. The latter option, requiring a simultaneous fit of many matrix elements for the sake of fixing a single parameter, may not be too practical. Another choice is to perform the spectroscopy fits on the two point functions to find the energy gaps and hold those values fixed in a fit for the coefficients of the exponentials.

In this work, a fit of the data to Eq. (5.2) is used for each correlation function, holding the ground state matrix element and effective energy gap fixed. Specifically, a fit is performed on N different effective bare matrix elements with different smearing setups, M_j^{eff} , to the form

$$M_j^{eff}(\nu, z^2, T) = M^0(\nu, z^2) \left(1 + e^{-\Delta_p T} [A_p^{(j)}(z^2) + B_p^{(j)}(z^2)T]\right) \quad (5.3)$$

with $2N+2$ fit parameters where $j = 1 \dots N$ labels the different smearings. The fit parameters will be chosen with a weighted χ^2 minimization which employs the full covariance matrix. The minimization of the χ^2 function is performed with the variable projection technique (varpro). Varpro [66] expedites the convergence of a non-linear least squares extremization problem when the non-linear function can be written as a linear combination of non-linear functions.

Consider a general model function

$$f(\alpha, \mathbf{c}, t) = \sum_i c_i F_i(\alpha, t) \quad (5.4)$$

where F_i are non-linear functions which depend on the model parameters α and have linear coefficients c_i . The full minimization problem at hand is given by

$$\min_{\alpha, \mathbf{c}} (y_i - f(\alpha, \mathbf{c}, t_i)) C^{-1} (y_i - f(\alpha, \mathbf{c}, t_i))^T \quad (5.5)$$

where y_i is a dataset corresponding to independent variables t_i with a covariance matrix C . The varpro method suggests splitting this minimization into an exact linear minimization problem and a non-linear minimization problem with fewer variables. The full problem has the same solution as the minimization of the smaller problem

$$\min_{\alpha} (y_i - f(\alpha, \mathbf{c}(\alpha), t_i)) C^{-1} (y_i - f(\alpha, \mathbf{c}(\alpha), t_i))^T \quad (5.6)$$

where only the α parameters are minimized. $\mathbf{c}(\alpha)$ is the solution to the linear minimization problem

$$\min_{\mathbf{c}} (y_i - f(\alpha, \mathbf{c}, t_i)) C^{-1} (y_i - f(\alpha, \mathbf{c}, t_i))^T \quad (5.7)$$

with fixed α .

For fitting matrix elements with Eq. (5.3), the non-linear functions F_i will be a single trivial 1 for the matrix element term and pairs of $e^{-\Delta_p T}$ and $T e^{-\Delta_p T}$ for each correlation function being used. These non-linear functions all share the single non-linear parameter $\alpha = \Delta_p$. The linear coefficients will be the matrix element M and the pairs of $A^{(j)}$ and $B^{(j)}$.

The different momentum smearing parameters only improve the signal-to-noise ratio for a certain range of momenta states and change their fit values for $A^{(j)}$ and $B^{(j)}$. Which correlation functions are used in the fit, as well as the T range of the fit, are varied to minimize the χ^2 per degree of freedom. All statistical errors are estimated using the jackknife resampling technique. An example of these fits for nucleon matrix elements are plotted in Figs. 5.2 and 5.1.

5.3 Renormalization of power divergence

Unlike in dimensional regularization, with a lattice regulator, the operator $O_{\text{WL}}(z)$ has a power divergence in $\frac{z}{a}$. The handling of this power divergence in lattice QCD renormalization schemes, such as the popular RI-MOM scheme, and the associated matching relationships have been generated a large amount of discussion, see [67] for a comparison of methods. In a lattice QCD calculation, renormalization constants require another calculation. An alternative to performing a full RI-MOM calculation is to, when possible, form ratios of matrix elements or renormalization constants which are finite in the continuum limit. In this spirit, a ratio, which has the same leading order in the OPE as the pseudo-ITD, will be constructed where all renormalization constants.

The renormalization constant of the quark fields can be canceled through the local vector current, $M^0(z \cdot p, z^2)|_{z=0}$, where M^0 is the bare matrix element. In the continuum limit, where the vector current is conserved, this matrix element should be equal to 1. Due

to lattice effects, the local vector current is not conserved and has a finite renormalization constant which has ap dependence. This renormalization constant does have the property $Z_V \rightarrow 1$ in the continuum limit $a \rightarrow 0$. Again the leading order behavior of the OPE for the ratio

$$\mathcal{M}^0(z \cdot p, z^2) = \frac{M^0(z \cdot p, z^2)}{M^0(z \cdot p, z^2)|_{z=0}} \quad (5.8)$$

will match the original pseudo-ITD. This ratio still contains the logarithmic and power divergences associated with the Wilson line operator.

The renormalization constant of the Wilson line operator will be canceled by ratios with the rest frame matrix element $\mathcal{M}^0(z \cdot p, z^2)|_{p=0}$. The imaginary component is consistent with zero for all z . The real component results, which are plotted in Fig. 5.4, has the expected exponential behavior expected from the non-perturbative effects generated by the Wilson line operator. The low z/a region exhibits a cusp as $z \rightarrow 0$, which is a signal for the power divergence. For the pseudo-ITD, the matching kernel is unity at $\nu = 0$, meaning the rest frame matrix element will be the integral of the PDF, which for the iso-vector flavor quark combination is 1, up to potential higher twist and discretization errors. The leading order behavior of the OPE for this ratio will be the same at the pseudo-ITD, so that this matrix element satisfies the properties required for the reduced pseudo-ITD.

Finally the reduced matrix element is defined by the double ratio

$$\mathfrak{M}(\nu, z^2) = \left(\frac{M^0(\nu, z^2)}{M^0(\nu, 0)|_{z=0}} \right) / \left(\frac{M^0(0, z^2)|_{p=0}}{M^0(0, 0)|_{p=0, z=0}} \right). \quad (5.9)$$

This double ratio not only takes care of cancelling the multiplicative renormalization constants. It also has the designed goal of cancelling some $O(z^2)$ higher twist contaminations. This feature was demonstrated in the quenched approximation [50], where at fixed Ioffe time and large z^2 , the reduced pseudo-ITD was independent of z^2 instead of showing the

expected polynomial behavior. The reduced pseudo-ITD is a renormalization scheme independent quantity which can be matched directly to the $\overline{\text{MS}}$ light cone PDF through the OPE through Eq. (4.26).

The form of this double ratio has an additional advantage of being exactly unity at $\nu = 0$ with no possible higher twist and lattice spacing errors. This feature explicitly sets the iso-vector quark PDF sum rule, $\int_0^1 dx (u(x) - d(x)) = 1$. Any error in the sum rule for a PDF determined from this reduced pseudo-ITD must be an artifact of the procedure for calculating the PDF.

5.4 Reduced pseudo-ITD results

In Fig. 5.5 and 5.6 the reduced pseudo-ITD, as defined in Eq. (5.9), as a function of p_3 and z_3 respectively. The real component curves all have a Gaussian shape which suggests that the renormalization of the Wilson line was indeed canceled. The curves seem similar in shape, but with decreasing width with increasing momentum. If instead the real component reduced pseudo-ITD was plotted as a function of Ioffe time as in Figs. 5.7 - 5.9, then the data appear on a more universal curve which is nearly independent of z_3^2 . This feature was to be expected since the z^2 dependence of $\mathfrak{M}(\nu, z^2)$ only begins at either $O(\alpha_S \ln(z^2))$ or $O(\Lambda_{\text{QCD}}^2 z^2)$.

The lack of any large polynomial z^2 behavior is a sign that the higher twist contamination is being cancelled in the ratio from Eq. (5.9). This absence of TMD dependence was first observed in [68]. The logarithmic dependence in z^2 is an unavoidable feature of the pseudo-PDF, and reduced pseudo-PDF, in a renormalizable field theory. As discussed in Sec. 4.4.1, the short distance z^2 behavior is intimately related to the μ^2 dependence of the PDF through the OPE. To approach the light cone limit $z^2 \rightarrow 0$, this logarithmic divergence in z^2 requires factorization.

A proper treatment of the z_3 behavior of the reduced Ioffe time pseudo-distribution is necessary for an accurate calculation of the PDF from lattice QCD results. Just as in the testing of renormalization schemes, a window of applicability must be demonstrated where the scale is free of both polynomial corrections, for our case $\Lambda^2 z^2$, and of lattice corrections, for our case $\frac{a}{z}$. Fortunately, the $\log z_3^2$ behavior predicted by perturbative evolution appears to account for all the small z_3 behavior. This feature is encouraging, because it hints that the neglected higher twist effects seem insignificant and that our results are in a regime where perturbation theory is applicable. More tests will need to be performed at several lattice spacings to confirm that the discretization effects are negligible. These sort of confirmations are necessary before any PDF calculation from a lattice QCD result using perturbative formula is believable.

5.5 Evolution and Matching

Even though the data follows a nearly z_3 independent curve, the DGLAP evolution of the PDF dictates a perturbatively calculable dependence on the scale z^2 . Understanding this z^2 dependence is particularly necessary for comparing the data to phenomenological fits which are renormalized with a single scale. The z^2 evolution and $\overline{\text{MS}}$ matching procedures could be performed in a single step by applying the kernel in Eq. (4.24) to each set of data with different z_3 independently. It is also possible to separate the z^2 evolution from the $\overline{\text{MS}}$ matching steps. As long as the steps at the same order in α_s , the one step and two step matching relationships should result in the exact same final $\overline{\text{MS}}$ ITD, though they may have different systematic errors.

Above a certain length scale, the perturbative evolution of the data ceases and a complete factorization of the ν and z^2 arguments appears, up to the neglected higher twist effects which are partly canceled by the ratio. This results in a z^2 independence

for the reduced pseudo-ITD for large z^2 . For the evolution of the data points in this regime, the initial scale could be treated as the scale when evolution first appeared to stop. In the quenched approximation, this scale was found to be $z^{-1} \lesssim 400$ MeV [50]. This scale is particularly low for using a perturbative evolution, and a non-perturbative evolution method would be preferable. Failing to account for this cessation of perturbative evolution would cause the longest distance points to be evolved away from the universal curve.

In order to perform the convolutions in Eq. (4.24), the reduced pseudo-ITD for constant z^2 are fit to a sixth degree polynomial and subsequently integrated over. The real and imaginary components are fit to the even and odd powers in the polynomial respectively, with three free parameters each

$$\begin{aligned}\text{Re } \mathfrak{M}(\nu) &\sim 1 + a\nu^2 + b\nu^4 + c\nu^6 \\ \text{Im } \mathfrak{M}(\nu) &\sim a'\nu + b'\nu^3 + c'\nu^5.\end{aligned}\tag{5.10}$$

In order to test the systematic effects of this choice, the real and imaginary components are also interpolated with a cubic spline. The results of these two integrations are consistent with each other. An example of the convolutions for two of the ensembles are shown in Figs 5.10-5.11. These convolutions follow the same trends as the model predictions shown in Fig. 4.2. Fig. 5.13 show the reduced pseudo-ITD evolved to $z^{-2} = 4e^{2\gamma_E+1}$ GeV². This particular scale was chosen, so that the reduced pseudo-ITD can be easily matched to the $\overline{\text{MS}}$ ITD at $\mu = 4$ GeV. The value of $\alpha_s(4 \text{ GeV}) = 0.229$ was taken from the evolution used by LHAPDF [69] for the dataset `cj15nlo` from the CTEQ-Jefferson Lab collaboration [2].

There are two ways in which these convolutions can be used to evolve and matching the reduced pseudo-ITD to the $\overline{\text{MS}}$ ITD. The most straightforward is a direct inversion of Eq. (4.26) for data with different z^2 independently. At NLO, the inverse of this formula is

found simply by switching $Q(\nu, \mu^2)$ and $\mathfrak{M}(\nu, z^2)$ and changing the sign of α_s resulting in

$$Q(\nu, \mu^2) = \mathfrak{M}(\nu, z^2) - \frac{\alpha_s C_F}{2\pi} \int_0^1 du \left[\ln \left(z^2 \mu^2 \frac{e^{2\gamma_E+1}}{4} \right) B(u) + L(u) \right] \mathfrak{M}(u\nu, z^2). \quad (5.11)$$

An alternative, but equivalent, approach is to perform the z^2 evolution of the reduced pseudo-ITD, for each z^2 independently, using Eq.(4.22) to the scale $z_0^2 = 4e^{-2\gamma_E-1}\mu^{-2}$. With all the data evolved to this common scale, Eq. (5.11) can be applied to match the pseudo-ITD to the $\overline{\text{MS}}$ ITD for the data originating with all z^2 simultaneously. The convolutions with L are performed by fitting the evolved reduced pseudo-ITD to the same polynomials as before in Eq. (5.10). The common scale was chosen such that the scale dependent logarithm in Eq. (5.11) vanishes when matching to the $\overline{\text{MS}}$ ITD for a particular μ .

In this work, the scale $\mu = 4 \text{ GeV}$ was chosen. The evolved reduced pseudo-ITD and the matched $\overline{\text{MS}}$ ITD are shown in Fig 5.13. It has been tested that the evolution and matching procedure performed in a single step or being performed in two steps result in a consistent $\overline{\text{MS}}$ ITD. For the remainder of this work, only the one step matching results will be used.

5.6 Presence of anti-quark distributions

Before attempting to perform a rigorous calculation of the valence quark PDF from the $\overline{\text{MS}}$ ITD, it is instructive to look at the differences between the real component and the imaginary component. As was shown in Eq. (2.23), the real component can be used to determine the valence distribution. Subsequently, the contribution of the valence distribution to the imaginary component can be removed resulting in the antiquark distribution.

The trend of the real component can be estimated from the PDF Ansatz in Eq. (2.25).

In Fig. 5.14, the sine transform of this function can be compared to the imaginary component and the discrepancy is due to the antiquark distribution. For the nucleon, in dynamical quark ensembles, the antiquark ITD,

$$Q_{\bar{q}}(\nu) = i \int_0^1 dx \sin(\nu x) \bar{q}(x), \quad (5.12)$$

tends to have negative values, but the signal is far too poor to determine a quantitative result. A negative trend for the isovector, $q = u - d$, quark combination would be a hint that $\frac{\bar{d}(x)}{\bar{u}(x)} > 1$ which is the violation of the Gottfried sum rule [70]. In the quenched approximation, the case is backwards as was shown in [50]. The relative size of $\bar{u}(x)$ and $\bar{d}(x)$ is dictated by the hadron's quark flavor content. For a nucleon, $\bar{u}(x) = 2\bar{d}(x)$. For this simplified analysis, the trend is identified “by eye”, in a more rigorous study the techniques in Ch. 7 will be necessary. Unfortunately, as will be seen, the current errors do not allow for determination of the anti-quark distribution.

FIG. 5.1: Typical fits of the reduced Ioffe time pseudo-distribution from the ensemble $a_{127}m_{415}L$. The left and right plots correspond to real and imaginary component of the bare matrix element respectively. The ranges of T shown is the range of data used in the fit. The points and bands over the same color correspond to a correlation fitted to Eq (5.3) and the resulting fit respectively. The grey band corresponds to the bare matrix element result.

FIG. 5.2: Typical fits of the reduced Ioffe time pseudo-distribution from the ensemble $a_{127m415}$. The left and right plots correspond to real and imaginary component of the bare matrix element respectively. The ranges of T shown is the range of data used in the fit. The points and bands over the same color correspond to a correlation fitted to Eq (5.3) and the resulting fit respectively. The grey band corresponds to the bare matrix element result.

FIG. 5.3: Typical fits of the reduced Ioffe time pseudo-distribution from the ensemble *a094m390*. The left and right plots correspond to real and imaginary component of the bare matrix element respectively. The ranges of T shown is the range of data used in the fit. The points and bands over the same color correspond to a correlation fitted to Eq (5.3) and the resulting fit respectively. The grey band corresponds to the bare matrix element result.

FIG. 5.4: The rest frame distribution $\mathcal{M}^0(z \cdot p, z^2)|_{p=0}$. The top left plot is from the ensemble *a127m415*, the bottom left plot is from the ensemble *a127m415L*, and the top right plot is from the ensemble *a094m390*. The cusp as $z \rightarrow 0$ is the signal for the power divergences which occur in perturbation theory. The large z limit reveals the exponential nature of the Wilson line renormalization constant.

FIG. 5.5: The reduced Ioffe time pseudo-distribution as a function of p . Here the momenta and displacement are labeled in lattice units, $p_3 = 2\pi p/L$ and $z_3 = za$. The left plots are the real component and the right are the imaginary component. The top plots are from the ensemble $a127m415$, the middle plots are from the ensemble $a127m415L$, and the bottom plots are from the ensemble $a094m390$.

FIG. 5.6: The reduced Ioffe time pseudo-distribution as a function of z . Here the momenta and displacement are labeled in lattice units, $p_3 = 2\pi p/L$ and $z_3 = za$. The left plots are the real component and the right are the imaginary component. The top plots are from the ensemble *a127m415*, the middle plots are from the ensemble *a127m415L*, and the bottom plots are from the ensemble *a094m390*.

FIG. 5.7: The real and imaginary components of the reduced Ioffe time pseudo-distribution on the ensemble *a127m415* as a function of $\nu = z \cdot p$.

FIG. 5.8: The real and imaginary components of the reduced Ioffe time pseudo-distribution on the ensemble *a127m415L* as a function of $\nu = z \cdot p$.

FIG. 5.9: The real and imaginary components of the reduced Ioffe time pseudo-distribution on the ensemble *a094m390* as a function of $\nu = z \cdot p$.

FIG. 5.10: The convolutions required for the evolution and matching of the reduced pseudo-ITD to the $\overline{\text{MS}}$ ITD on the ensemble $a127m415$. The reduced pseudo-ITD was interpolated by fitting a polynomial.

FIG. 5.11: The convolutions required for the evolution and matching of the reduced pseudo-ITD to the $\overline{\text{MS}}$ ITD on the ensemble $a127m415L$. The reduced pseudo-ITD was interpolated by fitting a polynomial.

FIG. 5.12: The convolutions required for the evolution and matching of the reduced pseudo-ITD to the $\overline{\text{MS}}$ ITD on the ensemble $a094m390$. The reduced pseudo-ITD was interpolated by fitting a polynomial.

FIG. 5.13: The evolution and matching of the pseudo-ITD to the ITD. The left and right plots show the real and imaginary components respectively. The top plots are from ensemble $a127m415$ and the lower plots are from ensemble $a127m415L$.

FIG. 5.14: Estimating the size of the antiquark distribution. The left and right columns are from ensemble $a127m415$ and $a127m415L$ respectively. The top plots are the real component, middle plots are the imaginary component, and the bottoms plots are the sine transform of the antiquark distribution. The trend for ensemble $a127m415$ uses $a = -0.2$ and $b = 2.5$. The trend for ensemble $a127m415L$ uses $a = -0.1$ and $b = 2.75$.

CHAPTER 6

Calculation of Ioffe time pseudo-SF

The next LCS to be calculated is the axial vector-vector current combination, $\tilde{\mathcal{M}}^{\mu\nu}$ mentioned in Sec. 4.5.1. The following description of the calculation in this chapter is taken from [51]. The numerical calculation is performed on the ensemble $a127m415L$. This particular current combination will be used to calculate the pion quark valence PDF for comparison with the results from the pion pseudo-ITD. As in the calculation of the pseudo-ITD, the momentum and separation will be chosen to be along the z direction. Based upon this choice, the particular Lorentz indices in the currents will be chosen to be $\mu = x$ and $\nu = y$ in order to isolate contributions from individual Ioffe time pseudo-SF, $\mathcal{T}_1(\nu, z^2)$ from Eq. (4.46).

6.1 Method of Calculation

The matrix elements for the two current operators will be calculated with the traditional ratio method described in Sec. 3.3.2. The correlation function containing the current operators, in this case a four point correlation function, will be calculated using a modified version of the sequential source technique. A sequential source is an object which contains

the source and sink interpolating fields and allows for contractions with the operators of interest. In contrast to the sequential operator technique described in Sec. 5.1, the sequential source method allows all operators of interest to be inserted without additional propagator insertions, but requires the calculation of a new sequential source for every new interpolating field, including when changing the separations between the interpolating fields. In addition, to minimize excited state contamination, the operator insertion time, t , will be fixed to be midway between the source and sink interpolators, *i.e.* $t = \frac{T}{2}$. Varying t would require recalculation of the modified sequential source which would become prohibitively expensive. The necessity to restrict t limits the ability of the modified sequential source technique to precisely control the excited state effects compared to a typical sequential source calculation. Unfortunately this cost is necessary for the calculation of four point functions.

For the calculation of any good LCS, the composite operators used have finite spatial extent z . Introduction of a heavy auxiliary quark field Q ($m_Q > m_l$), such that our operators are of the form $\mathcal{O}(t) = \mathcal{J}_\Gamma^\dagger(z, t) \mathcal{J}_\Gamma(0, t)$ with $\mathcal{J}_\Gamma = \bar{q}\Gamma Q + \bar{Q}\Gamma q$, limits the available momentum phase space of the quark propagating between the two currents thereby reducing the statistical noise. An auxiliary heavy quark has also been used in Ref. [71] to remove higher twist contamination in the calculation of moments of the PDF and the distribution amplitude (DA). For this calculation of the pion valence distribution, multiple auxiliary quark masses between the light and strange quark mass were tested. Only a slight improvement in the signal-to-noise ratio from the heavier masses was observed for the larger momenta. Using quark masses which are too large run the risk of $O(m_q a)$ discretization errors, but no mass dependence was observed. The auxiliary quark propagator is set to the strange quark mass for the remainder of this calculation.

As mentioned in Sec. 4.5, flavor changing currents have contributions for the PDFs of both quark flavors. When using an artificial quark flavor, this additional PDF is an

unwanted systematic error. Fortunately, in this calculation of a valence PDF, there is no valence PDF for the artificial auxiliary quark, meaning this systematic error is entirely removed. On the other hand a calculation of the other current combinations mentioned in Sec. 4.5, which match to the CP odd combination $q_v + 2\bar{q}$, will contain this systematic error. Additionally, when using a flavor changing current with artificial quarks, the set of Wick contractions for the higher twist “cat ears” diagram do not contribute. These contractions do not exist because the interpolating fields do not contain the artificial quarks, and that source of higher twist contamination is removed from the matrix element calculation.

The 4-point correlation function is constructed using a modified sequential source technique. The standard sequential source method does not work here, because this correlation function requires a quark propagator of fixed length to and from many locations on a given time slice. Such an all-to-all propagator is impractical to calculate. However, for the case of meson there is a straightforward implementation where momentum projections are performed at source and sink meson operators and the corresponding correlation functions are computed as chain of sequential sources as described below. The correlation function is expressed as follows

$$\begin{aligned}
C_{4\text{pt}}(z, p, T, t) &= \langle \Pi_p(\vec{z}, T) \mathcal{J}_\Gamma^\dagger(x_0 + z, t) \mathcal{J}_{\Gamma'}(x_0, t) \bar{\Pi}_p(\vec{y}, 0) \rangle \\
&= \sum_{\vec{z}, \vec{y}} e^{-i(\vec{z}-\vec{y}) \cdot \vec{p}} \langle \left[\tilde{d} \gamma^5 \tilde{u} \right] (\vec{z}, T) \left[\bar{Q} \Gamma u + \bar{u} \Gamma Q \right] (x_0 + z, t) \times \\
&\quad \left[\bar{u} \Gamma' Q + \bar{Q} \Gamma' u \right] (x_0, t) \left[\tilde{u} \gamma^5 \tilde{d} \right] (\vec{y}, 0) \rangle \\
&= \text{Tr} \left[I_u^p(x_0 + z, t; x_0, t) \Gamma \gamma^5 G_Q(x_0 + z, t; x_0, t)^\dagger \gamma^5 \Gamma' \right] + (\Gamma \rightarrow \Gamma' \text{ and } z \rightarrow -z) \quad (6.1)
\end{aligned}$$

where x_0 is a randomly determined source point, $\tilde{q}$ are momentum smeared quark fields, $G_Q(y'; x')$ is the flavor Q auxiliary quark propagator from x' to y' , and $I_q^p(y'; x')$ is the modified sequential source with flavor q -quarks and pions at momentum p . The modified

sequential source is constructed through sequential inversions of the light quark Dirac operator, reusing already calculated propagators as sources.

Heuristically, the modified sequential source is constructed by calculating the light quark propagator from a point-source located at one of the currents to the source interpolator, using this object as a source for a subsequent propagator to the sink interpolator, and lastly using this larger object as a source for propagation from the sink to the second current. This construction is done by solving the following sequence of systems of equations for G_q , H_q^p , and I_q^p , suppressing color and flavor indices,

$$\begin{aligned}
\sum_{x',s'} D_q(x,s;x',s') G_q(x',s';x_0,t) &= \delta(x-x_0)\delta(s-t) \\
\sum_{x',s'} D_q(x,s;x',s') H_q^p(x',s';x_0,t) &= \\
e^{-ix \cdot p} \sum_{x',x''} S(x;x') \gamma^5 S(x';x'') G_q(x'',s;x_0,t) \delta(s) &= \\
\sum_{x',s'} D_q(x,s;x',s') I_q^p(x',s';x_0,t) &= \\
e^{ix \cdot p} \sum_{x',x''} S(x;x') \gamma^5 S(x';x'') H_q^p(x'',s;x_0,t) \delta(s-T) &=
\end{aligned} \tag{6.2}$$

where D_q is the Dirac matrix for the light quarks and $S(x;x')$ represents the smearing procedure. The phase projects the interpolating fields onto definite momentum, while the signs of the momentum-smearing phases applied to the quark fields must be treated carefully to ensure Π and $\bar{\Pi}$ correctly project onto states with a given momentum. Note that the smearing procedures are applied once for each of the quark fields in the interpolating fields, and no smearing is applied at the current insertions. This procedure is shown diagrammatically in Fig 6.1. An advantage of this procedure is that the correlation function for any current pair with any separation can be calculated without additional Dirac matrix inversions. On the other hand, for each choice of momentum and source-sink separations,

the first two light quark Dirac matrix inversions must be recalculated.

FIG. 6.1: The four point-function is constructed by combining a heavy quark propagator, represented by the green line, and a modified sequential source, represented by the black lines. An advantage of this setup is that any pair of currents with any separation may be constructed without additional costly propagator inversions.

6.2 Extraction of Matrix elements

In this section, the numerical results of the calculation of the matrix elements of spatially-separated antisymmetric V-A currents is presented. As discussed in Section 6.1, the pion source-sink separation T is systematically increased, while holding the time $t = \frac{T}{2}$ fixed at which the currents are inserted. Figs. 6.2 and 6.3 show the calculation of the operator $\langle \pi(p) | \mathcal{J}_i(x_0 + z) \mathcal{J}_j(x_0) | \pi(p) \rangle$, where $\mathcal{J}_i$ and $\mathcal{J}_j$ are the antisymmetric V-A currents discussed in Section 4.5.1.

A correlated fit is performed to the jackknife ensemble ratios of four point to two point correlation functions for a given momentum p and spatial separation z between the currents. In order to extract the desired matrix element from fits to our data, we assume the following single exponential form for the ratios of four point to two point functions:

$$R(T) = \frac{C_4(T)}{C_2(T)} = A + B e^{-\Delta T/2} \quad (6.3)$$

where Δ is the effective energy gap between the ground-state and the excited states.

Therefore the ratio $R(T)$ will give the desired matrix element $\langle \pi(p) | \mathcal{J}_i(x_0 + z) \mathcal{J}_j(x_0) | \pi(p) \rangle$, up to an additional amplitude obtained from the fit to the two point function in the asymptotic limit of large T . Given the symmetries engineered into our calculation, namely the current always being inserted at $t = \frac{T}{2}$ and the source/sink interpolators being created in an identical manner, one can further assume that the excited state contamination is equal on the source and sink sides of the current.

In Figs. 6.2a and 6.2b we present representative fits according to Eq. (6.3) applied to the jackknife ensemble ratio of four point to two point correlation functions as a function of source-sink separation T , for $z = (0, 0, a, 0)$ and $z = (0, 0, 4a, 0)$ with momenta along the z -direction $p = 0.610$ GeV and $p = 1.525$ GeV, respectively. For $p = 0.610$ GeV, the largest source-sink separation we use is $T = 22a$. As the signal-to-noise ratio is significantly reduced for states with large momentum, we attempt to limit additional noise in our extracted matrix elements by considering smaller separations between the currents for states with large momenta. For instance, we limit the largest source-sink separation to be $T = 16a$ for $p = 1.525$ GeV. As seen from Fig. 6.2a, the ratio of the correlation function for $p = 0.610$ GeV has reasonable signal for almost all the source-sink separations. However, as expected for $p = 1.525$ GeV, after $T = 12a$, the lattice matrix elements become very noisy and have no effect on the fit. Reasonable statistical signal in such a relatively small window of source-sink separations, compared to lower momenta data, might be a cause for concern. However it is worth remembering that with such a coarse lattice spacing ($a = 0.127$ fm), $T = 12a$ is sufficiently large (~ 1.525 fm) to minimize any excited-state contamination. We present values of the fit parameters in Table 6.1.

With the fit to the data for momenta along the z -direction in the range $p \in [0.610 - 1.525]$ GeV and current separations $|z| \leq 4a$ in the z -direction, we obtain the matrix elements shown in Fig. 6.4. As discussed in Ch. 4, we only include $|z| \leq 4a$ in our analysis so that

p [GeV]	ξ	A	B	Δ_{eff}	$\chi^2/\text{d.o.f.}$
0.610 GeV	$1a$	0.102(5)	-0.028(11)	0.054(20)	1.21
0.610 GeV	$4a$	0.083(4)	-0.026(13)	0.062(42)	0.89
1.525 GeV	$1a$	0.097(8)	-0.267(513)	0.809(503)	0.15
1.525 GeV	$4a$	0.049(9)	-0.107(148)	0.450(420)	0.19

TABLE 6.1: The fit parameters of the V-A matrix elements with $z = (0, 0, a)$ for momenta along the z -direction $p = 0.610$ GeV and $p = 1.525$ GeV, respectively.

z is sufficiently smaller than $\frac{1}{\Lambda_{\text{QCD}}}$ and thereby ensuring the factorization of Eq. (4.7).

FIG. 6.2: Jackknife ensemble ratio data of the 4pt to 2pt correlation functions used in the extraction of antisymmetric V-A current matrix elements. Figs 6.2a and 6.2b show fits to the matrix elements for $p = 0.610$ GeV with spatial separation between the currents $z = 1a$ and $z = 4a$, respectively. The blue data points are obtained from the lattice QCD calculation, the green band shows the two-state fit to the data, and the red band shows the extracted value of the matrix element in the asymptotically large source-sink separation limit.

FIG. 6.3: Jackknife ensemble ratio data of the 4pt to 2pt correlation functions used in the extraction of antisymmetric V-A current matrix elements. Figs 6.3a and 6.3b show fits to the matrix elements for $p = 1.525$ GeV with spatial separation between the currents $z = 1a$ and $z = 4a$, respectively. The green data points are obtained from the lattice QCD calculation, the gray band shows the two-state fit to the data, and the red band shows the extracted value of the matrix element in the asymptotically large source-sink separation limit.

FIG. 6.4: The pseudo-ITSF of the axial vector-vector current combination.

CHAPTER 7

Extraction of PDF

The determination of the PDF from lattice results constitutes an ill-posed inverse problem to solve. The methods which resolve ill-posed inverse problems must be verified that they work in the specific case of inverse Fourier transforms. Much of the following discussion and tests in this chapter were taken from [72].

All current methods for extracting parton distribution functions (PDF) or distribution amplitudes (DA) from lattice QCD require calculating a hadronic matrix element which is an integral of the PDF. These can be a Fourier transform, Mellin transform, Laplace transform or something more complicated. The inversion of these integral relationships constitute an ill-posed inverse problem. The following will be largely concerned with the reconstruction of the PDF from the $\overline{\text{MS}}$ ITD, but many of the conclusions are also applicable to the reconstruction of quasi-PDFs from the lattice calculated matrix element which is also a Fourier transform. These techniques can also be used in the analysis of the Hadronic tensor and the reconstruction of the PDF through the moments, but their usefulness in these cases must be studied. In inverting Eq. (2.20) to obtain $q(x, \mu^2)$, there exist two challenges. First being that the required integral over ν does not extend over the

full Brillouin zone. The second is that in practice only a small number of points along ν can be computed. Only the second challenge exists for the calculation of the quasi-PDF. As we will discuss in more detail below, taken together these issues render the extraction highly ill-posed and we explore several regularization strategies on how to nevertheless reliably estimate the PDF from the data at hand.

It should be noted that though comparisons between this calculation and phenomenologically derived PDFs will be made, very few of the systematic errors in this calculation are taken into account. The most important appears to be the pion mass. Calculations of the local matrix element for the first moment of the PDF are systematically too large at heavy pion mass. This same feature occurs in this calculation, include the determination of the moments in Sec. 7.2. The first moment is determined by large x behavior of the PDF, so this region will be overestimated by this calculation. In order to satisfy the sum rule, the PDF must also be underestimated in some other region of the PDF.

7.1 Issue of Fourier Transformation

Though the extraction of the PDF from the ITD may seem straight forward at first glance, there exist significant technical issues. The limited range of separations and momenta available to a modern lattice calculation restricts the range in Ioffe time to a small window. Since this finite range is different from the full Brillouin zone, the inverse Fourier transform is ill-posed. To expose the issue, consider a naïve discretization of the Fourier transform to $N_x + 1$ points over a range $x \in [0, 1]$ implementing the trapezoid rule. The Fourier transform in Eq. (2.23) becomes

$$\mathcal{Q}(\nu) = \frac{1}{2} \cos(\nu x_0) q_v(x_0) + \left[\sum_{k=1}^{N_x-1} \Delta x \cos(\nu x_k) q_v(x_k) \right] + \frac{1}{2} \cos(\nu x_{N_x}) q_v(x_{N_x}) \quad (7.1)$$

with $\Delta x = 1/Nx$ and $x_k = k\Delta x$.

Assuming there exists an equal number of data points for $Q(\nu)$ over a range $\nu \in [0, \nu_{\max}]$, Eq. (7.1) can be written as a linear system of equations which can be solved exactly for the unknown $q_v(x_k)$. Let $\mathfrak{Q}$ be a vector of the data at evenly separated points $\nu_k = k\Delta\nu$ where $\Delta\nu = \nu_{\max}/N_x$ and let $\mathfrak{q}$ be a vector of the unknown valence PDF at the values x_k . The system of equations relating these two can be written in matrix form as

$$\mathfrak{Q} = \mathfrak{C}\mathfrak{q} \quad (7.2)$$

with the Coefficient matrix

$$\begin{aligned} \mathfrak{C}_{kl} &= \Delta x \cos(\nu_k x_l) = \frac{1}{N_x} \cos(\nu_k x_l) \quad \text{for } l \in [1, N_x - 1], \\ \mathfrak{C}_{kl} &= \frac{1}{2} \Delta x \cos(\nu_k x_l) = \frac{1}{2} \frac{1}{N_x} \cos(\nu_k x_l) \quad \text{for } l = 0, N_x. \end{aligned} \quad (7.3)$$

The ill-posed nature of the problem can be observed by studying the coefficient matrix. In an ill-posed situation, one will not be able to invert the Coefficient matrix. By examining the eigenvalues, one can see the dependence on the range of Ioffe time.

Consider the case of $N_x = 40$ and maximum Ioffe times $\nu_{\max} = 10, 20, 100$, and 40π . When the range of Ioffe time is below the full Brillouin zone, the set of cosine functions do not constitute a complete linearly independent basis of functions. When that happens, the columns of the Coefficient matrix become linearly dependent and eigenvalues become 0 or exponentially suppressed. The eigenvalues of the Coefficient matrices for the different $\nu_{\max}$ are shown in Fig 7.1. The case $\nu_{\max} = 40\pi$ is special, because it may be recast as a genuine discrete Fourier transform and the inverse should not be ill-posed. The eigenvalues for this case are all of the same order and do not show the exponential suppression that the cases with smaller $\nu_{\max}$ exhibit.

FIG. 7.1: The eigenvalues of the coefficient matrix from direct inversion (right) and derivative method inversion (left). In both cases, the ill-posed nature of the problem suppresses the eigenvalues exponentially except for the final range which is a genuine discretized Fourier transform.

The singular nature of the Coefficient matrix shows that the inversion problem is ill-posed and that the direct inversion of the Fourier transform is impractical for Ioffe time range much below the Brillouin zone. To demonstrate the effects of the ill-posedness, the inverse will be taken with an SVD cutoff for mock data. The “data” will come from the Fourier transform of the PDF model

$$q(x) = \frac{x^a(1-x)^b}{B(a+1, b+1)}, \quad (7.4)$$

with $a = -\frac{1}{4}$ and $b = 3$. Gaussian noise is added to the “data” in order to see how the possible fluctuations will effect the quality of the reconstruction. An important question is what quality of signal-to-noise is necessary for a precision reconstruction of the PDF.

It has also been suggested in [73], that the related extraction of the quasi-PDF can be improved by considering the derivative of the matrix element. This trick, an application of integration by parts which is called the “derivative method”, was demonstrated to create a more faithful reconstructed PDF from a model input ITD made from the CJ15 data. It is important to note that the test in [73] showed an improvement for an infinitely precise derivative of the ITD with range of Ioffe time from 0 to 12π , which is much larger than

what can be achieved in a modern lattice calculation. Also, it should be noted that in the original proposal of the derivative method, the boundary term from the integration by parts, $\left((-2\pi i p_z x)^{-1} e^{-i x p_z z} h(z)\right)\Big|_{-z_{\max}}^{z_{\max}}$ is neglected.

The derivative method can be written as attempting to invert the equation

$$\partial \mathcal{Q}(\nu) = - \int_0^1 dx x \sin(\nu x) q(x), \quad (7.5)$$

which can be discretized as before to

$$\partial \mathfrak{Q} = (\partial \mathfrak{C}) \cdot \mathfrak{q}, \quad (7.6)$$

with $\partial \mathfrak{C}$ being the coefficient matrix with matrix elements,

$$\begin{aligned} [\partial \mathfrak{C}]_{kl} &= -\Delta x x_l \sin(\nu_k x_l) = -\frac{1}{N_x} x_l \sin(\nu_k x_l) \quad \text{for } l \in [1, N_x - 1], \\ [\partial \mathfrak{C}]_{kl} &= -\frac{1}{2} \Delta x x_l \sin(\nu_k x_l) = -\frac{1}{2} \frac{1}{N_x} x_l \sin(\nu_k x_l) \quad \text{for } l = 0, N_x. \end{aligned} \quad (7.7)$$

As shown in Fig. 7.1, the eigenvalue spectrum of Coefficient matrix of the derivative method does not differ qualitatively from that of the direct inversion. The derivative method does not alleviate the ill-posed nature of the inverse problem, but instead pushes the artificial oscillations into the derivative of the PDF. As before, the test of the inversion with a SVD cutoff is performed on the mock data.

Some of the following methods perform more reliably in a preconditioned form to extract a rescaled PDF $h(x) = q(x)/p(x)$ where $p(x)$ is an appropriately chosen function to remove difficult to recover features. This preconditioning requires a rescaled kernel $K'(x) = K(x)p(x)$, where the kernel's other dependences are suppressed. Methods for extracting smooth and slowly varying functions, such as the Backus Gilbert method, are

significantly improved by using a preconditioning to remove the diverge at $x = 0$.

7.2 Calculation of pseudo-moments

Before figuring out how to invert the Fourier transform, information about the PDF can still be determined from the reduced pseudo-ITD. The pseudo-moments can be used to calculate PDF moments while avoiding entirely the inverse problem [60]. A discretized version of the relationship in Eq. (4.15) between the pseudo-moments and the reduced pseudo-ITD data can be written in the matrix form

$$\mathfrak{M} = \mathbf{C}\mathbf{b}, \quad (7.8)$$

where $\mathfrak{M}$ is a vector of N data points, $\mathbf{C}$ is known as the Vandermonde matrix, and $\mathbf{b}$ is a vector of M moments weighted by the factor of $i^n/n!$ mentioned in Eq. (4.15). The Vandermonde matrix is an $N \times M$ matrix of the form $\mathbf{C}_{in} = \nu_i^n$ where ν_i is the Ioffe time for the i -th data point in $\mathfrak{M}$. This relationship can also be split into a real and imaginary components which only contain even and odd powers of ν and result in even and odd pseudo-moments respectively. This equation is inverted for points with a fixed z^2 . The results of the first and second pseudo-moments as well as the matched PDF moments are shown in Fig 7.2. As described in Sec. 4.3, the pseudo-moments can then be matched to the $\overline{\text{MS}}$ moments.

At small separations, the pseudo-moments have small dependence on z^2 . After the application of the DGLAP evolution and matching relationships from Eq. (4.27), any residual z^2 dependence of the moments of the PDF, which would be caused by higher twist contaminations, appears negligible. Fig 7.2 has a comparison of this calculation of the $\overline{\text{MS}}$ PDF and the moments calculated from the CJ15 fits. Just as in previous

calculations of the local matrix element [74], the PDF moment at heavy pion mass are systematically higher than the phenomenologically expected result.

In principle, one can also use this technique to extract the higher moments. This procedure has been tested on the next two higher moments. Only the results with the largest few z^2 had statistical errors comparable to the lower moments. Since these data potentially have significant the neglected higher twist corrections, these results should be considered questionable. The range of Ioffe time for those large z data points had been sufficiently large to determine the second variable in the Taylor expansion in Eq. (5.10) with reasonable statistical precision. The small z data points appear almost entirely described by including only the terms proportional to a and a' . Further studies on finer lattice spacings will be required to extend the range of Ioffe time for low z data in order to constrain the higher moments and confirm the lack of higher twist contamination which was observed in the lower moments.

Reconstructing the PDF from its moments in itself is an inverse problem. Instead of inverting a Fourier transform, like most of the procedures discussed in this chapter, this problem is the inversion of a Mellin transform. The PDF can be parameterized by some function such as the PDF Ansatz in Eq. (2.25) or the moments can be parameterized by some function with a known inverse Mellin transform as was suggested in [49]. In this work, with at best two moments being constrained, there is not much hope for actually performing a reliable fit to any of these functional forms.

The determination of pseudo-moments is a useful calculation while studying these Ioffe time distributions. The inversion of the Fourier transform can be a dubious problem and the pseudo-moments allow for a quick sanity check that the data contains reasonable information. The residual z^2 dependence of the $\overline{\text{MS}}$ moments, derived from pseudo-moments, is a non-trivial check of the size of higher twist effects. The lack of any statistically meaningful z^2 dependence in the $\overline{\text{MS}}$ moments calculated on these ensembles

can be used to justify the validity of using the data with all z^2 in the following PDF extractions.

7.3 Fits of Model Ansatz

All solutions to the inverse problem require adding additional information to constrain a unique solution for the unknown function. In the global PDF fitting community, the most common choice for solving the inverse problem is to choose a physically motivated functional form for the PDF. By choosing a PDF parameterization with fewer parameters than existing data points, the inverse problem is regulated. This model Ansatz can be designed to explicitly show some limiting behaviors, physically motivated features, and satisfy some possible constraints. The better motivated the information used to create the Ansatz the more successful this technique will be, but any particular choice made will introduce a model dependent bias into the final result. The ill-posed inverse problem does not have a unique solution, and ultimately some bias must be introduced into the PDF determination. Ideally, several different parameterizations will be checked and compared, and in effect this cross checking has occurred amongst the phenomenological PDF fits each with different choices of models.

Generally, the phenomenological PDFs are fit to a functional form

$$f(x) = x^a(1-x)^b P(x), \quad (7.9)$$

where $P(x)$ is a yet to be specified function with more model parameters. There exist well known features of PDFs, such as vanishing as $x \rightarrow 1$ or diverging as $x \rightarrow 0$. These limiting behaviors can be seen through the signs of the model parameters a and b in Eq. (7.9). By separating these features, $P(x)$ is allowed to be a smoother and slower varying function

which is easier to extract.

One choice of $P(x)$ for the valence quark PDF employed by the CJ collaboration [2] and MSTW [5] is

$$P(x) = \frac{1 + c\sqrt{x} + dx}{B(a+1, b+1) + cB(a+1.5, b+1) + dB(a+2, b+1)}, \quad (7.10)$$

where $B(x, y)$ is the Beta function. This functional form explicitly sets the PDF's sum rule and allows all moments to be directly calculated as a ratio of sums of Beta functions. The choice used by MMHT [75] is

$$P(x) = A \left(1 + \sum_{i=1}^4 c_i T_i^{\text{Ch}}(1 - 2\sqrt{x}) \right), \quad (7.11)$$

where $T_i^{\text{Ch}}(y)$ are Chebyshev polynomials. Another choice used by the CTEQ [76] collaboration, given by

$$P(x) = A \exp [c_0 + c_1\sqrt{x} + c_2x + c_3x^2] \quad (7.12)$$

Due to the possibly large fluctuations allowed by the exponential and strong correlations between c_1 , c_2 , and c_3 , in a more recent work [77], this choice was replaced by

$$P(x) = A \sum_{i=0}^4 c_i b_i^4(\sqrt{x}) \quad (7.13)$$

which is a sum of Bernstein polynomials $b_n^\nu(y) = \binom{\nu}{n} y^\nu (1-y)^{n-\nu}$. In all of these parameterizations, the normalization is fixed in order to satisfy the sum rule. Finally, the NNPDF collaboration [6] uses a neural network to parameterize $P(x)$. This approach will be discussed in more detail in Sec. 7.4.2.

Though some of these choices, particularly the polynomials in $\sqrt{x}$, are similar, they have different systematic effects and biases, such as the correlations between fit parameters.

From the simplest polynomial in (7.10) to the extremely flexible neural network, all these functions produce statistically consistent PDFs within the range of x where experimental data exists and with comparable $\chi^2/\text{d.o.f.}$.

Due to the limited number of data points available to this lattice calculation, the number of fit parameters can be rather limited. Lack of control of systematic effects, such as the higher twist z^2 effects and discretization errors, will limit the χ^2 which can be achieved while fitting the lattice data.

7.3.1 Nucleon Reduced pseudo-ITD

For the calculation of the PDF from the ITD, different functional forms will be tried to parameterize $P(x)$ in Eq. (7.9). The statistical errors will be obtained with the jackknife resampling technique. In the dynamical quark ITDs, there does not appear to be a strong dependence on the initial separation with which the data point had originated. This feature indicates that large polynomial z^2 corrections do not exist. Therefor we can fit all z^2 separations simultaneously. The fits are performed using all separations or with separations $|z| \lesssim 0.5$ fm. Typically, the two results are consistent with each other, but with smaller statistical errors when all the data are used. All of the resulting fits are compared to phenomenologically determined PDFs from the NLO global fit CJ15nlo [2], and the NNLO global fits MSTW2008nnlo68cl_nf4 [5] and NNPDF31_nnlo_pch_as_0118_mc_164 [6] all evolved to 4 GeV.

The first functional form tested is the simplest PDF Ansatz in Eq. (2.25). On the other hand with only two parameters, one cannot expect $\chi^2/\text{d.o.f.}$ to be too small. Fig. 7.3-7.5 shows the results of these fits. This simple PDF parameterization does very well to reproduce the data even with 2 parameters. The qualitative features of the PDF, such as the divergence as $x \rightarrow 0$ and vanishing as $x \rightarrow 1$ do appear.

The next functional form tested is that used by the JAM collaboration which modifies the Ansatz by the factor Eq. (7.10). This functional form provides slightly more flexibility Eq. (2.25). Fig. 7.6- 7.8 shows the results of these fits. The fits with the full dataset perform very similarly to the Ansatz fit. On the other hand, the fact that the largest available Ioffe time for the fit of the $z < 4a$ data allows the PDF to fluctuate more. If the extrapolation of the fit result beyond the region of Ioffe time has large variance, the fit PDF will be poorly constrained.

The next test is to add more terms to the polynomial used by the JAM collaboration. The polynomial

$$P(x) = N^{-1} \left(1 + c x^{\frac{1}{2}} + d x + e x^{\frac{3}{2}} + f x^2 \right) \quad (7.14)$$

where

$$N = B(a+1, b+1) + c B(a+1.5, b+1) + d B(a+2, b+1) + e B(a+2.5, b+1) + f B(a+3, b+1) \quad (7.15)$$

will be used to fit the PDF. Fig. 7.9- 7.11 shows the results of these fits. The resulting PDFs from these fits all seem to carry some oscillatory features. As claimed, the fit for ensemble *a127m415* using only data from $z < 4a$, which has a large variance in Ioffe times, appear to have solutions with very large oscillations in the PDF. Constraining large Ioffe time results is very important for reliable extractions of the PDF with these model fits.

Next interpolating polynomial tried is a linear combination of Chebyshev polynomials. The fits are performed using the form

$$P(x) = \sum_{i=0}^3 c_i T_i^{\text{Ch}}(x). \quad (7.16)$$

Unlike the previously used functional forms, this one is not normalized by the sum rule. This constraint, though helpful, is not always necessary in these fits. The data at low Ioffe

time are close to 1 with small statistical errors. These precise data will help to push the sum rule to be satisfied in some cases. The results of these fits are shown in Figs. 7.12- 7.14.

The larger variance of the fit result at large Ioffe time, outside where the data exists, result in larger errors in the PDF. It will be seen in other functional forms when the variance grows significantly at large Ioffe times, the PDF will contain oscillatory solutions which generate large errors. In these functional forms, the large Ioffe time behavior of the PDF is largely governed by the low x parameter a . The large Ioffe time behavior of the model in Eq. (2.26) is given by

$$Q(\nu) \sim -\sin\left(\frac{\pi}{2}a\right) \frac{\Gamma(a+1)}{\nu^{a+1}} + b \cos\left(\frac{\pi}{2}a\right) \frac{\Gamma(a+2)}{\nu^{a+2}}. \quad (7.17)$$

The PDF must have a finite integral for the sum rules to be enforced. This feature restricts the power of the $x \rightarrow 0$ divergence to $a > -1$ and this ITD must vanish in the limit $\nu \rightarrow \infty$. All of the polynomials tried above, will only add terms which force the ITD to converge to 0 more rapidly. Any of these fit solutions which does not eventually converge to 0 should be rejected, because it cannot have a finite integral and violates the sum rule.

7.3.2 Axial vector-vector current pseudo-SF

The following analysis of the pseudo-SF was performed in [51].

With the limitations related to z^2 corrections, both from higher twist and discretization, in mind, in this preliminary calculation, we use the numerical fitting program ROOT [78] to fit bootstrap samples of the axial vector-vector matrix elements and obtain a LO $q_V^\pi(x)$ -distribution. The uncertainty band in the fit has been obtained from the fit results of the bootstrap samples. The functional form was the one used by the JAM collaboration given in Eq. (7.10) The integrations with the LO kernel in Eq. (4.48) are performed numerically. With the various sources of z^2 corrections not taken into account, we did not expect the

matrix elements as a function of Ioffe time to fall upon a single curve - consequently the $\chi^2/\text{d.o.f.}$ was close to 2.2.

We find that the term $c\sqrt{x}$ in Eq. (7.10) has no effect in the fit, as $d \simeq 0$. A similar zero-value for d was also found in Ref. [12] and other global fits to experimental data. We therefore adopt a simpler functional form for the PDF in our calculation

$$q_v^\pi(x) = \frac{x^\alpha(1-x)^\beta(1+\gamma x)}{B(\alpha+1, \beta+1) + \gamma B(\alpha+2, \beta+1)}, \quad (7.18)$$

where the beta functions in the denominator ensure the normalization condition of the sum rule is met. In Eq. (7.18), the $(1-x)^\beta$ allows a smooth interpolation to zero as $x \rightarrow 1$ and is inspired by the counting rule of perturbative QCD. The x^α term is motivated by the behavior predicted by Regge theory at small x . One could interpolate these two limits using a polynomial of x . However, due to present statistics and small range of z , we cannot quantitatively distinguish between different choices of polynomials. The following physically motivated and relaxed constraints are set

$$\alpha < 0, \quad 0 < \beta < 4. \quad (7.19)$$

The fit to the lattice QCD data using the LO kernel in Eq. (4.48) and the functional form of PDF in Eq. (7.18) is shown in Fig. 7.15 with the fit parameters,

$$\begin{aligned} \alpha &= -0.34(31) \\ \beta &= 1.93(68) \\ \gamma &= 3.05(2.50) \end{aligned} \quad (7.20)$$

The pseudo-ITSF from this fit is shown in Fig. (7.16a) where the values of the fit parameters

are indicated. We also show the $xq_V^\pi(x)$ -distribution in Fig. (7.16b). The perturbative kernel fixes the value of the pseudo-SF to be $\frac{1}{\pi^2}$ at $\nu = 0$ for any value of z , therefore the fitted value of $T_1(\nu, z^2)$ has zero uncertainty at this point.

For a proper extraction of the PDF and comparison to global fit results, one would need to extend the LO matching kernel in Eq. (4.48) to include a higher order scale dependent terms. A next-to-leading-order (NLO) matching kernel would include $O(\alpha_s)$ logarithmic z^2 and constant corrections. There also can exist higher twist effects which contaminate the results without sufficiently small z^2 . Extending the perturbative calculation beyond LO will not only lead to a more reliable extraction of the PDF, but also an understanding of power corrections and higher twist effects. A NLO matching kernel will give control over the corrections in z , both from DGLAP and higher twist effects. This calculation was performed on a fairly coarse lattice with a large minimum z , and in the future these corrections will need to be taken into account. There also exist potential discretization errors from the small separation size, as well as rotational symmetry breaking effects as observed in [79].

This first exploratory lattice QCD calculation of the pion PDF using spatially-separated current-current correlation function is performed at a relatively heavy pion mass ($m_\pi \simeq 416$ MeV). This calculation must be repeated on several other lattice ensembles to determine the pion mass dependence, quantify lattice artifacts such as finite lattice spacing and finite volume [80] corrections and obtain the PDF in the continuum limit. While such calculations are underway and will be presented in a future work, the limitations in our current extraction of the pion valence PDF do not preclude comparison with global fits, two different model calculations and recent lattice calculations of pion valence quasi-distribution.

For a comparison with the LO extraction of $q_V^\pi(x)$ from Drell-Yan experimental data

in Ref. [7], we evolve our lattice QCD determination of the PDF in LO to an evolution scale of $\mu^2 = 27 \text{ GeV}^2$ starting from initial scale of $\mu_0^2 = 1 \text{ GeV}^2$. With only a LO matching kernel, the initial scale μ_0 is chosen to be comparable to the $\frac{1}{|z|}$'s used in this calculation, but not low enough for perturbation theory to be doubted. With a NLO matching kernel, there will exist an explicit relationship between the scales z and μ_0 from the logarithmic terms. After the evolution, a shift in the peak of the $xq_v^\pi(x)$ -distribution toward smaller values of x and a more convex-up behavior of the distribution near $x = 1$ is seen as expected in our calculation. From the fit parameters in Eq. (7.18) ($\alpha = -0.34(31)$, $\beta = 1.93(68)$, and $\gamma = 3.05(2.50)$ at the initial scale), it is seen that this lattice QCD calculation of $q_v^\pi(x)$ is in agreement within uncertainty with the analysis in Ref. [9], where the authors included next-to-leading-logarithmic threshold soft-gluon resummation effects in the calculation of the Drell-Yan cross section. The large- x behavior is statistically consistent with the expectation based on perturbative QCD [81, 82, 83] but of course with large uncertainty. In contrast, the large- x behavior of this calculation has about $\sim 1\sigma$ difference from the two other NLO fits [8, 10] which obtained a harder $(1 - x)$ fall-off of the pion valence distribution.

It is seen in Fig. (7.17) that the large- x behavior of this calculation is statistically consistent with the Dyson-Schwinger model prediction [12] labeled as “DSE” in the momentum fraction region $x > 0.7$. On the other hand, this lattice QCD calculation of $q_v^\pi(x)$ is in statistical agreement with the light-front holographic QCD model calculation labeled as “LFHQCD” in the region $x < 0.5$, but shows a slightly softer fall-off at large- x in its central value. As mentioned earlier, in a future calculation, when all the systematics of this lattice QCD calculation are to be well understood and controlled in a proper way, the first-principles determination of large- x behavior of pion PDF such as this one can shed light for understanding different approximations used in an array of model calculations.

Even with the limitations mentioned above, this first exploratory lattice QCD deter-

mination of $q_v^\pi(x)$ using LCSs provides encouraging results and shows that this method has the potential to capture the essential dynamics dictating the behavior of hadron PDFs. Of notable interest, our calculation of $xq_v^\pi(x)$, shown in Figs. 7.16 and 7.17, illustrates a peak of the distribution in a region $x < 0.50$ at any scale μ^2 . This is consistent with all the global analyses of the pion valence distribution, wherein $xq_v^\pi(x)$ is peaked below $x = 0.50$. The readers are referred to [84, 85, 86] for recent other lattice calculations of the pion valence quasi-distribution.

7.4 Mock Data Tests

More sophisticated methods for solving the inverse problem often use some additional prior information to constrain the results in order to improve the robustness of the result and provide controlled systematic error estimates. Before applying inverse methods to the data, it is beneficial to study mock data sets first. The specific prior information and how it is given should be studied in a scenarios where the correct inverse is known. Using the software package LHAPDF [69] and the dataset `NNPDF31_nnlo_as_0118` from the NNPDF collaboration [6], we Fourier transform numerically the iso-vector quark PDFs and obtain data for the $\mathcal{Q}_R(\nu)$ at $Q^2 = 2 \text{ GeV}^2$. This data set will be called scenario A. A second dataset, scenario B, is created by multiplying the NNPDF data by Nx , where N is a normalization to fix the valence quark sum rule. This modification forces the PDF to vanish as x approaches zero. This scenario is reminiscent of the quasi-PDF case which is finite at low y before the matching procedure is applied. These data sets contain a set of replicas that can be used to obtain errors on the PDFs. The standard deviation over the ensemble of replicas represent the 68% confidence interval. The PDFs and Ioffe time PDFs for these two cases are plotted in Fig. 7.18. For the various tests, the Ioffe time PDF will be sampled in the ranges mentioned above to study the dependence on the maximum

available Ioffe time.

In the following, we discuss three extraction methods that do not presuppose a functional form of the encoded PDFs or have a very flexible functional form parametrizing the solution. Each of these methods regulates the inverse problem in different ways. The first, the Backus-Gilbert method, regulates the problem by minimizing the statistical variance of the solution. The second, a neural network parametrization, that provides a flexible parametrization given by specific choices of network geometry and activation functions. The third class of methods, is based on Bayesian inference, that rely on Bayes theorem to systematically incorporate prior information, such as e.g. positivity or smoothness of the solution. In addition, the Bayesian inference methods introduce a default model, which represent the prior information in that this model is the correct solution in the absence of any data. Non Bayesian approaches, while not explicitly mentioning prior information often incorporate additional knowledge about the problem at hand in an indirect fashion.

A preconditioning of the problem will be employed in order to improve the results of the Backus-Gilbert and neural network methods. In case that a relatively good guess of the form of the solution is available, the inverse problem can be recast into a different inverse problem which is easier to solve. The main ingredient is to define a rescaled kernel and rescaled PDF $h(x)$

$$K_j(x) \equiv \cos(\nu_j x) p(x) \quad \text{and} \quad h(x) \equiv \frac{q_v(x)}{p(x)}, \quad (7.21)$$

where $p(x)$ corresponds to an appropriately chosen function that makes the problem easier to solve. In particular, we wish to incorporate into $p(x)$ most of the non-trivial structure of $q(x)$, such that $h(x)$ is a slowly varying function of x and contains only the deviation of $q(x)$ from $p(x)$. In the following when preconditioning is used the preconditioning function

$p(x)$ will be of the form

$$p(x) = \frac{x^a(1-x)^b}{B(a+1, b+1)}. \quad (7.22)$$

With judicious choices of a and b , the preconditioning function can be used to remove any potential divergence at $x \rightarrow 0$ or to force the PDF to vanish as $x \rightarrow 1$. Different choices of a and b will be tested to study the dependence on that choice. In the Bayesian inference method such information would instead be supplied in the form of a default model.

Fig. 7.19 shows different functions which will be used for preconditioning function and default model. In both scenario A and B we find that the simple functional form (7.22) allows us to capture the overall features of the Ioffe time data very well. At intermediate- x values, the best fit in both scenarios deviates from the correct $q(x)$ but close to $x = 0$ provides a rather good description. This is exactly what we have in mind: the fit will provide us with prior information about the non-analytic behavior of the function, which we can use as default model in the Bayesian reconstruction. The Bayesian approach will then imprint the information encoded in the simulation data as deviations from the default model onto the end result.

7.4.1 Backus-Gilbert method

One approach to inverse problems, which has been used in a number of engineering and physics applications, is the Backus-Gilbert method [87, 88, 89, 90]. Like many approaches to the inverse problem, the Backus-Gilbert method provides a unique solution to the ill posed problem given some condition. This method differs from other regularizations, which impose a smoothness criterion on the resulting function $q(x)$, by imposing a minimization condition of the variance of the resulting function. The Backus-Gilbert method has been studied previously in [61] as a solution to an inverse problem for extracting PDFs from other types of lattice calculated hadronic matrix elements.

Let us start from the preconditioned expression with a rescaled PDF $h(x)$ that is only a slowly varying function of x . Hence our inverse problem becomes

$$d_j \equiv \mathcal{Q}_R(\nu_j) = \int_0^1 dx K_j(x) h(x). \quad (7.23)$$

The Backus-Gilbert method introduces a function $\Delta(x - \bar{x})$ that is written as

$$\Delta(x - \bar{x}) = \sum_j a_j(\bar{x}) K_j(x), \quad (7.24)$$

where $a_j(\bar{x})$ are unknown functions to be determined. It then estimates the unknown function as a linear combination of the data,

$$\hat{h}(\bar{x}) = \sum_j a_j(\bar{x}) d_j, \quad \text{or} \quad \hat{q}_v(\bar{x}) = \sum_j a_j(\bar{x}) d_j p(\bar{x}). \quad (7.25)$$

Given the above definitions, if $\Delta(x - \bar{x})$ were to be a Dirac δ function then $\hat{h}(\bar{x})$ would be equal to $h(\bar{x})$. If $\Delta(x - \bar{x})$ approximates a δ -function with a width σ , then the smaller σ is, the better the approximation of $\hat{h}(x)$ to $h(x)$. In other words, if $\hat{h}_\sigma(x)$ is the approximation resulting from $\Delta(x)$ with a width σ then

$$\lim_{\sigma \rightarrow 0} \hat{h}_\sigma(x) = h(x). \quad (7.26)$$

With this in mind the Backus-Gilbert method minimizes the width σ given by

$$\sigma = \int_0^1 dx (x - \bar{x})^2 \Delta(x - \bar{x})^2. \quad (7.27)$$

Note that other choices for the definition of the width can be used. This choice makes the resulting integrals easy to compute and the minimization problem becomes quadratic

in the unknown values a_j . Furthermore, if $\Delta(x)$ is to approximate a δ -function then one has to impose the constraint

$$\int_0^1 dx \Delta(x - \bar{x}) = 1. \quad (7.28)$$

Using a Lagrange multiplier λ one can minimize the width and impose the constraint by minimizing

$$\chi[a] = \int_0^1 dx (x - \bar{x})^2 \sum_{j,k} a_j(\bar{x}) K_j(x) K_k(x) a_k(\bar{x}) + \lambda \int_0^1 dx \sum_j K_j(x) a_j(\bar{x}). \quad (7.29)$$

By setting the derivative to zero with respect to $a_j(\bar{x})$ gives

$$\frac{\partial \chi[a]}{\partial a_j(\bar{x})} = 2 \int_0^1 dx (x - \bar{x})^2 \sum_k K_j(x) K_k(x) a_k(\bar{x}) + \lambda \int_0^1 dx K_j(x) = 0, \quad (7.30)$$

which results to

$$\int_0^1 dx (x - \bar{x})^2 \sum_k K_j(x) K_k(x) a_k(\bar{x}) = -\frac{1}{2} \lambda \int_0^1 dx K_j(x). \quad (7.31)$$

Let's now define the matrix $\mathbf{M}$ with matrix elements

$$M_{jk} = \int_0^1 dx (x - \bar{x})^2 K_j(x) K_k(x), \quad (7.32)$$

the vector $\mathbf{u}$ with components

$$u_j = \int_0^1 dx K_j(x), \quad (7.33)$$

and promote $a_j(\bar{x})$ to a vector $\mathbf{a}$. With these definitions the minimization condition takes the matrix equation form

$$\mathbf{M}\mathbf{a} = -\frac{1}{2} \lambda \mathbf{u} \quad (7.34)$$

or

$$\mathbf{a} = -\frac{1}{2}\lambda\mathbf{M}^{-1}\mathbf{u}. \quad (7.35)$$

Imposing the normalization condition for $\Delta(x)$ gives

$$\lambda = -2 [\mathbf{u}^T \mathbf{M}^{-1} \mathbf{u}]^{-1}, \quad (7.36)$$

therefore,

$$\mathbf{a} = \frac{1}{\mathbf{u}^T \mathbf{M}^{-1} \mathbf{u}} \mathbf{M}^{-1} \mathbf{u}. \quad (7.37)$$

An estimate of the unknown function may be obtained using Eq. (7.25). The width of the function $\Delta(x)$ which is otherwise known as the resolution function is also computable. The width of the distribution $\Delta(x - \bar{x})$ is given by

$$\sigma(\bar{x}) = \mathbf{a}^T \mathbf{M} \mathbf{a}. \quad (7.38)$$

Note that the width is dependent on $\bar{x}$ because $\mathbf{a}$ depends on $\bar{x}$ and represents the resolution of the method at $x = \bar{x}$. Features of the unknown function at scales shorter than $\sigma(\bar{x})$ will not be resolved and thus the method works better when the unknown function is smooth. This approach is known to fail when the matrix $\mathbf{M}$ becomes singular, and thus various techniques for regularizing the inversion of $\mathbf{M}$ have been proposed. As mentioned before, minimization of the variance of the resulting solution leads to a regularization of the matrix $\mathbf{M}$ with the covariance matrix of the data. This is implemented by the substitution

$$\mathbf{M} \rightarrow \mathbf{M} + \rho \mathbf{C}, \quad (7.39)$$

where $\mathbf{C}$ is the covariance matrix of the data and ρ is a small free parameter. The larger ρ the better conditioned the matrix $\mathbf{M}$ and the lower the resolution of the method is. A

similar method is the Tikhonov regularization [91, 92], where one makes the substitution

$$\mathbf{M} \rightarrow \mathbf{M} + \rho \mathbf{1}, \quad (7.40)$$

where $\mathbf{1}$ is the identity matrix and ρ a small adjustable parameter. In this case all the singular values of $\mathbf{M}$ are lifted resulting in a well-defined matrix $\mathbf{M}$. However, the Tikhonov regularization does not minimize the variance of the resulting solution of the inverse problem as the covariance matrix regularization does. Again larger values of ρ result in better regulated matrices at the cost of reduced resolution.

Another regularization method similar to Tikhonov is the method was adopted for the experiments. The singular vectors with singular values smaller than a chosen given cut-off ρ are projected out. For the problem at hand, the Tikhonov regularization and the SVD approach just described result in nearly identical solutions of the inverse problem. Furthermore, the covariance matrix regularization for appropriate values of the parameter ρ produces very similar results as the other two methods for our problem. Therefore, we decided to use our simple approach for regulating the matrix $\mathbf{M}$ and present results of our experiments using this SVD cutoff method.

The Backus-Gilbert method provides a unique solution to the inverse problem demanding that the solution maximizes a stability criterion. The method has a tunable free parameter, ρ , which provides a trade off between stability and resolution. This freedom can lead to a bias from the user, but the Backus-Gilbert method provides the variance and the resolution function as quantitative measures to assure a proper analysis is performed. It should be noted that the forms of preconditioning functions, $p(x)$, are restricted in the Backus-Gilbert method. The preconditioning function must be chosen such that integrals of the preconditioned kernels which define $\mathbf{M}$ and $\mathbf{u}$ remain finite. Furthermore, it should be noted that a good preconditioning function makes the remainder smooth, as a result

the demand for a small resolution is smaller the better the preconditioning is.

All integrals are computed numerically in double precision to tolerance 1.0×10^{-16} . In order to regularize the matrix $\mathbf{M}$ an SVD cut-off of $\rho = 1 \times 10^{-8}$ is used as discussed in pervious sections ¹. With this set up, a resolution function width, $\sigma(\bar{x})$, of $\mathcal{O}(10^{-2})$ is obtained for most of $x < 0.75$ gradually increasing after that reaching $\mathcal{O}(10^{-1})$ for $x = 1.0$. With the ensemble of replicas, one can also compute the covariance matrix of the data and therefore use the covariance matrix regularization for the Backus-Gilbert method. We have done that as well as employ the Tikhonov regularization and obtained results that are similar with the results we discuss here where the SVD cutoff method was used to regularize the matrix inverse. In order to obtain the statistical error-band and the mean for the reconstructed curve we take the standard deviation and the mean of the ensemble of reconstructed curves. The resulting statistical error-band represents a 68% confidence level.

After performing several numerical experiments on the Backus-Gilbert method, we have concluded that the use of a preconditioning function $p(x)$ is essential in obtaining an accurate determination of the PDF. The reconstructions of the PDF from data with $\nu_{max} = 10$ and no preconditioning function are shown in Fig. 7.20. The reconstructed PDF is consistent with the mock PDF for $x > 0.3$, but deviates from it in the low x region. This result is expected, because the low x region is dominated by the Ioffe time distribution at large Ioffe times which are not present in the input data. Also in Fig. 7.20, the fidelity of the reconstruction is tested by taking the Fourier transform of the reconstructed PDF and comparing to the data used to generate it which shows agreement across the range of Ioffe time.

As argued earlier, preconditioning is essential in improving the Backus-Gilbert extraction of the PDF. In order to test its effectiveness, the preconditioning function defined

¹We have also tested using the Moore-Penrose pseudo inverse instead and got similar results.

in Eq. (7.22) is used with a range of exponents a and b that allows for all the integrals that define $\mathbf{M}$ and $\mathbf{u}$ to be convergent. For scenario A we achieved the best reproduction with $a = -0.35$ and $b = 2$ while for scenario B with $a = 0.3$ and $b = 2$. In order to get a handle of possible systematics due to preconditioning we varied the exponents a and b in the intervals $[-0.25, -0.4]$ and $[1, 4]$ respectively for scenario A. Correspondingly for scenario B we varied the exponents in the intervals $[0.2, 0.35]$ and $[1.5, 3]$.

As one can see, for both scenarios, a wide range in x is well reproduced and deviations from the original data begins to appear for $x < 0.2$. Perhaps, these deviations are expected due to the rather small maximum value of ν used in this example, $\nu_{\max} = 10$. However, this aggressive cutoff for ν is realistic for lattice calculations. In addition, the number of selected points, which is 12, is also a number that is plausible for present lattice QCD calculations. From these results we can conclude that the fidelity of the reconstruction can be improved significantly by choosing appropriate preconditioning functions. The best reconstruction is obtained using a preconditioning function that is chosen to roughly fit the data. In this case the resulting reconstruction is indistinguishable from the original data for nearly all the range of x with the largest deviations occurring for $x < 0.1$. Therefore, we conclude that the Backus-Gilbert reconstruction with an appropriate preconditioning function is well suited for the reconstruction of the PDFs from the limited data for the Ioffe time PDFs that are provided by present day lattice QCD calculations.

Finally, we explore the dependence of the reconstruction on the maximum value of ν . Experiments are performed with $\nu_{\max} = 20$ and $\nu_{\max} = 100$. In all cases the data points were selected to be uniformly spread in the interval with a separation $\delta\nu = 10/11$. The results are shown in Fig. 7.22. The best result that is in agreement with the input PDF within errors across the whole range of x is obtained from the experiment with $\nu_{\max} = 100$. However, this is an unrealistically large range in ν unlikely to be accessible in contemporary lattice QCD calculations. As $\nu_{\max}$ is lowered then the reconstruction

deteriorates. However, even for $\nu_{max} = 20$ the results seem to agree with the input with the exception of one small region $x = [0.05, 0.1]$ where we observe a one σ deviation. Overall, the $\nu_{max} = 20$ result is only very slightly better than the case of $\nu_{max} = 10$ which was presented in Fig. 7.21. Therefore, we conclude that extending the range of ν slightly beyond $\nu = 10$ will have only small improvement in the determination of the PDF. Our conclusion from this analysis is that 12 points for $\nu \in [0, 10]$ seems to be adequate to obtain a good reconstruction of the underlying PDF. This finding is rather encouraging for present day calculations as it is possible to cover the range of $\nu \in [0, 10]$ with about 12 points in lattice QCD calculations.

7.4.2 Neural Network parameterization

Neural networks can be used as rather flexible parametrizations of unknown functions and thus have been used in the literature to address various inverse problems. Neural networks do not provide any explicit constraints such as a smoothness condition or minimization of a variance, as the Backus-Gilbert method does but instead it is the depth and structure of their layers that limits what functional forms may be encoded. Hence, neural networks provide a very general parametrization of the unknown PDF, without forcing a model dependent form such as Eq. (2.25), for a statistical regression. Neural networks have e.g. been used in the literature to extract PDFs from experimental data. The pioneering work of [93] is now one of the established approaches in obtaining PDFs from cross section data by the NNPDF collaboration [6, 94, 95, 96]. In the following, the use of neural networks in solving the inverse problem at hand is explored. The specific neural network implementation used in this work is known as a multilayer feedforward neural network.

A neural network is composed of a system of interconnected nodes, called the neurons. The connection between neuron pairs is called the synapsis. The output of each

neuron, called the activation, is typically a non-linear transformation of its input, which is called the activation function of that neuron. The input for each neuron is a weighted sum of the activations of the connected neurons shifted by some real number, which is called the threshold or bias of the neuron. The weights and the thresholds are the free parameters which will be determined by some process which is called the training procedure.

In a multilayer neural network, the neurons are organized into distinct layers whose neurons are not directly connected to each other. In a multilayer feedforward network, the layers are ordered and each neuron on a given layer is only connected to the neurons on the previous and subsequent layers. The geometry of these networks, shown in Fig. 7.23, is described by a list of numbers, $N_1, \dots, N_L$, giving the number of neurons on each of the L layers. The first layer, $\xi^{(1)}$, is a given input vector of length N_1 and the final layer, $\xi^{(L)}$ is the output vector of length N_L , also called the response of the neural network. All other layers are referred to as hidden layers. The activation of the i -th neuron on the l -th layer of the network is given by the recursive relationship

$$\xi_i^{(l)} = g_i^{(l)} \left(\sum_j^{N_{l-1}} w_{ij}^{(l)} \xi_j^{(l-1)} - \theta_i^{(l)} \right). \quad (7.41)$$

Specified by the geometry, activation functions, thresholds, and weights, a multilayer feedforward neural network can be considered a parametrization of a function from $\mathbb{R}^{N_1} \rightarrow \mathbb{R}^{N_L}$. For a given geometry and a set of activation functions, the thresholds and weights can be chosen for the network to approximate a given continuous function. For the case of reconstructing the rescaled PDF $h(x)$, which is a single valued function with a single argument, the geometry is restricted to have $N_1 = 1$ and $N_L = 1$. It should be noted that NNPDF has also used $\log \frac{1}{x}$ as a second argument in the input layer [97], which is important to reproduce the small x behavior of the PDF. However, in our case our data are not sensitive to small x and thus using $\log \frac{1}{x}$ at the input layer is not essential.

A neural network can be used to perform a regression by choosing the thresholds and weights with a training procedure. During the training procedure, the weights and thresholds are modified to minimize some error function, which describes the difference between the response of the neural network and some desired output. When using a neural network to perform a statistical regression, a common choice of error function is the χ^2 function, e.g.

$$\chi^2(\{w\}, \{\theta\}) = \sum_{k=1}^N \left(\mathcal{Q}_k - \int_0^1 dx K_k(x) h(x; \{w\}, \{\theta\}) \right)^2 / \sigma_k^2, \quad (7.42)$$

where $\mathcal{Q}_k$ are N data points with standard deviations σ_k and h is the output layer of the neural network given an input layer x , weights $\{w\}$, and thresholds $\{\theta\}$. Using a neural network to parametrize the unknown function may result in a $\chi^2(\{w\}, \{\theta\})$ with a large number of local minima. Some of these local minima are trivial multiplicities due to symmetries a neural network has under permutations of the weights and thresholds. However, the possibility of multiple non-trivial local minima exists resulting in many realizations of the network that reproduce the data equally well. In these cases special care has to be taken to avoid “over-fitting” and several methods to do so have been developed in the literature [97].

These roughly equivalent minima can be found by a training procedure such as a genetic algorithm. A genetic algorithm is an iterative process based upon the idea of natural selection. Each iteration, also called a generation, begins with a sample of possible networks, called a population. A fitness function is evaluated for each of the networks, which in this case is the error function $\chi^2(\{w\}, \{\theta\})$. Those networks which are the “least fit”, i.e. largest χ^2 , are removed from the population. The surviving population is then “mutated” by randomly changing their parameters, i.e. weights and thresholds, to create the starting population for a new generation. This procedure is iterated for enough generations that a

final population covers a sufficient number of minima with sufficiently small values of the error function.

The genetic algorithm used in this study is based upon simulated annealing. The initial population is created from N_{rep}^0 sets of initial weights and thresholds which are generated from a random normal distribution with a wide initial search width, σ_0 .

$$\{w\}_i^{(0)}, \{\theta\}_i^{(0)} \sim \mathcal{N}(0, \sigma_0) \quad (7.43)$$

where $\{w\}_i^{(0)}$ and $\{\theta\}_i^{(0)}$ are the parameters of the i -th neural network in the initial population. The weights and thresholds of this initial population of neural networks is the starting point for minimizing with respect to χ^2 to some new values $\{w'\}_i^{(0)}$ and $\{\theta'\}_i^{(0)}$. The resulting neural networks which are the least fit, i.e. that have the largest χ^2 , are removed from the population and not used in the next generation. The surviving neural networks are then mutated by adjusting their weights and thresholds by some small, Gaussian distributed amount with a mutation width σ ,

$$\{w\}_i^{(g)} \sim \{w'\}_i^{(g-1)}(1 + \eta\mathcal{N}(0, \sigma)) \quad \{\theta\}_i^{(g)} \sim \{\theta'\}_i^{(g-1)}(1 + \eta\mathcal{N}(0, \sigma)), \quad (7.44)$$

where η is the relative size of the mutation, $\{w\}_i^{(g)}$ and $\{\theta\}_i^{(g)}$ are the initial parameters of the i -th neural network in the g -th generation and $\{w'\}_i^{(g)}$ and $\{\theta'\}_i^{(g)}$ are the parameters of the i -th surviving neural network in the g -generation after minimization. This procedure of minimization and removal is repeated for N_{gen} generations. After enough generations have passed, the population of neural networks will all exist in various minima of χ^2 which can be used to estimate the PDF.

Many other training procedures exist which all have different methods of identifying minima, both stochastic and deterministic. An example of such a procedure is the ap-

proach the NNPDF collaboration is taking in obtaining PDFs from experimental data [97]. Though different training procedures vary in their efficiency for finding minima, they should eventually all converge on the same set of minima of the error function.

Neural networks provide a rather general solution to the inverse problem. The large number of highly interconnected parameters allows for model independent result, but the complexity of the hidden layers do not give any insight to what effects underlie that result.

In this section, the neural network method described above is tested to reconstruct the mock PDFs. The data used for this reconstruction is the mock scenario A and B in the smallest range of ν , $\nu \in [0, 10]$, discretized into $N_\nu = 12$ points. For this study, we chose the hyperbolic tangent as activation function for all the nodes in the hidden layers. The activation function for the final layer is linear with the threshold value fixed to zero. For a single hidden layer of size N_2 , the neural network parametrizes the PDF as

$$q(x) = \sum_{i=1}^{N_2} w_{1,i}^{(3)} \tanh(w_{i,1}^{(2)} x + \theta_i^{(2)}), \quad (7.45)$$

while for two hidden layers of sizes N_2 and N_3

$$q(x) = \sum_{j=1}^{N_3} w_{1,j}^{(4)} \tanh \left(\sum_{i=1}^{N_2} w_{j,i}^{(3)} \tanh(w_{i,1}^{(2)} x + \theta_i^{(2)}) + \theta_j^{(3)} \right). \quad (7.46)$$

The neural network behavior is governed by the collective interactions of all the weights, thresholds, and connections, not on any individual neurons. Slight modifications of the geometry, such as adding or removing a neuron, will have little change in the final result. To test this point, 1-3-1, 1-4-1, and 1-2-2-1 geometries are tried, all resulting in similar PDFs. No preconditioning function will be used in this test. A preconditioning function of the form in Eq. (7.22) can be set to prefer certain features, such as the PDF vanishing in the limit $x \rightarrow 1$ by using $b > 0$ or a divergence as $x \rightarrow 0$ by using $a < 0$.

Remarkably the neural network was able to reproduce both these high and low x features without the need of any constraint or prior information.

The training procedure described above has a number of tunable parameters, but the end result does not depend very strongly on any particular parameter. For this study, the initial search width, σ_0 , is set to 5, the mutation width, σ , is set to 1, and the mutation size, η , is set to 0.05. From most starting points in any given generation, the minimization procedure sends the χ^2 to very small values. In case this does not occur, the minimization is restarted. Three checks on χ^2 are performed and if the value is too large then the mutation and minimization steps are performed again. First if the χ^2 is greater than 10^{-2} , then the mutation and minimization steps are repeated with the same mutation width and size. Second if the χ^2 is greater than 10^{-1} , then the mutation and minimization steps are repeated with the same mutation width and a size of 0.025. Third if the χ^2 is still greater than 10^{-1} , then the mutation and minimization steps are repeated with the same mutation width and a size of 0.0125. The above minimization strategy was used in two stages. First we used the global average of all available data and began with 45 initial neural networks and removed 2 every generation for 10 generations until 25 neural network replicas remained, $\{q_i(x)\}$. The response of the surviving population is shown in Fig. 7.24 for each of the geometries studied. The standard deviation of the response of these neural networks is used to estimate the systematic error in this method of regression.

$$\sigma_{sys} = \text{StdDev}[q_i(x)] \quad ; \quad i = 1 \dots N_{rep} \quad (7.47)$$

In the second stage, we perform the minimization on each of the $N = 100$ mock Ioffe time PDF replicas ² in order to estimate the statistical fluctuations of the resulting

²These the replicas the NNPDF collaboration provides in their data set. In a realistic lattice calculations these replicas would be a set of jackknife or bootstrap samples of the matrix elements.

function. Starting with each of the networks in the final population, each PDF replica is used as the training data for another minimization to create the PDF replica dependent networks of the final population, $\{q_i^{(b)}(x)\}$. The variance of the response averaged over the members of the final population across the different PDF replica dependent networks is used to estimate the statistical error,

$$\sigma_{stat} = \text{StdDev}\left[\frac{1}{N_{\text{rep}}} \sum_{i=1}^{N_{\text{rep}}} q_i^{(b)}(x)\right] \quad ; \quad b = 1 \dots N. \quad (7.48)$$

The average response of the PDF replica dependent networks and the combined systematic and statistical errors are shown in Fig. 7.25. As anticipated, the errors tend to grow at small- x . Since the input data is over a truncated region of ν , the information of the low- x behavior is lost. The neural networks are able to precisely reconstruct the large- x region while in the low- x region the flexible parametrization allows for a wide range of functions that can reproduce the data. One can also see that the various network geometries all give comparable results for the reconstructed PDF.

Next we consider how a preconditioning function affects the reconstructions. Using the function defined in Eq. (7.22), the preconditioning function for scenario A with $a = -0.25$ and $b = 2$ and for scenario B with $a = 0.3$ and $b = 2$. Both of these cases were tested on the 1-3-1 geometry. The reconstructed PDFs, shown in Fig. 7.27, are not significantly different for the majority of large x region. For scenario A, the low x behavior is slight improved compared to the case without preconditioning. This behavior is perhaps to be expected, because the preconditioned neural network is parameterizing a much slower varying function in that region compared to the divergent PDF.

Our conclusion is that for ranges of ν which are realistic in modern lattice calculations, a neural network is capable of reconstructing the PDF for a wide range of x without the need of preconditioning, though the neural networks can benefit from it. This study of

neural network parametrization of the PDF is by no means complete. There are a number of choices that were made which may not be optimal. It is clear that a more dedicated study of this approach is required in order to understand its full potential.

7.4.3 Bayesian Reconstruction

A final approach to inverse problems we discuss, which has proven to be versatile in practice, is Bayesian inference. It acknowledges the fact that the inverse problem is ill-defined and a unique answer may only be provided once further information about the system has been made available. This method does not require any explicit constraints, though it may benefit from them, nor does it require any functional form, even one as flexible as a neural network. This method finds the most probable value of $q(x)$ given the data and whatever prior information is provided. Other regularization methods may often be rewritten in terms of Bayesian reconstruction with the constraints treated as prior information.

Formulated in terms of probabilities, one finds in the form of Bayes theorem that

$$P[q|\mathcal{Q}, I] = \frac{P[\mathcal{Q}|q, I]P[q|I]}{P[\mathcal{Q}|I]}. \quad (7.49)$$

It states that the so called posterior probability, $P[q|\mathcal{Q}, I]$, for a test function q to be the correct x -space PDF, given lattice calculated Ioffe time PDF $\mathcal{Q}$ and additional prior information I may be expressed in terms of three quantities. The likelihood probability $P[\mathcal{Q}|q, I]$ denotes how probable it is to find the data $\mathcal{Q}$ if q were the correct PDF. Obtaining the most probable q by maximizing the likelihood is nothing but a χ^2 fit to the $\mathcal{Q}$ data, which as we saw from the direct inversion is by itself ill-defined. The second term of importance is the prior probability $P[q|I]$, which quantifies, how compatible our test function q is with respect to any prior information we have. In particular such information

can be related to the appearance of non-analytic behavior of $q(x)$ at the boundaries of the x interval or the positivity of the PDF. $P[\mathcal{Q}|I]$, the so called evidence represents a q independent normalization.

For sampled (lattice) data, due to the central limit theorem, the likelihood probability may be written as the quadratic distance functional $P[\mathcal{Q}|q, I] = \exp[-L]$,

$$L = \frac{1}{2} \sum_{k,l} (\mathcal{Q}_k - \mathcal{Q}_k^q) C_{kl}^{-1} (\mathcal{Q}_l - \mathcal{Q}_l^q), \quad (7.50)$$

where $\mathcal{Q}_k^q$ are the Ioffe-time data arising from inserting the test function q into Eq.(2.23) and

$$C_{kl} = \frac{1}{N_m(N_m - 1)} \sum_h (\mathcal{Q}_k^h - \mathcal{Q}_k)(\mathcal{Q}_l^h - \mathcal{Q}_l), \quad (7.51)$$

denotes the covariance matrix of the N_m measurements of simulation data $\mathcal{Q}_k^h$.

For the regularization of the inversion task, we now have to specify an appropriate prior probability $P[q|I] = \exp[\alpha S[q, m]]$. Prior information enters in two ways here: on the one hand the (unique) extremum of the S functional is given by the so called default model m , which, by means of Bayes theorem, represents the most probable answer in the absence of data. On the other hand the functional form of S encodes which solutions are admissible to the inverse problem, e.g. enforcing positivity.

In a Bayesian approach the preconditioning discussed in the beginning of this section is naturally incorporated not via a modification of the Kernel but via the specification of an appropriate default model. Actually dividing out a fitted Ansatz $p(x)$ from the Kernel and using a constant default model is equivalent to working with the original Kernel and simply assigning $m(x) = p(x)$. In other words, in the Bayesian approach we select the most appropriate default model from a best fit of Eq. (2.26) to the Ioffe time data and

estimate the dependence of the end result $q_{\text{Bayes}}(x)$ on the choice of m by repeating the reconstruction with default models arising from a slight variation of the best fit parameters in Eq. (2.25).

Based on different sets of prior information, different regularization functionals have been derived in the literature. Note that due to Bayes theorem, in the “Bayesian continuum limit” i.e. the combined limit of the number of supplied datapoints becoming infinite and the uncertainty on the data going to zero, all reconstructions will lead to the same solution. A good choice of regulator functional and the availability of valid prior information can help us approach this solution more closely even for data that are coarse and noisy. Indeed if an accurate approximate solution is already available we can use Bayesian methods that imprint this prior information strongly onto the end result (steep $S[q]$), while in cases where only unreliable prior information is available one wishes to use a method that let’s the data “speak” as freely as possible without distorting them through the default model (weaker $S[q]$).

The reconstruction of PDF’s from Ioffe-time PDF data benefits from the availability of good prior information in the form of the empirically obtained Eq.(7.22). This situation is quite different from other inverse problems, e.g. the reconstruction of hadronic spectral functions, where little to no relevant information about the spectral structures of interest is known. We therefore foresee that methods with a steep prior probability will fare better than those that are explicitly designed to minimize the impact of the default model.

The first example is the well known Maximum Entropy Method [98, 99], which features the Shannon-Jaynes entropy as regulator

$$S_{SJ}[q, m] = \sum_n \Delta x_n \left(q_n - m_n - q_n \log\left(\frac{q_n}{m_n}\right) \right). \quad (7.52)$$

Its functional form has been derived using arguments from two-dimensional image recon-

struction and it is designed to introduce as least as possible additional correlations into the end result, beyond what is encoded in the input data. The standard MEM implementation restricts the space of solutions to those close to the default model, meaning that for a small number of available input points its result will lie close to the function m . While it has been shown [100] that this implementation of the MEM in general fails to recover the global extremum of the posterior, for a default model that is close to the correct result, the outcome is often satisfactorily accurate.

We contrast the MEM to another Bayesian approach simply named Bayesian reconstruction (BR) [101], which has been derived by requiring positive definiteness of the resulting q , smoothness of q , where the data $\mathcal{Q}$ do not provide constraints on its form, as well as independence of the resulting q on the units used to express $\mathcal{Q}$. The resulting regulator reads

$$S_{BR}[q, m] = \sum_n \Delta x_n \left(1 - \frac{q_n}{m_n} + \log\left(\frac{q_n}{m_n}\right) \right). \quad (7.53)$$

It has been shown that this regulator also leads to a unique extremum of the posterior and since in contrast to the MEM no flat direction appears in S_{BR} , the search space does not need to be restricted. The BR method had been developed in particular for inverse problems, where only scarce or unreliable prior information is available, imprinting the form of m as weakly as possible onto the end result. We expect thus that in the present case with good prior information available, this method will produce less accurate results than the MEM, as it does not leverage the information in m as strongly.

Note that in the definition of $P[q|I]$ we introduced a further parameter α , a so called hyperparameter, which weighs the influence of simulation data and prior information. It has to be taken care of self-consistently. In the MEM it is selected, such that the evidence has an extremum. In the BR method, we marginalize the parameter α a priori, i.e. we

integrate the posterior with respect to the hyperparameter, assuming complete ignorance of its values $P[\alpha] = 1$.

Up to this point we have only considered regulators for the reconstruction of positive definite functions. It is fathomable though that one has to deal with PDF reconstructions where positive definiteness does not hold, such as for spin dependent PDFs. In that case the regulator (7.53) is not applicable. A choice of regulator often employed in the literature in such a situation is the quadratic one, which corresponds to a modified form of Tikhonov regularization

$$S_{QDR}[q, m] = \sum_n \Delta x_n \left(q_n - m_n \right)^2. \quad (7.54)$$

It is a comparatively strong regulator and imprints the form of the default model significantly onto the end result. Marginalizing the hyperparameter α with a flat prior probability $P[\alpha] = 1$ is hence not possible in this case. Indeed integrating over the normalized prior probability with respect to α leads to a delta function that fixes the end result to $\rho = m$ everywhere. Instead one uses the "historic MEM" approach [98, 99], where α is chosen such that for the corresponding reconstruction q^α the likelihood takes on the value $L[q^\alpha] = N_\nu$. As in the following we are able to utilize default models, which are already close to the correct result, the use of this regulator is justified.

Again we can compare this approach to one specifically designed to keep the influence of the default model to a minimum, i.e. to imprint its functional form as weakly as possible on the end result. To this end we resort to an extension of the BR method (7.53) to non-positive functions. The corresponding

$$S_{BRg}[q, m] = \sum_n \Delta x_n \left(-\frac{|q_n - m_n|}{h_n} + \log\left(\frac{|q_n - m_n|}{h_n} + 1\right) \right), \quad (7.55)$$

keeps the advantageous properties of the original BR prior, e.g. smoothness and scale invariance at the price of having to introduce another default model related function h . It may be thought of as the confidence we have in our default model. Note that for $h = m$ one obtains a regulator, which for $q > 0$ and $q > m$ takes on the form of the standard BR prior and mirrors it for $q < m$. In order to obtain the full uncertainty budget of the end result obtained with the generalized BR prior, one has to vary not only the default model but also the function h . Eq. (7.55) also admits for an a priori integrating out of α .

Once L , S and m have been provided, the most probable PDF q , given simulation data and prior information is obtained by numerically finding the extremum of the posterior

$$\left. \frac{\delta P[q|\mathcal{Q}, I]}{\delta q} \right|_{q=q_{\text{Bayes}}} = 0. \quad (7.56)$$

It has been proven that if the regulator is strictly concave, as is the case for all the regulators discussed above, there only exists a single unique extremum in the space of functions q on a discrete ν interval.

Note that as long as positive definiteness is imposed on the end result, the space of admissible solutions is significantly reduced. On the other hand, regulators admitting also q functions with negative contributions have to distinguish between a multitude of oscillatory functions, which if integrated over, resemble a monotonous function to high precision. We will observe the emergence of ringing artifacts for the quadratic and generalized BR prior below.

Other regularizations may be reformulated in the spirit of the Bayesian strategy. To give meaning to the inversion task they introduce in addition to the likelihood another functional which encodes prior information or prior expectations on the end result. In the standard Tikhonov approach e.g. the condition is that the values of q shall have the smallest possible magnitude. While they do not rely on an explicitly formulated default

model, they contain the prior information implicitly in the form of their regularization functional. Otherwise they could not provide a unique regularized answer to the task. The benefit of the Bayesian method is that all prior information is made explicit, so that the dependence of the end result on the regularization may be thoroughly and fully tested.

Finally, mock data tests of the Bayesian strategy outlined above are carried out to determine the feasibility of extracting the x -space PDF $q(x)$ from the Ioffe-time data $Q(\nu)$ in realistic settings for applications to lattice generated data. The x interval is discretized in $N_x = 2000$ steps to have fine enough resolution of the possibly highly oscillating cosine. To remain as close as possible to the situation encountered with real data, we first carry out a simple fit on the noisy mock Ioffe-time data, based on the simple Ansatz Eq. (2.25) via the expression in Eq. (2.26).

Note that in order to use the best fit $q(x)$ as default model we need to modify its functional form at the boundaries of the x -interval in both cases. If there is e.g. a pole at the origin, then the value of m is not well defined there. On the other hand if the function vanishes exactly at the origin, then the prior assumption of positive definiteness in some of the Bayesian methods is not fulfilled. Therefore we set $m(0) = q_{\text{best fit}}(\Delta x)/2$. To further understand how strongly our end result depends on the choice of default model, we also carry out reconstructions with other functions than the best fit one, two of which are obtained from varying the best fit parameters by factors of two and one function simply being the constant $m(x) = 1$.

We start with the most realistic and most challenging setting, where the input data ν is in the range I_1 , i.e. $\nu = [0, 10]$ discretized with $N_\nu = 12$ points, very similar to what is currently available in lattice QCD simulations. Due to the availability of a good approximate solution obtained from the fit with eq.(7.22), provided as default model, we first consider the Maximum Entropy Method, whose results are presented in Fig.7.28

The original mock $q(x)$ is shown on the top of Fig. 7.28 as a gray dashed line, and the

MEM reconstruction based on the best fit default model in red. The red errorband denotes the statistical uncertainty, while the blue band arises from the default model dependence. In the bottom row panels the input Ioffe-time data (gray) as well as the data corresponding to the Bayesian reconstruction (red) are shown. We find that the reproduction for both scenario A and scenario B is excellent, i.e. the MEM is able to utilize the approximate solution provided by the default model to regularize the inverse problem and reproduce the input PDF within uncertainties. If one inspects the reconstruction in scenario B carefully around the maximum one finds that the MEM solution slightly underestimates the mean below $x = 0.25$ and overestimates it for $x > 0.25$. The deviation however is minute.

Let us compare how the BR method fares. In this method the information of the default model is more weakly incorporated. The corresponding results are shown in Fig.7.29 and yield, as expected, a less accurate reproduction of the input data. While for scenario A it reproduces the PDF within uncertainties, the reconstruction of scenario B shows undulations around the correct result. Close to the maximum at intermediate x the mean is found to be underestimated.

This result is not surprising but the appearance of differences in the results from different Bayesian regulators urges us to figure out how far we have to improve the input data for both results to provide a reproduction of the correct result within uncertainties. To this end we have also carried out reconstructions based on the larger interval $\nu = [0, 20]$ with the same number of $N_\nu = 12$ Ioffe-time data-points. The results of the MEM and BR method are shown in Fig.7.30 and reflect that the improvement in input data has significantly reduced the errors of the reconstruction so that now both methods lie within the uncertainty of the original input data. It is still the case that the MEM utilizes the provided prior information to a stronger extent than the BR method and produces slightly more accurate results in this setting. A further extension of the interval to $\nu = [0, 10]$ and increasing the number of available points to $N_\nu = 100$ confirms this trend, improving the

accuracy of the reconstructions in particular around the maximum of the PDF in scenario B.

Let us turn our attention now to the Bayesian reconstruction methods, which do not require the positivity of the PDF. The first uses the quadratic prior combined with the “historic MEM” procedure to choose the hyperparameter α . Results for both the realistic PDF mock scenario A (left column) and the mock scenario B (right column) are given in Fig. 7.31. Since this method also imprints the default model information strongly on the end result, it is not surprising that the reconstruction of scenario A also works excellently here. For scenario B we find that while the overall shape has been reproduced in an acceptable manner, the small $x < 0.1$ region is overestimated, while close to the maximum of the PDF the solution underestimates the correct value. It is reassuring to see that with only $N_\nu = 12$, the presence of accurate prior information is able to efficiently suppress the many possible oscillatory solutions, which could also reproduce the provided input data.

Let us consider as a last item the reconstruction based on the generalized BR method with $h = m$, whose results are given in Fig. 7.32. To understand the performance of this method, we note that it has been designed for cases, where no reliable prior model is available and where thus the influence of m on the end result needs to be held as small as possible. We find that for both scenarios A and B the reconstructions are, as expected, less accurate than for the quadratic prior. In particular we see that excursions into negative values occur at some x , which in turn lead to overestimation of values at different x . Note that the obtained reconstructions however lead to Ioffe-time data that reproduce the input data very accurately.

It is possible to improve the accuracy of the generalized BR method by changing the value of the default model confidence function h , which essentially makes the regulator steeper and steeper, i.e. imprinting the default model more and more strongly. Our intention of showing the generalized BR method with a weak regulator is to emphasize the

role accurate prior information has in approaching the correct reconstruction result.

In summary, we find that due to the availability of accurate prior information on the true shape of the PDF via fits with eq.(7.22), Bayesian methods that are designed to imprint prior information strongly on the end result outperform those that minimize the influence of the default model. I.e. in case of a rather small interval $\nu = [0, 10]$ and number of data points $N_\nu = 12$, the Maximum Entropy Method provides the most accurate results for both scenario A and B (Fig. 7.28). The BR method is competitive in scenario A but shows a larger deviation from the true result for scenario B (Fig.7.29). Increasing the extent of the ν interval and the number of data points consistently shows that both methods approach the correct solution in the “Bayesian continuum limit”.

When comparing methods without the requirement of positivity we find that the quadratic prior still fares excellently with scenario A but shows some deviations, in particular at small x values in scenario B. As expected and by construction, the generalized BR method with the choice $h = m$ shows the weakest dependence on the used default model but at the same time also provides the least accurate reconstructions.

The success of the MEM relies on the availability of very good prior information. On the other hand the BR method is constructed to imprint the default model on the end result in a weaker fashion than the MEM (lower curvature of the prior functional). Hence we plan to deploy it together with the MEM when investigating actual lattice QCD data in the future. The reason is that as long as the asymptotics of the PDF at $x = 0$ and $x = 1$ are adequately provided, the BR method is expected to be able to maintain its reconstruction quality in case that in a genuine lattice data based reconstruction the quality of the default model at intermediate x values will be worse than in the above tests.

If on the other hand a non-positive q function is concerned and accurate prior information is available, then we can utilize the quadratic prior even in cases that only small number of data points are available. The generalized BR method to become competi-

tive requires a more careful treatment of the default model confidence function h in this context.

7.4.4 Restricted χ^2 sampling

In order to clarify the role of different prior information on the success of the reconstruction, we perform a further numerical experiment on our mock data sets. Here we consider positivity as the only property of the PDF known a priori. The ill-posedness of the inversion is related to the many local extrema in the likelihood. If we work with a constant prior probability, the resulting posterior probability will just be a flat distribution, unable to provide meaningful insight. This fact has been explicitly tested and confirmed. It is interesting to then ask, whether the restriction to positive PDF's limits the number of local extrema in L , such that all of their contributions taken together in a statistical fashion will lead to a meaningful posterior?

To explore this idea, the MC-Stan library, a modern tool for statistical inference, can be used. It implements an efficient hybrid Monte-Carlo algorithm, specifically a no-U-Turn sampler, which allows the sampling of a wide variety of joint probability distributions of random variables. The applicability of the machinery of HMC arises from identifying its Hamiltonian with the logarithm of the joint probability distribution (for technical details see Ref. [102]). Using its high level programming language, the test can be formulated by identifying each of the (appropriately decorrelated) Ioffe-time mock data points with an individual Gaussian distribution. Its mean is expressed in term of the PDF $q(x)$ and the cosine integral Kernel, the spread by the corresponding data uncertainty. Positivity of the PDF is enforced by restricting its sampling to values larger than zero a priori. The complete STAN model reads

```
data {
```

```

    int NNu; int Nx;  matrix[NNu, Nx] Kernel;

    vector[NNu] Q;

    vector[NNu] Uncertainty;
}

parameters {
    vector<lower=0>[Nx] q;
}

model {
    Q ~ normal(Kernel * q, Uncertainty);
}

```

For the Kernel we discretize the x-range with $N_x = 1000$ points and consider two sets of input data for each scenario A and B. As realistic test case we use again the a $\nu \in [0, 10]$ with $N_\nu = 12$ points (see Fig. 7.33), exactly as has been used in the Bayesian reconstruction, as well as the ideal case of $N_\nu = 100$ points with $\nu_{\max} = 100$ (see Fig. 7.34). The sampling is performed over 20 chains of HMC trajectories, each with 1500 steps in MC time, 500 of which are discarded as warmup phase. The mean value of $q(x)$ over the 20 chains is shown as the red data points, while its erroband is obtained from the usual variance among different chains. As we do not imprint any further prior information, execept positivity, and we assume that positive holds exactly, there is no further systematic error to be considered for this reconstruction.

We find that positivity alone is a powerful regulator, as sampling of the restricted χ^2 functional reveals. Even though we did not specify any further constraints, the posterior probability for each $q(x_i)$ leads to a well defined mean, its variance of course depends on the number and quality of the provided input data. In the ideal case of $\nu_{\max} = 100$ we manage to reproduce the PDF for scenario A excellently, for scenario B however ringing artifacts

remain clearly visible. On the other hand $\nu_{\max} = 20$, as expected leads to results that are less accurate than those obtained by the advanced Bayesian reconstruction methods or the neural network.

We thus learn that the additional prior information encoded in the Bayesian reconstruction manages to provide meaningful regularization to the problem beyond positivity. At the same time it also tells us that the neural network approach is able to extract more relevant prior information for a regularization than just positivity, i.e. it goes beyond a simple χ^2 fit.

7.5 Backus Gilbert

The Backus Gilbert method is the fastest of the three methods being tested, particularly when using Jackknife resampling to estimate statistical errors. The $\mathbf{a}$ are calculated, using Eq. (7.37), independently of the values of the data only requiring the Ioffe times associated with the data. Only the final matrix multiplication needs to be performed for each jackknife sample. This feature is unlike the neural networks or Bayesian inference where the entire process must be repeated for each jackknife sample. This first test of the BG method will use the data from the pseudo-ITD, subsequently matched to the ITD, from the dynamical quark ensembles. Unlike the mock data, the actual lattice data has multiple values at a single Ioffe time. It has been checked that performing BG on this set, or a set where points with common Ioffe time are averaged, make no discernible difference.

The preconditioning functions will be chosen by modifying the results of the fits to the functional form in Eq. (2.25). For use as the preconditioning function, the fit values for a will be lowered and b will be raised such that the preconditioned PDF will converge to 0 at $x = 0$ and $x = 1$. BG performs better in regions where the unknown function is smooth and slowly varying. In these tests, only a single preconditioning function will be

used, but for a proper study of the systematic errors in this analysis, several should be tested as before. The values of the preconditioning are list in Tab. 7.1.

ensemble	a	b
$a127m415$	-0.25	2.25
$a127m415L$	-0.25	2.75
$a094m390$	-0.25	2.5

TABLE 7.1: The parameterized used in the preconditioning function, with the functional form in Eq. (2.25), for the BG on each ensemble. The results of the fit to this functional form were modified such that the preconditioned PDF will vanish at $x = 0$ and $x = 1$.

In the following the inversion of the matrix $\mathbf{M}$ will be regulated utilizing the covariance matrix as in Eq. (7.39). The value of ρ is varied in order to check the systematic effects on the choice. Figs 7.35- 7.38 show the results of the BG inversion. The low ρ BG PDFs, typically $\rho < 0.5$, have not completely regulated in the inverse problem and the results have noticeable oscillatory behavior. The reconstructions with $\rho > 1$ show no sign of these oscillations which hint that the inverse problem has been resolved.

As with the mock data, the width σ , shown in Fig 7.44, increases with x . The large x region is where the PDF is flattest and sampling wider distributions should not create large systematic errors, as it would where the PDF changes more rapidly. The covariance regulated BG is designed to trade off between the size of the width σ and the size of the variance of the resulting function. This tradeoff is controlled by the parameter ρ . As can be seen in the results, by increasing ρ , the variance will systematically decrease as the width increases. The low ρ BG PDFs where noticeable oscillatory behavior still occur have a larger variance than the large ρ results. On the other hand, the reconstructions with $\rho > 1$ have larger widths σ but typically show no clear sign of oscillations. The BG method does not require the same ρ to be used for all x . A more sophisticated analysis could use the larger ρ for the low x region where the width is still small and a smaller ρ ,

with its smaller widths, in the large x region where the oscillatory behavior of the result is not strong.

For the region $x > 0.1$, the covariance regulated BG PDFs are consistently much larger than the phenomenological PDFs, except for when the oscillatory behavior is strong. The larger x regions are what control the lowest moments, which at these heavy pion masses are known to be larger than the phenomenological result. Even so, the reconstructed BG ITD consistently overestimates the results in the low Ioffe time region, which is controlled by the sum rule in Eq. (2.24) as well lowest moments. In a future calculation, a modification of the BG procedure could contain another Lagrange multiplier which would force the results to satisfy the sum rule by forcing the reconstructed BG ITD to be unity at $\nu = 0$.

Next shown in Figs. 7.41-7.43 are the Backus Gilbert inversions when a Tikhonov regularization is used for inverting the matrix $\mathbf{M}$ as in Eq. (7.40). The widths for these inversions are shown in Fig. 7.45. In general, the Tikhonov regulated BG PDFs look similar to the covariance regulated PDFs. The reconstructed Tikhonov regulated BG ITDs are still systematically larger than the original data, but the discrepancy is less than the covariance regulated BG ITDs. At large ρ , the results are statistically consistent with the sum rule. For the $a127m415L$ ensemble, the oscillatory behavior persist for $\rho < 0.01$. The widths are the quantitative tool for comparing the BG results with different regulators. The widths can be compared for values of ρ where in both cases the oscillations have clearly ceased. The widths of the Tikhonov regulated result at $\rho = 0.1$ is consistently larger than for the covariance regulated BG at $\rho = 5$.

The Tikhonov regulated BG does not explicitly attempt to tradeoff between the width σ and the variance of the resulting PDF. This feature does, to some extent, still appear to exist in the results. The Tikhonov regulated BG PDF at $\rho = 10^{-4}$ does consistently have larger variance than the result at $\rho = 0.1$. This feature, may be an artifact of the oscillatory behavior which still remains at low ρ , not an inherent property of the Tikhonov

regulated BG.

The BG method has very few constraints explicitly put into it. The results contain well known features of the PDF vanishing as $x \rightarrow 1$ or diverging as $x \rightarrow 0$. However, a judicious choice of the preconditioning function can be used to favor these effects, but they cannot be explicitly forced into the result. With so little specified, the qualitative similarities to the phenomenological PDFs is very encouraging that this method may be useful for extracting PDFs from lattice data. This method also contains a number of choices, all of which must be changed in order to check for potential biases. For example, the consistency of results must be confirmed when changing the particular regulator for inverting the $\mathbf{M}$ matrix, the maximum width σ one allows, and the shape of the preconditioning function.

7.6 Comparison of PDF extractions

Due to the limitations of lattice calculations, the determination of a PDF does not have a unique result. Some choice must be made on how to select a particular solution, and this choice cannot be made with adding a bias into the results. It is important to make sure that the PDFs from different solutions to the inverse problem is in agreement.

Fig 7.46 shows a comparison of the different PDF fits and BG results for each of the ensembles. As can be seen, most of the PDF results are in statistical agreement for a large window in x . The agreement breaks down as the $x \rightarrow 1$ and $x \rightarrow 0$ limits are approached. The $x \rightarrow 0$ limit represents how that particular inverse method attempted to cope with this unconstrained region and a disagreement is expected. The $x \rightarrow 1$ limit, where the PDF value is smallest, is where the PDF is very small. In this region, a slight discrepancy can lead to a large relative disagreement.

Unlike the other PDF results, the covariance regulated BG results systematically diverge from the fit to the JAM functional form in Eq. (7.10) without growing errors. Since

this method explicitly attempts to minimize the resulting PDF's variance, it is possible that its statistical errors are misleading. An analysis of the systematic effects of, amongst other things, choosing a particular preconditioning function must be performed before the errors are considered reasonable.

The fits to functional forms with more parameters, those in Eqs. (7.14) and (7.16), tend to have consistently worse statistical errors than the fits to the functional forms in Eqs. (2.25) and (7.10). The fits to these more complicated functions did not improve the χ^2 as well. Since they do not improve the quality of the fit, it is hard to justify including these extra parameters, even if they do add flexibility to the functional form and thereby alleviate some potential bias.

FIG. 7.2: The moments of the pseudo-PDF compared to phenomenologically determined PDF moments from the NLO global fit CJ15nlo [2], and the NNLO global fits MSTW2008nnlo68cl_nf4 [5] and NNPDF31_nnlo_pch_as_0118_mc_164 [6] all evolved to 4 GeV. The top, middle, and bottom plots are from the ensembles $a_{127m415}$, $a_{127m415L}$ and $a_{094m390}$ respectively. The left and right columns show the first and second moments respectively. Only the lowest two moments have signal for most z . The higher moments only have signal for the largest z where the maximum Ioffe time used allows for identifying more than the leading behavior in the Taylor expansion.

FIG. 7.3: The nucleon valence distribution obtained from the ensemble *a127m415* using the fit form in Eq. (2.25). The upper and lower plots are from including all the data in the fit or only using data originating with $z < 4a$ respectively. The $\chi^2/\text{d.o.f.}$ for the fit with all the data is 2.1(7). The $\chi^2/\text{d.o.f.}$ for the fit with points with $z < 4a$ is 5(1). The uncertainty band is obtained from the fits to the Jackknife samples of the data.

FIG. 7.4: The nucleon valence distribution obtained from the ensemble $a127m415L$ using the fit form in Eq. (2.25). The upper and lower plots are from including all the data in the fit or only using data originating with $z < 4a$ respectively. The $\chi^2/\text{d.o.f.}$ for the fit with all the data is 1.5(5). The $\chi^2/\text{d.o.f.}$ for the fit with points with $z < 4a$ is 1.8(8). The uncertainty band is obtained from the fits to the Jackknife samples of the data.

FIG. 7.5: The nucleon valence distribution obtained from the ensemble $a094m390$ using the fit form in Eq. (2.25). The upper and lower plots are from including all the data in the fit or only using data originating with $z < 4a$ respectively. The $\chi^2/\text{d.o.f.}$ for the fit with all the data is 1.8(5). The $\chi^2/\text{d.o.f.}$ for the fit with points with $z < 4a$ is 3.0(9). The uncertainty band is obtained from the fits to the Jackknife samples of the data.

FIG. 7.6: The nucleon valence distribution obtained from the ensemble $a127m415$ fit to the form used by the JAM collaboration in Eq. (7.10). The upper and lower plots are from including all the data in the fit or only using data originating with $z < 4a$ respectively. The $\chi^2/\text{d.o.f.}$ for the fit with all the data is 2.1(8). The $\chi^2/\text{d.o.f.}$ for the fit with points with $z < 4a$ is 6(1). The uncertainty band is obtained from the fits to the Jackknife samples of the data.

FIG. 7.7: The nucleon valence distribution obtained from the ensemble $a127m415L$ fit to the form used by the JAM collaboration in Eq. (7.10). The upper and lower plots are from including all the data in the fit or only using data originating with $z < 4a$ respectively. The $\chi^2/\text{d.o.f.}$ for the fit with all the data is 1.6(5). The $\chi^2/\text{d.o.f.}$ for the fit with points with $z < 4a$ is 2.2(9). The uncertainty band is obtained from the fits to the Jackknife samples of the data.

FIG. 7.8: The nucleon valence distribution obtained from the ensemble $a094m390$ fit to the form used by the JAM collaboration in Eq. (7.10). The upper and lower plots are from including all the data in the fit or only using data originating with $z < 4a$ respectively. The $\chi^2/\text{d.o.f.}$ for the fit with all the data is 1.9(5). The $\chi^2/\text{d.o.f.}$ for the fit with points with $z < 4a$ is 3.4(1.2). The uncertainty band is obtained from the fits to the Jackknife samples of the data.

FIG. 7.9: The nucleon valence distribution obtained from the ensemble $a127m415$ fit to the form using the polynomial in Eq. (7.14). The upper and lower plots are from including all the data in the fit or only using data originating with $z < 4a$ respectively. The $\chi^2/\text{d.o.f.}$ for the fit with all the data is 2.3(9). The $\chi^2/\text{d.o.f.}$ for the fit with points with $z < 4a$ is 7(2). The uncertainty band is obtained from the fits to the Jackknife samples of the data.

FIG. 7.10: The nucleon valence distribution obtained from the ensemble $a127m415L$ fit to the form using the polynomial in Eq. (7.14). The upper and lower plots are from including all the data in the fit or only using data originating with $z < 4a$ respectively. The $\chi^2/\text{d.o.f.}$ for the fit with all the data is 1.7(6). The $\chi^2/\text{d.o.f.}$ for the fit with points with $z < 4a$ is 2.7(9). The uncertainty band is obtained from the fits to the Jackknife samples of the data.

FIG. 7.11: The nucleon valence distribution obtained from the ensemble $a094m390$ fit to the form using the polynomial in Eq. (7.14). The upper and lower plots are from including all the data in the fit or only using data originating with $z < 4a$ respectively. The $\chi^2/\text{d.o.f.}$ for the fit with all the data is 2.0(1.2). The $\chi^2/\text{d.o.f.}$ for the fit with points with $z < 4a$ is 4.2(1.3). The uncertainty band is obtained from the fits to the Jackknife samples of the data.

FIG. 7.12: The nucleon valence distribution obtained from the ensemble $a127m415$ fit to the form using the polynomial in Eq. (7.16). The upper and lower plots are from including all the data in the fit or only using data originating with $z < 4a$ respectively. The $\chi^2/\text{d.o.f.}$ for the fit with all the data is 2.3(9). The $\chi^2/\text{d.o.f.}$ for the fit with points with $z < 4a$ is 7(2). The uncertainty band is obtained from the fits to the Jackknife samples of the data.

FIG. 7.13: The nucleon valence distribution obtained from the ensemble $a127m415L$ fit to the form using the polynomial in Eq. (7.16). The upper and lower plots are from including all the data in the fit or only using data originating with $z < 4a$ respectively. The $\chi^2/\text{d.o.f.}$ for the fit with all the data is 1.4(5). The $\chi^2/\text{d.o.f.}$ for the fit with points with $z < 4a$ is 3(1). The uncertainty band is obtained from the fits to the Jackknife samples of the data.

FIG. 7.14: The nucleon valence distribution obtained from the ensemble $a094m390$ fit to the form using the polynomial in Eq. (7.16). The upper and lower plots are from including all the data in the fit or only using data originating with $z < 4a$ respectively. The $\chi^2/\text{d.o.f.}$ for the fit with all the data is 4.9(8). The $\chi^2/\text{d.o.f.}$ for the fit with points with $z < 4a$ is 7.5(7.0). The uncertainty band is obtained from the fits to the Jackknife samples of the data.

FIG. 7.15: The pseudo-ITSF of the axial vector-vector current combination fit to the PDF functional form in Eq. (7.18) using the leading order kernel in Eq. (4.48). The uncertainty band is obtained from the fits to the Jackknife samples of the data.

FIG. 7.16: The pion valence distribution obtained using the LO perturbative kernel in Eq. (4.48) derived in Sec. 4.5.1 and the functional form of the PDF in Eq. (7.18). Figure (7.16a) shows the pion valence distribution $q_v^\pi(x)$ and Figure (7.16b) shows the $xq_v^\pi(x)$ -distribution. The uncertainty band is obtained from the fits to the Jackknife samples of the data.

FIG. 7.17: Comparison of pion $xq_v^\pi(x)$ -distribution with the leading-order (LO) extraction from Drell-Yan data [7] (gray data points with uncertainties), next-to-leading order (NLO) fits [8, 9, 10] (orange band, magenta curve, and red band), and model calculations [11, 12] (black and blue lines). This lattice QCD calculation of $q_v^\pi(x)$ is evolved from an initial scale $\mu_0^2 = 1 \text{ GeV}^2$ at LO. All the results are at evolved to an evolution scale of $\mu^2 = 27 \text{ GeV}^2$.

FIG. 7.18: The two mock datasets used for testing the methods. The PDF on the left. The Ioffe time PDF on the right. Scenario A (red band) will test the case of a PDF which diverges at small x and scenario B (blue band) will test the case of a PDF which converges at small x .

FIG. 7.19: Best fit PDF (red solid line) from (left) realistic PDF data [mock scenario A] and (right) a modified scenario with the PDF vanishing at the origin [mock scenario B]. The actual mock PDF in the former cases is given as gray solid line. To determine the dependence of our results on the choice of default model, three further choices for m are plotted, two arising from varying the best fit parameters by factors of 2, one being the constant default model $m=1$.

FIG. 7.20: x -space PDF's reconstructed using the Backus-Gilbert (BG) method from $N_\nu = 12$ Ioffe-time data points on the interval $\nu = [0, 10]$ (top) as well as the input data (gray crosses) compared to the data arising from the reconstructed PDF (red crosses) in the bottom panels. The plots in the left column show the results for mock data based on a phenomenological PDF, while the right column from the modified scenario where the PDF vanishes at the origin. In both scenarios no preconditioning has been employed.

FIG. 7.21: x -space PDF's reconstructed using the Backus-Gilbert (BG) method from $N_\nu = 12$ Ioffe-time data points on the interval $\nu = [0, 10]$ (top) as well as the input data (gray crosses) compared to the data arising from the reconstructed PDF (red crosses) in the bottom panels. The plots in the left column show the results for mock data based on a phenomenological PDF, while the right column from the modified scenario where the PDF vanishes at the origin. The reconstruction was performed with preconditioning exponents $a = -0.35$ and $b = 2$ for scenario A and $a = 0.3$ and $b = 2$ for scenario B.

FIG. 7.22: x -space PDF's reconstructed using the Backus-Gilbert (BG) method from $N_\nu = 23$ Ioffe-time data points on the interval $\nu = [0, 20]$ (top) and from $N_\nu = 112$ Ioffe-time data points on the interval $\nu = [0, 100]$ (bottom). The plots in the left column show the results for mock data based on a phenomenological PDF, while the right column from the modified scenario where the PDF vanishes at the origin. The reconstruction was performed with preconditioning exponents $a = -0.35$ and $b = 2$ for scenario A and $a = 0.3$ and $b = 2$ for scenario B. It is noteworthy that by having lattice data up to $\nu_{\max} = 20$ one can achieve a very good reconstruction with the BG method.

FIG. 7.23: A neural network can be used as a general parametrization of an unknown function from $\mathbb{R}^{N_1} \rightarrow \mathbb{R}^{N_L}$. For the case of a PDF, a single valued function of a single argument, the input and output layers have only one neuron.

FIG. 7.24: The genetically trained neural nets. The blue band is the original data. The curves are the responses for the final population of genetically trained neural networks. The left column is with NNPDF data. The right column is with modified data. The first row has a network geometry of 1-3-1. The second row has a geometry of 1-4-1. The third row has a geometry of 1-2-2-1.

FIG. 7.25: The genetically trained neural nets. The blue band is the original data. The red band is the reconstructed PDF with statistical and systematic errors. The left column is with NNPDF data. The right column is with modified data. The first row has a network geometry of 1-3-1. The second row has a geometry of 1-4-1. The third row has a geometry of 1-2-2-1.

FIG. 7.26: The genetically trained neural nets. The grey points are the original data while the red points are the ones reconstructed by the NN. The left column is with NPDF data. The right column is with modified data. The first row has a network geometry of 1-3-1. The second row has a geometry of 1-4-1. The third row has a geometry of 1-2-2-1.

FIG. 7.27: The PDF reconstructions from preconditioned Neural Networks (top). The grey points are the original data while the red points are the ones reconstructed by the NN (bottom). The left is scenario A and the right is scenario B.

FIG. 7.28: x -space PDF's reconstructed using the Maximum Entropy method (MEM) from $N_\nu = 12$ Ioffe-time data points on the interval $\nu = [0, 10]$ (top) as well as the input data (gray crosses) compared to the data arising from the reconstructed PDF (red crosses) in the bottom panels. The plots in the left column denote the results for mock data based on a phenomenological PDF, while the right column from the modified scenario where the PDF vanishes at the origin. (top) The original mock PDF is given as gray dashed line and the result based on the BR method with the best fit default model as solid red line. Statistical uncertainty is visualized by a red, systematic uncertainty from the choice of default model by a blue errorband respectively. (bottom) The red solid line corresponds to the relative deviation of the reconstruction from the input mock PDF, the red errorband encodes the combined uncertainty of the BR reconstruction, the gray errorband the uncertainty of the original mock PDF.

FIG. 7.29: x -space PDF's reconstructed using the Bayesian reconstruction (BR) method from $N_\nu = 12$ Ioffe-time data points on the interval $\nu = [0, 10]$ (top) as well as the input data (gray crosses) compared to the data arising from the reconstructed PDF (red crosses) in the bottom panels.

FIG. 7.30: x -space PDF's reconstructed using the Maximum Entropy Method (top) as well as the Bayesian reconstruction (BR) method (bottom) from $N_\nu = 12$ Ioffe-time data points on the interval $\nu = [0, 20]$.

FIG. 7.31: x -space PDF's reconstructed using a quadratic prior Bayesian (QDR) method from $N_v = 12$ Ioffe-time data points on the interval $\nu = [0, 10]$ (top) as well as the input data (gray crosses) compared to the data arising from the reconstructed PDF (red crosses) in the bottom panels.

FIG. 7.32: x -space PDF's reconstructed using the generalized Bayesian reconstruction (BRg) method from $N_\nu = 12$ Ioffe-time data points on the interval $\nu = [0, 10]$ (top) as well as the input data (gray crosses) compared to the data arising from the reconstructed PDF (red crosses) in the bottom panels.

FIG. 7.33: Reconstructed PDF (red dashed) from restricted sampling of the χ^2 functional based on $N_\nu = 12$ Ioffe-time data points on the interval $\nu = [0, 10]$. Mock scenario A is shown on the left, scenario B on the right. While positivity alone already provides a powerful regularization, for this realistic scenario we do observe deviations from the correct result, especially for scenario B.

FIG. 7.34: Reconstructed PDF (red dashed) from restricted sampling of the χ^2 functional based on $N_\nu = 100$ Ioffe-time data points on the interval $\nu = [0, 100]$. Mock scenario A is shown on the left, scenario B on the right. In this almost ideal setting, positivity alone provides a regularization that is sufficient for a reasonable determination of the PDF of scenario A, while it is unable to suppress ringing artifacts in scenario B.

FIG. 7.35: Reconstructed PDF from covariance regulated Backus Gilbert for ensemble $a_{127m415}$.

FIG. 7.36: Reconstructed PDF from covariance regulated Backus Gilbert for ensemble $a_{127m415}$.

FIG. 7.37: Reconstructed PDF from covariance regulated Backus Gilbert for ensemble $a127m415L$.

FIG. 7.38: Reconstructed PDF from covariance regulated Backus Gilbert for ensemble $a_{127m415L}$.

FIG. 7.39: Reconstructed PDF from covariance regulated Backus Gilbert for ensemble $a094m390$.

FIG. 7.40: Reconstructed PDF from covariance regulated Backus Gilbert for ensemble $a094m390$.

FIG. 7.41: Reconstructed PDF from Tikhonov regulated Backus Gilbert for ensemble $a127m415$.

FIG. 7.42: Reconstructed PDF from Tikhonov regulated Backus Gilbert for ensemble $a_{127m415L}$.

FIG. 7.43: Reconstructed PDF from Tikhonov regulated Backus Gilbert for ensemble $a094m390$.

FIG. 7.44: Widths σ from covariance regulated Backus Gilbert inversions. The top left and top right plots are from ensembles $a_{127m415}$ and $a_{127m415L}$ respectively and the lower plot is from ensemble $a_{094m390}$.

FIG. 7.45: Widths σ from Tikhonov regulated Backus Gilbert inversions. The top left and top right plots are from ensembles $a127m415$ and $a127m415L$ respectively and the lower plot is from ensemble $a094m390$.

FIG. 7.46: Comparisons of the PDF extractions. The top, middle and bottom plots are from ensembles $a127m415$, $a127m415L$, and $a094m390$ respectively. The plots on the right column have been normalized by taking the central value from the fit to the functional form in Eq. (7.10).

CHAPTER 8

Summary and Outlook

In this work, the first calculation of “Good Lattice Cross Sections” have been performed. The leading twist matching kernels between the LCSs and the PDF, which had been previously determined using collinear factorization, are applied to the lattice QCD data. The methods for extracting the PDFs from this matched lattice data have been implemented and studied. This work leads the way for the computation of parton distributions directly from QCD. Possibly in the future, the phenomenological extractions from experimental data can be augmented by including results calculated from LCSs.

On a Euclidean lattice, the light cone matrix elements which define the PDFs are inaccessible. All possible separations between fields must be space-like. A certain class of operators with space-like separated fields are identified which have a similar OPE to the PDF. The matching kernels for the pseudo-ITD to the $\overline{\text{MS}}$ ITD at NLO and the matching kernels for a pseudo-ITSF to the $\overline{\text{MS}}$ PDF at LO were applied to the results of a Lattice QCD calculation. Using these kernels, one can see that the real and imaginary components of the pseudo-ITD will give the quark valence distribution $q_v(x)$ and the combination $q(x) + \bar{q}(x)$ respectively. For the two current matrix elements, most of the

ones considered would only give access to the $q(x) + \bar{q}(x)$. The antisymmetric combination of the axial vector and vector currents was shown to give access to only the quark valence distribution.

The matrix elements for the nucleon and pion pseudo-ITD were calculated on several ensembles of 2+1 flavors of clover improved Wilson fermions using the novel Feynman-Hellmann summation technique. This method allowed for excellent control of excited state effects. The pseudo-ITD was renormalized and had higher twist contaminations reduced by forming the ratio in Eq. (5.9). The reduced pseudo-ITD showed no signs of large higher twist effects even for long separations $z \sim \Lambda_{\text{QCD}}^{-1}$. Instead the reduced pseudo-ITD appears independent of the scale z^2 . A future calculation with finer lattice spacings would allow for several points at a single Ioffe time with large and small separations to study the z^2 dependence between the perturbative and non-perturbative regimes.

The pion pseudo-ITSF for the combination of the axial vector and vector currents was calculated on an ensemble with 2+1 flavors of clover improved Wilson fermions with heavier than physical pion mass. Alongside this particular choice, the correlation functions of other current pairs were calculated and will be presented in a future work. These two current matrix elements do lack in the statistical precision that the pseudo-ITD has, partly due to the lack of volume averaging when forming the correlation function. The modified sequential source technique used also is incapable of calculating nucleon matrix elements. A calculation using distillation can be performed to overcome this difficulty, and we plan to pursue this avenue in the future.

Several methods were studied for extracting the PDF from these matrix elements. In the first method the data was fit to functional forms similar to those used by phenomenological PDF fits. The results are qualitatively similar to the phenomenological results even without many systematic errors under control. The PDF typical is overestimated at large x due to the heavy quark masses used in this work. In a future publication, results with

near physical pion mass will be presented.

More advanced methods were studied using mock data generated from the a phenomenological extraction performed by the NNPDF collaboration. The effects of having discretized data on finite ranges of Ioffe time were studied as well as the effects of preconditioning the ill-posed inverse problem. Even with short ranges of Ioffe time, $\nu_{\max} = 10$, the mock data PDF was accurately reconstructed for a wide range of $x > 0.2$. The Backus Gilbert and neural network method have been applied to the lattice calculated matrix elements with results comparable to the fits.

This work focused on the calculation of the quark valence PDFs only. For the nucleon, this quantity is known to high precision from phenomenological fits which makes it an ideal testing ground for lattice methods. The success of this calculation motivates the calculation of parton distributions with less precise phenomenological determinations, such as polarized PDFs, meson Distribution Amplitudes, and gluon PDFs. Lattice results with more systematic errors under control can also be used to compliment the experimental datasets for phenomenological fits. A combined analysis of lattice and experimental results could lead to higher precision PDFs than either individually could achieve.

BIBLIOGRAPHY

- [1] R. Edwards, B. Joó, K. Orginos, D. Richards, and F. Winter, U.S. 2+1 flavor clover lattice generation program (2016), `unpublished`.
- [2] A. Accardi, L. T. Brady, W. Melnitchouk, J. F. Owens, and N. Sato, *Phys. Rev. D* **93**, 114017 (2016), `1602.03154`.
- [3] C. Alexandrou, K. Cichy, M. Constantinou, K. Jansen, A. Scapellato, and F. Steffens, *Phys. Rev. Lett.* **121**, 112001 (2018), `1803.02685`.
- [4] J.-W. Chen, L. Jin, H.-W. Lin, Y.-S. Liu, Y.-B. Yang, J.-H. Zhang, and Y. Zhao (2018), `1803.04393`.
- [5] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt, *Eur. Phys. J. C* **63**, 189 (2009), `0901.0002`.
- [6] R. D. Ball et al. (NNPDF), *Eur. Phys. J. C* **77**, 663 (2017), `1706.00428`.
- [7] J. S. Conway, C. E. Adolphsen, J. P. Alexander, K. J. Anderson, J. G. Heinrich, J. E. Pilcher, A. Possoz, E. I. Rosenberg, C. Biino, J. F. Greenhalgh, et al., *Phys. Rev. D* **39**, 92 (1989).
- [8] K. Wijesooriya, P. E. Reimer, and R. J. Holt, *Phys. Rev. C* **72**, 065203 (2005).
- [9] M. Aicher, A. Schafer, and W. Vogelsang, *Phys. Rev. Lett.* **105**, 252003 (2010), `1009.2481`.

- [10] P. C. Barry, N. Sato, W. Melnitchouk, and C.-R. Ji (Jefferson Lab Angular Momentum (JAM) Collaboration), Phys. Rev. Lett. **121**, 152001 (2018).
- [11] G. F. de Teramond, T. Liu, R. S. Sufian, H. G. Dosch, S. J. Brodsky, and A. Deur (HLFHS), Phys. Rev. Lett. **120**, 182001 (2018), 1801.09154.
- [12] C. Chen, L. Chang, C. D. Roberts, S. Wan, and H.-S. Zong, Phys. Rev. **D93**, 074021 (2016), 1602.01502.
- [13] K. G. Wilson, Phys. Rev. D **10**, 2445 (1974), URL <https://link.aps.org/doi/10.1103/PhysRevD.10.2445>.
- [14] A. V. Belitsky, X.-d. Ji, and F. Yuan, Phys. Rev. **D69**, 074014 (2004), [hep-ph/0307383](https://arxiv.org/abs/hep-ph/0307383).
- [15] J. D. Bjorken, Phys. Rev. **179**, 1547 (1969), URL <https://link.aps.org/doi/10.1103/PhysRev.179.1547>.
- [16] R. P. Feynman, Conf. Proc. **C690905**, 237 (1969).
- [17] A. V. Manohar, P. Nason, G. P. Salam, and G. Zanderighi, JHEP **12**, 046 (2017), 1708.01256.
- [18] V. Braun, P. Gornicki, and L. Mankiewicz, Phys. Rev. **D51**, 6036 (1995), [hep-ph/9410318](https://arxiv.org/abs/hep-ph/9410318).
- [19] V. N. Gribov and L. N. Lipatov, Sov. J. Nucl. Phys. **15**, 438 (1972), [Yad. Fiz.15,781(1972)].
- [20] G. Altarelli and G. Parisi, Nucl. Phys. **B126**, 298 (1977).
- [21] Y. L. Dokshitzer, Sov. Phys. JETP **46**, 641 (1977), [Zh. Eksp. Teor. Fiz.73,1216(1977)].

- [22] A. Vogt, S. Moch, and J. A. M. Vermaseren, Nucl. Phys. **B691**, 129 (2004), [hep-ph/0404111](#).
- [23] S. Moch, J. A. M. Vermaseren, and A. Vogt, Nucl. Phys. **B688**, 101 (2004), [hep-ph/0403192](#).
- [24] K. G. Wilson, Phys. Rev. D **10**, 2445 (1974).
- [25] A. Zichichi, ed., *New Phenomena in Subnuclear Physics*, vol. 13, Plenum (Plenum, New York, USA, 1977), ISBN 9781461342106, 9781461342083.
- [26] K. Symanzik, Nucl. Phys. **B226**, 187 (1983).
- [27] K. Symanzik, Nucl. Phys. **B226**, 205 (1983).
- [28] B. Sheikholeslami and R. Wohlert, Nucl. Phys. **B259**, 572 (1985).
- [29] G. S. Bali, B. Lang, B. U. Musch, and A. Schäfer, Phys. Rev. **D93**, 094515 (2016), [1602.05525](#).
- [30] C. Bouchard, C. C. Chang, T. Kurth, K. Orginos, and A. Walker-Loud, Phys. Rev. **D96**, 014504 (2017), [1612.06963](#).
- [31] K. Orginos, A. Radyushkin, J. Karpie, and S. Zafeiropoulos, Phys. Rev. **D96**, 094503 (2017), [1706.05373](#).
- [32] C. C. Chang et al., Nature **558**, 91 (2018), [1805.12130](#).
- [33] S. Capitani and G. Rossi, Nucl. Phys. **B433**, 351 (1995), [hep-lat/9401014](#).
- [34] G. Beccarini, M. Bianchi, S. Capitani, and G. Rossi, Nucl. Phys. **B456**, 271 (1995), [hep-lat/9506021](#).
- [35] C. Monahan and K. Orginos, Phys. Rev. **D91**, 074513 (2015), [1501.05348](#).

- [36] W. Detmold, W. Melnitchouk, and A. W. Thomas, Eur. Phys. J.direct **3**, 13 (2001), hep-lat/0108002.
- [37] X. Ji, Phys. Rev. Lett. **110**, 262002 (2013), 1305.1539.
- [38] R. A. Briceño, M. T. Hansen, and C. J. Monahan, Phys. Rev. **D96**, 014502 (2017), 1703.06072.
- [39] X. Xiong, X. Ji, J.-H. Zhang, and Y. Zhao, Phys. Rev. **D90**, 014051 (2014), 1310.7471.
- [40] C. Alexandrou, K. Cichy, M. Constantinou, K. Hadjiyiannakou, K. Jansen, H. Panagopoulos, and F. Steffens, Nucl. Phys. **B923**, 394 (2017), 1706.00265.
- [41] I. W. Stewart and Y. Zhao, Phys. Rev. **D97**, 054512 (2018), 1709.04933.
- [42] T. Izubuchi, X. Ji, L. Jin, I. W. Stewart, and Y. Zhao, Phys. Rev. **D98**, 056004 (2018), 1801.03917.
- [43] W. Wang, J.-H. Zhang, S. Zhao, and R. Zhu (2019), 1904.00978.
- [44] A. Radyushkin, Phys. Lett. **B767**, 314 (2017), 1612.05170.
- [45] A. V. Radyushkin, Phys. Rev. **D96**, 034025 (2017), 1705.01488.
- [46] A. V. Radyushkin, Phys. Lett. **B788**, 380 (2019), 1807.07509.
- [47] Y.-Q. Ma and J.-W. Qiu, Phys. Rev. **D98**, 074021 (2018), 1404.6860.
- [48] Y.-Q. Ma and J.-W. Qiu, Phys. Rev. Lett. **120**, 022003 (2018), 1709.03018.
- [49] A. J. Chambers, R. Horsley, Y. Nakamura, H. Perlt, P. E. L. Rakow, G. Schierholz, A. Schiller, K. Somfleth, R. D. Young, and J. M. Zanotti, Phys. Rev. Lett. **118**, 242001 (2017), 1703.01153.

- [50] K. Orginos, A. Radyushkin, J. Karpie, and S. Zafeiropoulos, Phys. Rev. **D96**, 094503 (2017), 1706.05373.
- [51] R. S. Sufian, J. Karpie, C. Egerer, K. Orginos, J.-W. Qiu, and D. G. Richards (2019), 1901.03921.
- [52] K.-F. Liu and S.-J. Dong, Phys. Rev. Lett. **72**, 1790 (1994), hep-ph/9306299.
- [53] G. Martinelli, Nucl. Phys. Proc. Suppl. **73**, 58 (1999), hep-lat/9810013.
- [54] A. V. Radyushkin, Phys. Lett. **B781**, 433 (2018), 1710.08813.
- [55] V. S. Dotsenko and S. N. Vergeles, Nucl. Phys. **B169**, 527 (1980).
- [56] R. A. Brandt, F. Neri, and M.-a. Sato, Phys. Rev. **D24**, 879 (1981).
- [57] T. Ishikawa, Y.-Q. Ma, J.-W. Qiu, and S. Yoshida, Phys. Rev. **D96**, 094019 (2017), 1707.03107.
- [58] A. Radyushkin, Phys. Rev. **D98**, 014019 (2018), 1801.02427.
- [59] J.-H. Zhang, J.-W. Chen, and C. Monahan, Phys. Rev. **D97**, 074508 (2018), 1801.03023.
- [60] J. Karpie, K. Orginos, and S. Zafeiropoulos, JHEP **11**, 178 (2018), 1807.10933.
- [61] J. Liang, K.-F. Liu, and Y.-B. Yang, EPJ Web Conf. **175**, 14014 (2018), 1710.11145.
- [62] S. Duane, A. Kennedy, B. J. Pendleton, and D. Roweth, Physics Letters B **195**, 216 (1987), ISSN 0370-2693, URL <http://www.sciencedirect.com/science/article/pii/037026938791197X>.
- [63] S. Borsanyi et al., JHEP **09**, 010 (2012), 1203.4469.

- [64] M. Constantinou and H. Panagopoulos, Phys. Rev. **D96**, 054506 (2017), 1705.11193.
- [65] M. Constantinou and H. Panagopoulos, EPJ Web Conf. **175**, 06025 (2018), 1711.00543.
- [66] G. Golub and V. Pereyra, SIAM Journal on Numerical Analysis **10**, 413 (1973), <https://doi.org/10.1137/0710036>, URL <https://doi.org/10.1137/0710036>.
- [67] C. Alexandrou, K. Cichy, M. Constantinou, K. Hadjiyiannakou, K. Jansen, A. Scapellato, and F. Steffens (2019), 1902.00587.
- [68] B. U. Musch, P. Hagler, J. W. Negele, and A. Schafer, Phys. Rev. **D83**, 094507 (2011), 1011.1213.
- [69] A. Buckley, J. Ferrando, S. Lloyd, K. Nordström, B. Page, M. Rüfenacht, M. Schönherr, and G. Watt, Eur. Phys. J. **C75**, 132 (2015), 1412.7420.
- [70] K. Gottfried, Phys. Rev. Lett. **18**, 1174 (1967), URL <https://link.aps.org/doi/10.1103/PhysRevLett.18.1174>.
- [71] W. Detmold and C. J. D. Lin, Phys. Rev. **D73**, 014501 (2006), hep-lat/0507007.
- [72] J. Karpie, K. Orginos, A. Rothkopf, and S. Zafeiropoulos, JHEP **04**, 057 (2019), 1901.05408.
- [73] H.-W. Lin, J.-W. Chen, T. Ishikawa, and J.-H. Zhang (LP3), Phys. Rev. **D98**, 054504 (2018), 1708.05301.
- [74] C. Alexandrou, EPJ Web Conf. **137**, 01004 (2017), 1612.04644.
- [75] L. A. Harland-Lang, A. D. Martin, P. Motylinski, and R. S. Thorne, Eur. Phys. J. **C75**, 204 (2015), 1412.3989.

- [76] J. Gao, M. Guzzi, J. Huston, H.-L. Lai, Z. Li, P. Nadolsky, J. Pumplin, D. Stump, and C. P. Yuan, Phys. Rev. **D89**, 033009 (2014), 1302.6246.
- [77] S. Dulat, T.-J. Hou, J. Gao, M. Guzzi, J. Huston, P. Nadolsky, J. Pumplin, C. Schmidt, D. Stump, and C. P. Yuan, Phys. Rev. **D93**, 033006 (2016), 1506.07443.
- [78] R. Brun and F. Rademakers, in *AIHENP'96 Workshop, Lausanne* (1996), vol. 389, pp. 81–86.
- [79] G. S. Bali, V. M. Braun, B. Gläsel, M. Göckeler, M. Gruber, F. Hutzler, P. Korcyl, A. Schäfer, P. Wein, and J.-H. Zhang, Phys. Rev. **D98**, 094507 (2018), 1807.06671.
- [80] R. A. Briceño, J. V. Guerrero, M. T. Hansen, and C. J. Monahan, Phys. Rev. **D98**, 014511 (2018), 1805.01034.
- [81] G. R. Farrar and D. R. Jackson, Phys. Rev. Lett. **43**, 246 (1979).
- [82] E. L. Berger and S. J. Brodsky, Phys. Rev. Lett. **42**, 940 (1979).
- [83] S. J. Brodsky, M. Burkardt, and I. Schmidt, Nuclear Physics B **441**, 197 (1995), ISSN 0550-3213.
- [84] J.-W. Chen, L. Jin, H.-W. Lin, Y.-S. Liu, A. Schäfer, Y.-B. Yang, J.-H. Zhang, and Y. Zhao (2018), 1804.01483.
- [85] N. Karthik, T. Izubichi, L. Jin, C. Kallidonis, S. Mukherjee, P. Petreczky, C. Shugert, and S. Syritsyn, PoS **LATTICE2018**, 109 (2018), 1811.06075.
- [86] T. Izubuchi, L. Jin, C. Kallidonis, N. Karthik, S. Mukherjee, P. Petreczky, C. Shugert, and S. Syritsyn (2019), 1905.06349.
- [87] G. Backus and F. Gilbert, Geophysical Journal International **16**, 169 (1968).

- [88] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery (1992).
- [89] B. B. Brandt, A. Francis, H. B. Meyer, and D. Robaina, Phys. Rev. **D92**, 094510 (2015), 1506.05732.
- [90] R.-A. Tripolt, P. Gubler, M. Ulybyshev, and L. Von Smekal (2018), 1801.10348.
- [91] M. V. Ulybyshev, C. Winterowd, and S. Zafeiropoulos, EPJ Web Conf. **175**, 03008 (2018), 1710.06675.
- [92] M. Ulybyshev, C. Winterowd, and S. Zafeiropoulos, Phys. Rev. **B96**, 205115 (2017), 1707.04212.
- [93] S. Forte, L. Garrido, J. I. Latorre, and A. Piccione, JHEP **05**, 062 (2002), hep-ph/0204232.
- [94] R. D. Ball et al. (NNPDF), JHEP **04**, 040 (2015), 1410.8849.
- [95] R. D. Ball, L. Del Debbio, S. Forte, A. Guffanti, J. I. Latorre, J. Rojo, and M. Ubiali, Nucl. Phys. **B838**, 136 (2010), 1002.4407.
- [96] R. D. Ball, V. Bertone, F. Cerutti, L. Del Debbio, S. Forte, A. Guffanti, J. I. Latorre, J. Rojo, and M. Ubiali (NNPDF), Nucl. Phys. **B849**, 112 (2011), [Erratum: Nucl. Phys.B855,927(2012)], 1012.0836.
- [97] J. Rojo, in *13th Conference on Quark Confinement and the Hadron Spectrum (Confinement XIII) Maynooth, Ireland, July 31-August 6, 2018* (2018), 1809.04392.
- [98] J. Skilling and S. F. Gull, *Bayesian maximum entropy image reconstruction* (Institute of Mathematical Statistics, Hayward, CA, 1991), vol. Volume 20 of *Lecture Notes–Monograph Series*, pp. 341–367, URL <https://doi.org/10.1214/lnms/1215460511>.

- [99] M. Asakawa, T. Hatsuda, and Y. Nakahara, Prog. Part. Nucl. Phys. **46**, 459 (2001), [hep-lat/0011040](#).
- [100] A. Rothkopf, J. Comput. Phys. **238**, 106 (2013), [1110.6285](#).
- [101] Y. Burnier and A. Rothkopf, Phys. Rev. Lett. **111**, 182003 (2013), [1307.6106](#).
- [102] B. Carpenter, A. Gelman, M. D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, and A. Riddell, Journal of Statistical Software **76** (2017).