

THE UNIVERSITY
of ADELAIDE

Novel Techniques for Global Analysis of Parton Distribution Functions

by

Nicholas Hunt-Smith

In fulfilment of the requirements for the degree of

Doctor of Philosophy

February 2025

Department of Physics

School of Physics, Chemistry and Earth Sciences

The University of Adelaide

Contents

Declaration xvii

Acknowledgements xix

Abstract xxi

Introduction 1

I The Standard Model 3

1 QCD in action 5
- 1.1 Standard Model Overview . . . 5
- 1.2 QCD Formalism and Factorization . . . 8
- 1.3 PDF Parametrization . . . 9
- 1.4 Deep Inelastic Scattering . . . 11

2 Global Analysis of PDFs 15
- 2.1 Bayesian inference . . . 16
- 2.2 Markov chain Monte Carlo . . . 18
- 2.3 Hybrid Markov chain Monte Carlo . . . 19
- 2.4 Nested sampling . . . 20
- 2.5 The Hessian method . . . 22
- 2.6 Data resampling . . . 24

3 On the determination of uncertainties in parton densities 27
- 3.1 Comparing methodologies . . . 28
  - 3.1.1 Construction of toy data . . . 28
  - 3.1.2 Fit results . . . 28
- 3.2 Neural networks, partitioning and cross validation . . . 32
  - 3.2.1 Neural networks with cross validation . . . 32
  - 3.2.2 Comparing NN methods to parametric methods . . . 37

3.3 Conclusions and outlook 39

4 High-x Polarized DIS Analysis 43
4.1 Observables 44
4.1.1 Polarized inclusive DIS 45
4.1.2 Polarized semi-inclusive DIS 49
4.1.3 W production in polarized hadron collisions 51
4.1.4 Jet production in polarized hadron collisions 52
4.1.5 Lattice QCD data 53
4.2 Methodology 54
4.3 Global QCD Analysis 56
4.3.1 Quality and stability of the fit 56
4.3.2 Helicity-dependent parton distributions 60
4.3.3 Resolving the sign of Δ_g 63
4.3.4 Higher Twist Effects 68
4.4 Conclusion 70

5 Accelerating Markov Chain Monte Carlo sampling with diffusion models 73
5.1 A brief introduction to diffusion models 75
5.2 A diffusion-model assisted Markov Chain Monte Carlo sampler 76
5.3 Analytic test functions 79
5.3.1 2D Himmelblau function 80
5.3.2 10D Gaussian mixture 80
5.3.3 4D EggBox function 83
5.4 Physics example: global parton distribution function fit 84
5.5 Conclusions 86

II Beyond the Standard Model 89

6 Global QCD analysis and dark bosons 91
6.1 Dark photons 92
6.1.1 Dark photon background 93
6.1.2 Methodology 95
6.1.3 Global QCD analysis results 96
6.2 Z' bosons 102
6.3 Outlook 104

7 A comparison of Bayesian sampling algorithms for particle astrophysics applications 107
7.1 Sampling algorithms . . . 108
7.2 Methodology . . . 111
7.2.1 Evaluating Nested Sampling implementations . . . 111
7.2.2 Diagnostics & Metrics . . . 112
7.3 Results . . . 114
7.3.1 Eggbox . . . 114
7.3.2 Cosmology — Planck . . . 121
7.3.3 Particle physics — LHC . . . 126
7.4 Conclusions . . . 128

Summary and Outlook 131

A Target mass corrections in collinear factorization 133

B Data to theory comparisons 137

C Accelerating Markov Chain Monte Carlo sampling with diffusion models 157
C.1 Hyperparameters . . . 157
C.2 4D Rosenbrock function . . . 159

D A comparison of Bayesian sampling algorithms for particle astrophysics applications 161
D.1 Effective sample size . . . 161
D.1.1 MCMC . . . 161
D.1.2 Nested sampling . . . 162
D.2 Posterior mean and standard deviation . . . 163
D.2.1 Eggbox . . . 163
D.2.2 Rosenbrock . . . 163
D.2.3 Rastrigin . . . 164
D.2.4 Spike-slab function . . . 164
D.2.5 Cosmology-Planck . . . 164
D.2.6 Particle physics - LHC . . . 165
D.3 MCMC: Number of walkers . . . 169

List of Tables

4.1 Dependence of the quality of the global fit (including lattice QCD data) on the minimum value of W^2 used for the inclusive DIS data. The quoted reduced χ^2 values and number of data points $N_{\rm dat}$ for each cut refer only to the inclusive polarized DIS data. 58

4.2 Summary of $\chi^2_{\rm red}$ values and Z-scores for all reactions included in the global fit and the number of data points $N_{\rm dat}$ fitted. The asterisks on the unpolarized and SIA data indicate that these replicas were obtained from the earlier JAM analysis in Ref. [54].The daggers in the $N_{\rm dat}$ column indicate that those datasets have fewer points for the $W^2 > 10$ GeV^2 cut - specifically, $N_{\rm dat} = 365$ for Polarized DIS, $N_{\rm dat} = 700$ for total Polarized data, and $N_{\rm dat} = 7215$ for total of all data. 60

4.3 Truncated and full moments of PDFs at $Q^2 = 10$ GeV^2. The asterisk * indicates the value is fixed in the fit. 63

4.4 Comparison of the $\chi^2_{\rm red}$ values for $\Delta g > 0$ and $\Delta g < 0$ solutions across various datasets. *The "baseline" and "+LQCD" cases have 365 data points in the polarized inclusive DIS dataset, compared with 1735 for the "+ high-x DIS" set. 66

6.1 Comparison of the χ^2 values per degree of freedom, $\chi^2_{\rm dof}$, with ("dark") and without ("base") dark photon modifications for various datasets, taking into account a correction for missing higher order uncertainties (see text). 98

6.2 Couplings of the dark photon and Z' boson to the Standard Model fermions. The second column only shows the vector couplings, as the effect of the axial-vector couplings is negligible. 104

7.1 Algorithm performance metrics for the $10D$ Eggbox test function. . . . 115

7.2 Summary of algorithm performance for $10D$ Rosenbrock test function. 118

7.3 Summary of algorithm performance for $10D$ Rastrigin test function. . . 120

7.4 Summary of algorithm performance for $10D$ spike-slab test function . . 122

7.5 Prior ranges for the Lambda CDM model parameters. 124

7.6 Summary of algorithm performance for CMB likelihood 125
7.7 Summary of algorithm performance for MSSM7 likelihood 129

B.1 Reduced χ^2 values for longitudinal DIS polarization asymmetries A_1 and $A_\parallel$ for p, d, and ^{3}He (or "neutron") targets. The number of data points (N_{dat}) corresponds to a cut of $W^2 > 4$ GeV^2 on the DIS data. 140
B.2 Reduced χ^2 values for transverse DIS polarization asymmetries A_2, $A_\perp$, and $\widetilde{A}_\perp$ for p, d, and ^{3}He targets. The number of data points (N_{dat}) corresponds to a cut of $W^2 > 4$ GeV^2 on the DIS data. 141
B.3 Reduced χ^2 values for longitudinal SIDIS asymmetries A_1^h for production of hadrons h ($= \pi^\pm, K^\pm, h^\pm$) from p, d, and ^{3}He targets. 142
B.4 Reduced χ^2 values for jet and weak boson production in polarized pp collisions. Inclusive jets data are binned according to y_{jet} and p_T, while W/Z production are binned according to p_T and the electron rapidity, η_e. 143

C.1 MCMC-diffusion hyperparameter values assigned for each test problem. 158

D.1 Posterior mean for $10D$ Eggbox test function. The true mean for each parameter is 15.7 ± 9.3. 163
D.2 Posterior mean for $10D$ Rosenbrock test function. 163
D.3 Posterior mean for $10D$ Rastrigin test function. The true mean for each parameter is 0. 164
D.4 Posterior mean and error for $10D$ spike-slab test function. 164
D.5 Posterior mean and error for the CMB likelihood. 165
D.6 Posterior mean and error for the MSSM7 likelihood. 166
D.7 Marginal Wasserstein-1 distances for each parameter from reference posterior for CMB for cheapest converging NS runs. A '–' indicates none of the completed runs converged. 167
D.8 Marginal Wasserstein-1 distances for each parameter from reference posterior for MSSM7 for cheapest converging NS runs. A '–' indicates none of the completed runs converged. 167
D.9 Marginal Wasserstein-1 distances for each parameter from reference posterior for Eggbox for cheapest converging NS runs. A '–' indicates none of the completed runs converged. 169
D.10 Marginal Wasserstein-1 distances for each parameter from reference posterior for rastrigin for cheapest converging NS runs. A '–' indicates none of the completed runs converged. 169
D.11 Marginal Wasserstein-1 distances for each parameter from reference posterior for rosenbrock for cheapest converging NS runs. A '–' indicates none of the completed runs converged. 170

D.12 Marginal Wasserstein-1 distances for each parameter from reference posterior for spikeslab for cheapest converging NS runs. A '–' indicates none of the completed runs converged. 170
D.13 Number of MCMC walkers for each test problem. 170

List of Figures

1.1 One-loop quantum corrections to m_H^2 from (a) a Dirac fermion, f, and (b) a scalar, S [31]. 7

3.1 *(Left)* Quark distributions q_1 and q_2 in our toy model as a function of x. *(Right)* Example of toy "cross section" datasets σ_1 and σ_2 for the pseudodata (filled and open circles) and cross sections calculated from the input q_1 and q_2 quark distributions (blue and orange curves). 29

3.2 Comparison of fit results for the DR, HMC, NS and Hessian methods. The top row contains cross section and PDF distributions, with the middle and bottom rows show ratios to the true values for each distribution. 31

3.3 An example of an NN overfitting a single set of our toy data. 33

3.4 Comparison of training (blue) and validation (green) set loss functions for an overfitting example. 34

3.5 Comparison of the σ_1 replica set for the NN fit with (red band) and without (yellow band) cross validation. 34

3.6 Variation of the 1σ uncertainty with x across several f values from 0.1 to 0.9 for a single dataset of 50 points. 35

3.7 Total uncertainty ratio to $f = 0.5$ for different partition fractions as a function of number of data points. 36

3.8 Comparison of a single σ_1 replica set between NN and parametric DR methods (left panel), and the ratio of each predicted point to the true underlying law (right panel). 38

3.9 Study of confidence interval coverage for the NN (red lines) and parametric DR (blue lines) methods as a function of x. 39

4.1 Kinematics of polarized datasets included in this analysis with scale Q^2 displayed versus the relevant scaling variable. *(Top panel)* For polarized DIS and SIDIS the four-momentum transfer squared Q^2 is versus the Bjorken scaling variable; for jet production the transverse momentum squared is versus the Feynman variable; and for vector boson production the mass squared of the boson is versus x_F. For DIS, cuts on $W^2 > 4$ GeV^2 (dashed line) and $W^2 > 10$ GeV^2 (dot-dashed line) also shown for reference. *(Bottom panel)* The scale Q^2 is shown versus the fragmentation variable $z = P \cdot p_h / P \cdot q$ for polarized SIDIS, and versus $z = 2\, p_h \cdot q / Q^2$ for single-inclusive annihilation (SIA) of e^+e^- pairs into π, K and unidentified hadrons h. 57

4.2 Stability of truncated moments with respect to the cut used for the polarized inclusive DIS data, from $W^2_{\rm min} = 3.5$ GeV^2 to 10 GeV^2, at $Q^2 = 10$ GeV^2: (top row) lowest moments of PDFs Δu^+, Δd^+, Δs^+, and Δg; (middle row) triplet g_A, octet a_8, and singlet $\Delta\Sigma$ axial charges; (bottom row) proton and neutron d_2 moments. 59

4.3 Polarized PDFs versus x for various parton flavors with a $W^2 > 4$ GeV^2 cut on the inclusive polarized DIS data (red bands) compared with a stronger $W^2 > 10$ GeV^2 cut (yellow bands) at a scale $Q^2 = 10$ GeV^2. Kinematically the $W^2 > 10$ GeV^2 cut constrains the PDFs up to $x = 0.54$, while the $W^2 > 4$ GeV^2 cut allows PDFs to be constrained up to $x = 0.71$. 62

4.4 Ratios of unpolarized to polarized u and d quarks versus x for various parton flavors with a $W^2 > 4$ GeV^2 cut on the inclusive polarized DIS data (red bands) compared with a stronger $W^2 > 10$ GeV^2 cut (yellow bands) at a scale $Q^2 = 10$ GeV^2. 64

4.5 Polarized PDFs compared to other PDF fitting groups, computed at $Q^2 = 10$ GeV^2. NNPDF refers to the `NNPDFpol1.1` version. 65

4.6 Theory versus data comparisons for selected polarized jet DSAs $A_{LL}^{\rm jet}$ from the STAR 2015 ($\sqrt{s} = 200$ GeV) [123] and STAR 2013 ($\sqrt{s} = 510$ GeV) [124] datasets, for positive (red) and negative (yellow) Δg replicas. To the baseline fit (left column) are added lattice data (middle), and further additional high-x polarized DIS data (right). 66

4.7 Comparison of χ^2 contributions for positive (red dots) and negative (blue dots) Δg mean theory predictions for LQCD data versus eigendirection; χ^2_i represents the i-th lattice data point. 67

4.8 Monte Carlo replicas for the polarized gluon PDF $x\Delta g$ and the unpolarized gluon PDF xg at $Q^2 = 10$ GeV^2. The "baseline" and "+ LQCD" cases are cut off at $x \approx 0.53$, corresponding to a $W^2 > 10$ GeV^2 cut for polarized DIS. 68

4.9 Histogram of negative Δg replicas for the Higgs production asymmetry A_{LL}^{H} at RHIC ($\sqrt{s} = 510$ GeV), for the "baseline" analysis and after the addition of LQCD data. 69
4.10 xg_2^{HT} for proton and neutron evaluated at $Q^2 = 10$ GeV^2. 69
4.11 Polarized structure functions, for a $W^2 > 4$ GeV^2 cut on the inclusive polarized DIS data (red bands) compared with a stronger $W^2 > 10$ GeV^2 cut (yellow bands) at a scale $Q^2 = 10$ GeV^2. 70
4.12 d_2 moment for proton and neutron. 70
4.13 Burkhardt-Cottingham (BC) sum rule for proton and neutron. 71

5.1 2D Himmelblau function sampled with MCMC-diffusion (10,000 samples). Black lines depict jumps between modes due to the diffusion model. 81
5.2 Corner plot of MCMC-diffusion for a 10D Gaussian mixture with 2 components (5,000 samples). The true marginalized posteriors are shown in red, blue contours represent sample density. 82
5.3 Acceptance rate with MCMC-diffusion for a 10D Gaussian mixture with 2 components. 83
5.4 Comparison of MCMC-diffusion and pure MH for the 4D EggBox function (0.1M samples). Coloured contours represent sample density for each algorithm, with more identified modes representing greater success at characterising the shape of the true posterior. 85
5.5 Comparison of MCMC-diffusion and the emcee algorithm for the 4D EggBox function (0.1M samples). Coloured contours represent sample density for each algorithm, with more identified modes representing greater success at characterising the shape of the true posterior. 86
5.6 Comparison of fit results for MCMC-diffusion and pure MH (10,000 samples) to the many-sample limit (1M samples). The top row contains cross section and parton distributions, while the middle and bottom rows show ratios to the true values for each distribution. 87

6.1 Kinetic mixing of the dark photon A_D with SM mediators γ and Z in $e^{\pm}p$ DIS [270]. 92
6.2 The excluded region in the dark photon parameter space, assuming that the SM is the correct theory of Nature. 97
6.3 Results of an hypothesis test for the likelihood that the SM is the correct theory to describe this data, compared with the case where a dark photon is included. The hypothesis that the SM is the correct theory is excluded at 6.5σ for the best dark photon fit at the red point. 98
6.4 Points in DIS kinematics (Q^2, x) exhibiting the greatest improvement in residuals (> 0.3 in red, < 0.3 in blue) between dark and baseline average theory. 99

6.5 A comparison of the PDF replicas with ("dark", red bands) and without ("baseline", blue bands) the dark photon modifications at the input scale, $Q^2 = 1.69$ GeV^2. 101
6.6 Profile likelihood for the dark photon mass m_{A_D} versus the mixing parameter ϵ. 102
6.7 Exclusion limits for g_z at the 95% CL (90% CL for BaBar). The BaBar constraints are derived by rescaling the limits of the dark photon mixing parameter ϵ [262]. The upper bound on g_z from neutrino–quark $(\nu - q)$ scattering is taken from Ref. [320]. 105

7.1 Corner plot of the $10D$ Eggbox function in eq. (7.2). The plot shows the posteriors from the best performing MCMC and NS implementations. 116
7.2 Corner plot of the $10D$ Rosenbrock function. 117
7.3 Corner plot of the $10D$ Rastrigin function. 119
7.4 Corner plot of the $10D$ spike-slab function. 121
7.5 Validation plot of the network, showing the true log likelihood on the x-axis and the predicted value on the y-axis. 123
7.6 Corner plot for the CMB Likelihood. 126
7.7 Corner plot for the MSSM7 Likelihood. 128

B.1 Proton A_1 polarization asymmetry from COMPASS [179, 180], EMC [175], SMC [176, 177], and the SANE experiment at Jefferson Lab [373] versus Bjorken x, compared with the JAM fit (with 1σ uncertainty bands). Data in different Q^2 bins are offset by i for clarity. 144
B.2 Polarized DIS data for the observable $A_\parallel^p$ from HERMES [188], SLAC [181, 184, 186], and Jefferson Lab [191, 192] plotted as a function of x against the JAM result (colored 1σ bands). Data in different Q^2 bins are increased by i for clarity. 145
B.3 Polarized DIS data for the proton A_2 asymmetry from HERMES [196] and SANE [373] plotted as a function of x against the JAM result (red 1σ bands). Data in different Q^2 bins are increased by i for clarity. . . 146
B.4 Polarized DIS data for the observable $A_\perp$ (left panel) and $\widetilde{A}_\perp$ (right panel) from SLAC [184, 194, 195] plotted as a function of x against the JAM result (colored 1σ bands). Data in different Q^2 bins are increased by i for clarity. 146
B.5 Polarized DIS data for the observable A_1^D from COMPASS [178] and SMC [176, 177] plotted as a function of x against the JAM result (red 1σ bands). Data in different Q^2 bins are increased by i for clarity. . . 147

B.6 Polarized DIS data for the observable $A_{\parallel}^{D}$ from from HERMES [188], SLAC [184, 185], and Jefferson Lab [191, 193] plotted as a function of x against the JAM result (colored 1σ bands). Data in different Q^2 bins are increased by i for clarity. 147

B.7 Polarized DIS data for the observable $A_{\perp}^{D}$ (left panel) and $\widetilde{A}_{\perp}^{D}$ (right panel) from SLAC [184, 194, 195] plotted as a function of x against the JAM result (colored 1σ bands). Data in different Q^2 bins are increased by i for clarity. 148

B.8 Polarized DIS data for the observable $A_1^{n/^3\mathrm{He}}$ (left panel) and $A_{\parallel}^{^3\mathrm{He}}$ (right panel) from HERMES [187], SLAC [182, 183], and Jefferson Lab [189, 190] plotted as a function of x against the JAM result (colored 1σ bands). Data in different Q^2 bins are increased by i for clarity. 149

B.9 Polarized DIS data for the observable $A_2^{^3\mathrm{He}}$ (left panel) and $A_{\perp}^{^3\mathrm{He}}$ (right panel) from SLAC [182, 183] and Jefferson Lab [189, 190] plotted as a function of x against the JAM result (colored 1σ bands). Data in different Q^2 bins are increased by i for clarity. 150

B.10 Polarized SIDIS asymmetries $A_1^{\pi^{\pm}}$ from HERMES [198] (red circles) and COMPASS [199, 200] (green triangles) plotted as a function of x and compared to the JAM fit (1σ colored bands). 151

B.11 Polarized SIDIS asymmetries $A_1^{K^{\pm}}$ from HERMES [198] (red circles) and COMPASS [199, 200] (green triangles) plotted as a function of x and compared to the JAM fit (1σ colored bands). 152

B.12 Polarized proton (top row), deuteron (middle row), and ^{3}He (bottom row) SIDIS asymmetries $A_1^{h^{\pm}}$ from HERMES [198] (red circles), COMPASS [199] (green triangles), and SMC [374] (blue stars) as a function of x and compared to the JAM fit (1σ colored bands). 153

B.13 Polarized double longitudinal jet asymmetries A_{LL}^{jet} from STAR [201, 202, 203, 121, 122, 123, 124] (black circles) plotted as a function of p_T^{jet} and compared to the JAM fit (red 1σ bands). Each subplot shows the year when the data was taken and the pseudorapidity bins η_{jet} or $|\eta_{\mathrm{jet}}|$. . . . 154

B.14 Single spin asymmetries A_L^W and $A_L^{W/Z}$ versus lepton pseudorapidity η. *Left panel*: Asymmetries from STAR [204] (black circles) at $\sqrt{s} = 510$ GeV and integrated over $p_T > 25$ GeV, compared with the full JAM fit (red solid lines and 1σ uncertainty bands). *Right panel*: Asymmetries from PHENIX [205, 206] at $\sqrt{s} = 510$ GeV and integrated over $p_T > 16$ GeV (blue points) or $p_T > 30$ GeV (purple points), compared with the full JAM fit (red points). The W^+/Z asymmetries are shown with circles, while the W^-/Z asymmetries are shown with squares (we show points instead of bands due to the wide binning of the PHENIX data). 155

C.1 Comparison of MCMC-diffusion and pure MH for the 4D Rosenbrock function (0.5M samples). The true marginalized posteriors are shown in red, coloured contours represent sample density. 158
C.2 Increase in diffusion sample acceptance rate with number of retrains for the 4D Rosenbrock function (0.5M samples). 159

D.1 Upper bound on Wasserstein distance from reference posterior for each problem. The black dashed line is the assumed rule of thumb for convergence for each problem in the unit-cube parameterised domain. . . . 168

Declaration

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

I acknowledge the support I have received for my research through the provision of an Australian Government Research Training Program Scholarship.

Nicholas Hunt-Smith 24/2/2025

Acknowledgements

Obviously my supervisors are such an important part of making this all happen, Tony and Martin have been very supportive throughout my PhD and enabled me through their different expertise and connections to work on a wide variety of projects at the same time. Your understanding and flexibility has helped me be as productive as possible, and I've always felt that if I'm stuck on anything you'll be free for a quick email or meeting to clear things up. I also want to give a massive thanks to my JLab collaborators Nobuo, Chris and of course Wally, who I've had the most continuous contact with across the years even with the continental distance between us. My PhD would never have reached the heights it did without this help, it has made each project that we've worked on together really exciting. Wally has essentially been an extra supervisor who I can rely on at any time to give a knowledgeable and clear answer to any question I have. I know that every plot I make from now on will be extra high quality thanks to all those hours Wally and I have spent tweaking axes labels and font sizes!

Thank you to my fellow PhD students, I feel a real connection to the amazing physics community that exists at Adelaide which I believe is very rare and to be treasured. Despite my continual comings and goings I always felt welcome and supported, and I have so many happy memories of chats over lunch and at the desk, officeball breaks and taking in as much of the little night life that exists in Adelaide. I've definitely missed seeing you all and I hope we'll continue to catch up when we can wherever we are in the world. Thanks in particular to Mel, Meera, Josh and Shanette, and of course to Bill who has awakened my inner card game nerd and kept me sane throughout my PhD with lots of laughs but also deep conversations. Best of luck to all of you in your future careers either in or out of physics!

For my family, we've been through some of the toughest times in our lives in the last few years, and it's only thanks to our resilience that I've been able to get through the PhD with as little disruption as possible. Thank you to my mum, my sister Annabel and my Uncle Kim for checking in on how my work is going and riding the highs and lows with me, let's hope for some stability in our lives very soon. I hope that my grandparents are proud of me wherever they are, I miss you both very much.

To my partner Roya, the PhD has been a constant presence hanging over our lives

from the very start, and I can't thank you enough for joining me on this journey and guiding me through the most difficult moments. Very few would have been able to make it through so easily with all the uncertainty about what the reality of the PhD would be and then all the long-distance travel and upheaval, I'm so glad that we've gotten to the end of it even stronger than before. Thank you for sharing my frustration when a piece of code doesn't work, for listening to me rant about the stupdity of the latest referee report, and for being there at the end of each day to help me close the laptop and disengage from the day's work. I love you, and I can't wait to move on to the next stage of life with you.

Abstract

In the Standard Model (SM) of particle physics, the strong force stands as one of the central components that help us interpret our world down to the smallest scale. However, the theoretical framework of Quantum Chromodynamics (QCD) which governs the strong force is made significantly more difficult to apply as one reduces the energy of the particles involved. At this energy scale we can no longer operate entirely within the basic perturbative theory that we understand very well. One way of dealing with these non-perturbative effects inherent to many QCD interactions is to make use of global QCD analysis, whereby we take as much experimental particle collision data as we can and feed it into a model containing parton distribution functions (PDFs), which describe the makeup of various composite particles.

In this thesis we will investigate many different approaches to global QCD analysis, comparing how each method performs for various toy problems as well as realistic fits. We will develop new ways of improving the efficiency of such algorithms before testing them against one another. We will also perform multiple state-of-the-art global analyses. aiming to further constrain spin-dependent PDFs including new data at high-x as well as from lattice QCD. We will operate beyond the confines of the SM to search for potential dark matter candidates, placing limits on dark boson models using our knowledge of fundamental physics from global QCD analysis. In discussing these various topics, we will demonstrate the central role QCD has in many different fields within particle physics, as well as the necessity of sophisticated algorithmic approaches in order to further our understanding of hadronic structure.

Introduction

The prevailing theory which best represents our current knowledge of the fundamental building blocks that make up the world around us is the Standard Model of Particle Physics (SM) [1]. The SM has been enormously successful due in part to its predictive power, having preempted the discovery of the gluon, the W, Z and Higgs bosons as well as the top and charm quarks. A key pillar of the SM is Quantum Chromodynamics (QCD), which describes one of the four known fundamental forces of nature, that of the strong force. The strong force is essential to the stability of composite hadronic particles such as protons and neutrons that we encounter in abundance on a daily basis. Through the process of confinement, the partons (quarks and gluons) that make up these hadrons are held together by the strong force. However, as one increases the energy of a given particle collision, the interaction between partons reduces dramatically, to the point where the partons can be treated as if they were free particles. This property is called *asymptotic freedom* [2, 3, 4], and it allows us to accurately model interactions between partons within the perturbative regime for these high-energy reactions.

But what about low-energy reactions involving confined partons in hadrons? For these cases, we need to make use of non-perturbative approaches, along with QCD factorization. By factorizing the cross section into a perturbative (or "hard") and non-perturbative (or "soft") part, we can simplify the problem significantly. The "soft" part must define the underlying structure of the given hadron, which is determined by parton correlation functions. The most significant of these for the purposes of this thesis are the parton distribution functions (PDFs). Understanding the spin structure of the nucleon and the hadronization of partons, interpreting experimental data, precisely measuring SM parameters and searching for signals of physics beyond the SM, are just some of the phenomenological problems that rely on knowledge of PDFs.

This thesis is split into two parts. Part I will operate entirely within the confines of the SM, looking at improving our current understanding of PDFs. In Chapter 1 we trace a path from the mathematical basis of the SM through to the crucial role PDFs play in non-perturbative QCD, using deep-inelastic scattering as a concrete example. In Chapter 2 we explore the various methodologies employed by PDF fitting groups in the literature in detail, followed in Chapter 3 by a series of tests comparing these methods on a toy example of a PDF global fit. We then present a real global analysis

of PDFs in Chapter 4 using the state-of-the-art JAM fitting framework, performing the most sophisticated and complete version of said analysis to date, with a focus on high-x polarized deep-inelastic scattering data. In Chapter 5 we develop a new Bayesian sampling algorithm called `MCMC-diffusion`, combining Markov Chain Monte Carlo sampling with a diffusion model proposal distribution. We show increased performance with this algorithm over the standard Metropolis-Hastings algorithm for a toy PDF global analysis.

Part II explores various extensions of the SM by observing how PDFs and other global fitting methods can inform our study of such models. Chapter 6 looks at dark bosons, specifically how we can constrain dark photon and Z' models using a JAM global analysis of PDFs, before performing such an analysis to rule out new regions of parameter space. We also note an improvement in the global fit for certain values of the dark photon parameters. We then turn to the Λ-Cold Dark Matter model and the Minimal Supersymmetric Standard model in Chapter 7, comparing various Bayesian sampling algorithms to determine which ones perform best in the setting of a BSM global fit.

Part I

The Standard Model

Chapter 1

QCD in action

1.1 Standard Model Overview

The basic structure of the SM relies on Quantum Field Theory (QFT), in which each particle is an excitation of an associated dynamical field that permeates space-time. The SM consists of a set of field theories which are gauge-invariant, meaning that their dynamics are unchanged under a local gauge transformation. The gauge transformations themselves form various symmetry groups, each of which has an associated Lie algebra with various representations. The specific fields which represent the fundamental forces of nature, along with the matter content of the SM, are contained in these representations of the appropriate Lie algebra. In the SM, the symmetry group $SU(3) \times SU(2) \times U(1)$ describes the three different types of fundamental interactions between SM particles, namely strong, weak and electromagnetic. The electromagnetic and weak forces have a special connection, forming their own electroweak symmetry group $SU(2)_L \times U(1)$. These forces can only be treated separately after a process called Electroweak Symmetry Breaking (EWSB). EWSB allows us to generate the observed masses of certain particles at low energies while keeping the underlying structure of the electroweak gauge interactions intact [5, 6, 7]. The energy scale of this symmetry breaking mechanism is set at approximately 246 GeV, corresponding to the vacuum expectation value (VEV) of the CP-even component of the electroweak complex doublet scalar field. Once this energy threshold is crossed, three massless Goldstone bosons are generated before being absorbed to give the W and Z bosons their mass. The all-important Higgs boson is left as a fundamental scalar particle from the remaining component of the complex doublet. It is only with the inclusion of the Higgs boson that all fundamental fermions receive their mass, as dictated by Yukawa interactions.

The strong force is the primary focus of this thesis, and is described by the remaining non-Abelian SU(3) gauge field theory known as QCD. Embedded within the SU(3) gauge group are the spin-$\frac{1}{2}$ fermionic quarks, where the quark fields ψ are said to

have "colour charge". Within quantum electrodynamics (QED), the mathematical formalism describing the electromagnetic force, we have the concept of electric charge - this colour charge can be thought of as its QCD counterpart. In addition to the $N_f = 6$ different flavours of quarks in the SM (up, down, strange, charm, bottom and top), quarks also come in three colours $N_C = 3$ (often labelled red, green and blue). The spin-1 gauge bosons in QCD A_μ are called gluons, acting as the colour force carriers. In the adjoint representation of the SU(3) gauge group, the dimension is $N_C^2 - 1 = 8$, meaning that the gluons form a colour octet. Essentially, each type of gluon can be thought of as a 3×3 matrix that can rotate a quark's colour in SU(3) space.

A key theme across the SM is the use of global fitting as an indispensible tool, allowing us to combine all relevant experimental data and learn the unknown parameters of a particle physics model [8, 9, 10, 11, 12]. Along with global fits in QCD [13, 14, 15, 16], this tool is also prevalent in neutrino physics [17, 18, 19, 20], effective field theories [21, 22, 23, 24], and in beyond the Standard Model phenomenology [25, 26, 27, 28, 29, 30]. There are two dominant approaches to global fitting that are widely used across all sciences: frequentist and Bayesian inference. The latter is common in particle physics, and will be used extensively throughout this thesis.

Although the SM has been very successful in explaining a wide range of phenomena, there exist several observations that do not appear to be fully understood within the confines of the SM. The rotation curves of galaxies as well as the cosmic microwave background (CMB) tell us that there is far more matter in the universe than is predicted by the SM - in fact, only 5% of the mass-energy content can be accounted for by the SM. The remaining 95% consists of an unknown substance we call dark matter (DM), as well as dark energy. The SM also cannot explain how there can be so much more matter than antimatter in the universe. At some point in the early universe a process we call baryogenesis must have taken place to create this matter-antimatter asymmetry, yet the electroweak mechanism offered up by the SM is insufficient to account for the amount of excess matter we currently observe.

On a more abstract level we have the hierarchy problem, which questions why the observed Higgs mass ($\approx$ 125 GeV) is so low, particularly compared to the Planck scale ($\approx 10^{19}$ GeV). The Higgs mass receives enormous quantum corrections from any particle that couples to the Higgs field - for example, a one-loop diagram containing a Dirac fermion f (see Fig. 1.1a) interacting with the Higgs field through a Lagrangian term $-\lambda_f H \bar{f} f$ would introduce a quantum correction of [31]

$$\Delta m_H^2 \approx \frac{|\lambda_f|^2}{8\pi^2}\Lambda^2, \tag{1.1}$$

where Λ is the momentum cutoff for loop integrals. Similarly, a one-loop diagram containing a a complex scalar particle S (see Fig. 1.1b) interacting with the Higgs field through a Lagrangian term $-\lambda_S |H|^2 |S|^2$ would introduce a quantum correction of

$$\Delta m_H^2 \approx \frac{\lambda_S}{16\pi^2}\left[\Lambda^2 - 2m_S^2 \ln\left(\frac{\Lambda}{m_S}\right)\right]. \tag{1.2}$$

The issue is that if we choose Λ to be large, say as large as the Planck scale, the correction from either of these processes will be many orders of magnitude larger than the observed Higgs mass. Eqs. (1.1) and (1.2) naturally suggest to us that m_H would be expected to take on a value at a similar scale to Λ, and the fact that it is so much lower makes the Higgs mass a prime example of *fine tuning*. Important parameters that appear to be picked out so carefully can be extremely philosophically troubling for many physicists, so much so that alternative theories are often devised in order to avoid any such "unnatural" model behaviour.

It seems necessary to introduce some additional physics beyond the standard model (BSM) if we are to address these issues. There are many such solutions, however we will touch on only a couple of those relevant to our future discussion in Part II.

One way to sidestep the hierarchy problem is to adopt a supersymmetric extension to the SM. The simplest such model is the minimal superysymmetric model (MSSM), which involves doubling the number of particles by introducing a superpartner for every SM particle. By including two complex scalar superpartners for every fermion with $\lambda_S = |\lambda_f^2|$, the quadratic contributions to eqs. (1.1) and (1.2) will exactly cancel out, leaving us with only the logarithmic dependency on Λ. In this way, the large quantum corrections to the Higgs mass are naturally removed, avoiding the issue of fine tuning in the SM. While supersymmetry may appear elegant on paper, no supporting experimental evidence has yet been found.

The dark photon is a simple modification one can make to the SM to begin to describe the role DM might play in our universe. As an additional $U(1)$ gauge boson, the dark photon can be thought of as a dark sector counterpart to the regular SM photon, acting as a force carrier for some unknown dark force. An important difference is that the dark photon has a mass, making it more akin to a dark Z boson. The dark photon would not be a DM candidate on its own, instead it would be expected to interact with other DM particles, providing us with a "portal" into the dark sector.



(a) (b)

Figure 1.1: One-loop quantum corrections to m_H^2 from (a) a Dirac fermion, f, and (b) a scalar, S [31].

Many accelerator-based searches for the dark photon have been undertaken, with none observing a signal to date.

Having explored the SM at large, we will now hone in on the QCD component which forms the backbone of this thesis.

1.2 QCD Formalism and Factorization

The QCD Lagrangian encodes the dynamics of the interactions between quarks and gluons, and it can be written as

$$\mathcal{L}_{\text{QCD}} = \sum_q \overline{\psi}_q^i (i\gamma^\mu D_\mu^{ij} - \delta^{ij} m_q)\psi_q^j - \frac{1}{4} F_{\mu\nu}^{(a)} F^{(a)\ \mu\nu}, \tag{1.3}$$

where γ^μ are the Dirac γ-matrices, the sum is over all quark flavours $q = u, d, s, c, b, t$, m_q is the mass of a given quark flavour, i, j correspond to the fundamental representation colour indices and a corresponds to the adjoint representation colour indices. The covariant derivative is given by

$$D_\mu^{ij} = \delta^{ij}\partial_\mu - igA_\mu^{ij}, \tag{1.4}$$

where g is the quark-gluon coupling. The gluon field strength tensor is defined as

$$F_{\mu\nu}^{(a)} = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a + g f_{abc} A_\mu^b A_\nu^c, \tag{1.5}$$

where f_{abc} are the SU(3) structure constants.

A key concept in perturbative QCD is that of renormalization, which allows one to calculate physical observables at a specific renormalization scale μ_R. In particular, the renormalized coupling α_S is a function of μ_R, and if one wants to determine the magnitude of the strong force for a given interaction then one must take μ_R^2 close to Q^2, the momentum transfer for the scattering process one wishes to study. Using the renormalization group equations, one finds that

$$\alpha_S(Q^2, \mu_R^2) = \frac{\alpha_S}{1 + \frac{\alpha_S}{4\pi}\left(\frac{11}{3}N_C - \frac{2}{3}N_f\right)\ln\left(\frac{Q^2}{\mu_R^2}\right)}. \tag{1.6}$$

Eq. (1.6) expresses mathematically two key concepts within QCD. The first is that of asymptotic freedom - due to the running of the QCD coupling, at high energies/short distances the interaction between quarks goes to zero. This means the quarks can be treated as entirely free particles, and their behaviour can be fully described in the perturbative regime. However, in the case of low energies/large distances, α_S goes to infinity. This rapid increase in the coupling means that quarks are primarily observed bound together in stable composite particles. The confinement of quarks in this way is

the second important conclusion that can be drawn from Eq. (1.6). Given the size of α_S at these low energies we clearly cannot rely on perturbative techniques alone. One approach is to split the problem into a perturbative "hard" part and a non-perturbative "soft" part, through a process called QCD factorization [32].

At the core of the non-perturbative aspect of QCD are the parton distribution functions (PDFs), which can be denoted by $f_i(x, \mu_F)$. These 1D functions define the probability of obtaining a given parton i at a certain longitudinal momentum fraction x of a hadron's total longitudinal momentum, computed at a specific factorization scale μ_F. The PDFs can only be inferred from first principles alone by discretizing spacetime using lattice QCD, however they can also be inferred through a variety of techniques including fitting to experimental data.

1.3 PDF Parametrization

Although we don't know the exact form of the PDFs at this stage, thankfully we do know how to evolve them with respect to the factorization scale μ_F. This is due to the crucial DGLAP equations, defined as [33, 34, 35]

$$\mu_F^2 \frac{\partial f_i(x_i, \mu_F^2)}{\partial \mu_F^2} = \sum_j \int_{x_i}^1 \frac{dz}{z} P_{ij}(z) f_j\left(\frac{x_i}{z}, \mu_F^2\right). \tag{1.7}$$

Here we can see QCD factorization in action, with interplay between the perturbative "splitting functions" P_{ij} corresponding to the emission of a collinear gluon by a quark/antiquark (which retains momentum fraction z after this emission), and the non-perturbative PDFs $f_i(x_i, \mu_F^2)$. If we require that the perturbative corrections be small by associating μ_F with a specific energy scale corresponding to a particular experimental process Q^2, then we can relate the PDFs to data obtained from such an experiment.

As well as knowing the evolution of the PDFs with respect to the energy scale, we can also place several constraints on the PDFs in the form of sum rules that will help us in our aim of determining the PDFs. Quarks within a given hadron can be split into two categories: sea quarks and valence quarks. Sea quarks arise out of $q\bar{q}$ pairs popping into existence from the QCD vacuum due to quantum effects, and may annihilate at any time. Valence quarks are then defined as the total quark content left in a given hadron after removing these sea quarks, which equates to the difference in total quarks and antiquarks. For example, the valence up quark distribution would be:

$$u_V(x, \mu_F^2) = u(x, \mu_F^2) - \bar{u}(x, \mu_F^2). \tag{1.8}$$

Here we have changed notation to show the specific PDF flavours rather than using a flavour index, meaning $u(x, \mu_F^2) = f_1(x_i, \mu_F^2)$. If we now turn to a specific case, the

proton is said to have 2 valence up quarks and 1 valence down quark, with no strange valence. For the proton, this therefore implies

$$\int_0^1 dx\ u_V(x,\mu_F^2) = 2, \tag{1.9}$$

$$\int_0^1 dx\ d_V(x,\mu_F^2) = 1, \tag{1.10}$$

$$\int_0^1 dx\ s_V(x,\mu_F^2) = 0. \tag{1.11}$$

The total momentum fraction of all the partons within the proton including the gluon g weighted by the PDFs must of course add up to 1, meaning that

$$\int_0^1 \mathrm{d}x\ x\left(u^+(x,\mu_F^2) + d^+(x,\mu_F^2) + s^+(x,\mu_F^2) + g(x,\mu_F^2)\right) = 1, \tag{1.12}$$

where $q^+ = q + \bar{q}$ and we neglect heavier quarks which contribute a negligible amount in the proton. Similar sum rules apply for other hadrons which we will not highlight here. Many scattering experiments use neutrons as a target instead of protons, however to accurately characterise neutron structure we do not require an entire separate set of PDF functions. Making use of isospin symmetry, all of the neutron PDFs are identical to the proton PDFs apart from the neutron up u_n and down d_n PDFs, which can be related to the proton u_p and d_p PDFs according to

$$u_n(x,\mu_F^2) = d_p(x,\mu_F^2), \tag{1.13}$$

$$d_n(x,\mu_F^2) = u_p(x,\mu_F^2). \tag{1.14}$$

Note that connecting the proton and neutron PDFs in this way involves an approximation, where we neglect the mass difference between the up and down quarks [36].

Having established the fundamental rules that any set of PDFs must follow, we can turn to the precise mathematical form the PDFs have. Since the PDFs remain a non-perturbative quantity to be constrained, the best approach we have is to maintain a highly flexible parametrization without introducing too many parameters to the point where the problem becomes computationally intractable. The most common method, and the one that will be employed for much of this thesis, is the below parametrization:

$$f(x,\mu_0) = Nx^\alpha(1-x)^\beta(1+\gamma\sqrt{x}+\eta x), \tag{1.15}$$

computed at the input scale μ_0 with free fitting parameters N, α, β, γ, and η for each parton flavor, f. This shape reflects the power-like behaviour that is expected as $x \to 0$ and $x \to 1$, and can take on a variety of forms within those limits while

introducing only 5 free parameters per parton flavor. Note that for particularly poorly constrained PDFs the η and/or γ parameters may be set to zero in order to simplify the problem. There also exist more complex parametrizations, chief among which is the neural network parametrization by the NNPDF fitting group, see Sec. 3.2.2 for further discussion. The neural network approach introduces many more parameters, and still partially relies on the form in Eq. (1.15) for computational efficiency.

1.4 Deep Inelastic Scattering

So far we know the general form of the PDFs, and how to evolve them from a perturbative input scale μ_0 up to the characteristic energy scale of a given scattering process Q through the use of the DGLAP equations. We will now study the deep inelastic scattering (DIS) process in greater detail, in order to see exactly how the evolved PDFs can be related to a specific experimental observable.

Deep inelastic scattering involves a high-energy lepton l with four-momentum k scattering off a nucleon/nucleus target N with four-momentum P, altering the lepton's four-momentum which we now call k' after the interaction. In this process l may exchange a photon γ, a $W^{\pm}$ boson or a Z boson with the target N, and the four-momentum of this exchanged particle q must be equal to the change in four-momentum of the lepton, i.e. $q = k - k'$. The recoiling system X with invariant mass W consists of all remnants of the target N after the collision, such that the process can be described as $lN \to l'X$. For the purposes of the following discussion, N is assumed to be a proton of mass M.

We can first define several invariant quantities, assuming massless leptons:

$$Q^2 \equiv -q^2, s \equiv (P+k)^2, x_{\mathrm{bj}} \equiv \frac{Q^2}{2p\cdot q} = \frac{Q^2}{2M\nu}, y \equiv \frac{P\cdot q}{P\cdot k} = \frac{Q^2}{x_{\mathrm{bj}}(s-M^2)}, \tag{1.16}$$

where s is the squared centre-of-mass energy of the reaction, Q^2 is the squared four-momentum transfer, x_{bj} is the Bjorken scaling variable, y is the elasticity and ν is the change in energy of the lepton l after the reaction in the rest frame of the target N. For unpolarized DIS, the cross section can be written as:

$$\frac{\mathrm{d}^2\sigma}{\mathrm{d}x_{\mathrm{bj}}\mathrm{d}y} = \frac{2\pi y\alpha^2}{Q^4}\sum_j \eta_j L^{\mu\nu}_{U,j} W^U_{\mu\nu}, \tag{1.17}$$

where α is the electromagnetic fine structure constant, $L^{\mu\nu}_{U,j}$ is the unpolarized leptonic tensor, $W^U_{\mu\nu}$ is the unpolarized hadronic tensor and j runs over the different available channels for each process. For neutral current (NC) interactions, $j = \gamma, \gamma Z, Z$ and for charged current (CC) interactions $j = W$. The η_j corresponding to each channel are:

$$\eta_\gamma = 1, \tag{1.18}$$
$$\eta_{\gamma Z} = \frac{G_F M_Z^2}{2\sqrt{2}\pi\alpha} \frac{Q^2}{Q^2 + M_Z^2}, \tag{1.19}$$
$$\eta_Z = \eta_{\gamma Z}^2, \tag{1.20}$$
$$\eta_W = \frac{1}{2}\left(\frac{G_F M_W^2}{4\pi\alpha} \frac{Q^2}{Q^2 + M_W^2}\right)^2, \tag{1.21}$$

where G_F is the Fermi constant, and M_Z and M_W are the masses of the Z and W bosons respectively. In the case of NC processes, the unpolarized leptonic tensors are

$$L_{\mu\nu}^{U,\gamma} = 2(k_\mu k'_\nu + k'_\mu k_\nu - g_{\mu\nu} k \cdot k'), \tag{1.22}$$
$$L_{\mu\nu}^{U,\gamma Z} = g_V^e L_{\mu\nu}^{U,\gamma}, \tag{1.23}$$
$$L_{\mu\nu}^{U,Z} = \left[(g_V^e)^2 + (g_A^e)^2\right] L_{\mu\nu}^{U,\gamma}, \tag{1.24}$$

where $g_V^e = -\frac{1}{2} + 2\sin^2\theta_W$ and $g_A^e = -\frac{1}{2}$ are the respective vector and axial couplings of the electron to the Z. The unpolarized hadronic tensor is defined as

$$W_{\mu\nu}^U = \left(\frac{q_\mu q_\nu}{q^2} - g_{\mu\nu}\right) F_1^p + \left(P^\mu + \frac{q^\mu}{2x_{\rm bj}}\right)\left(P^\nu + \frac{q^\nu}{2x_{\rm bj}}\right)\frac{F_2^p}{P\cdot q} - i\epsilon_{\mu\nu\alpha\beta}\frac{q^\alpha P_\beta}{2P\cdot q}F_3^p, \tag{1.25}$$

where $\epsilon_{\mu\nu\alpha\beta}$ is the Levi-Civita tensor and we introduce the proton structure functions $F_{1,2,3}^p$. F_1^p and F_2^p are parity-conserving, since their coefficients are symmetric under a $\mu \leftrightarrow \nu$ transformation, while F_3^p is parity-violating. These structure functions play a key role as they contain the all-important PDFs. Expanding in terms of $\frac{M^2}{Q^2}$ and $\frac{\Lambda_{\rm QCD}^2}{Q^2}$, the NC structure functions can be written in terms of the PDFs q, g as:

$$F_1^p = \frac{1}{2}\sum_q e_q^2[C_{1,q}^{\rm DIS} \otimes q^+ + C_{1,g}^{\rm DIS} \otimes g] + \mathcal{O}\left(\frac{M^2}{Q^2}, \frac{\Lambda_{\rm QCD}^2}{Q^2}\right), \tag{1.26}$$
$$F_2^p = x\sum_q e_q^2[C_{2,q}^{\rm DIS} \otimes q^+ + C_{2,g}^{\rm DIS} \otimes g] + \mathcal{O}\left(\frac{M^2}{Q^2}, \frac{\Lambda_{\rm QCD}^2}{Q^2}\right), \tag{1.27}$$
$$F_3^p = \sum_q e_q^2[C_{3,q}^{\rm DIS} \otimes q^-] + \mathcal{O}\left(\frac{M^2}{Q^2}, \frac{\Lambda_{\rm QCD}^2}{Q^2}\right), \tag{1.28}$$

where e_q is the charge of a given quark with flavor q, and $q^\pm = q \pm \bar{q}$. The DIS perturbative hard scattering coefficients $C_{i,f}^{\rm DIS}$ are given at next-to-leading order (NLO) by:

$$C_{i,f}^{\mathrm{DIS}} = C_{i,f}^{\mathrm{DIS}(0)} + \frac{\alpha_s(\mu_R)}{4\pi} C_{i,f}^{\mathrm{DIS}(1)} + \mathcal{O}(\alpha_s^2), \tag{1.29}$$

with the coefficients defined as in [37]. Just as for the factorization scale μ_F, the renormalization scale μ_R is taken to be equal to Q for the DIS process. Performing the contraction in Eq. (1.17), we obtain the NC cross section:

$$\frac{\mathrm{d}^2\sigma^{\mathrm{NC}}}{\mathrm{d}x_{\mathrm{bj}}\mathrm{d}y} = \frac{4\pi\alpha^2}{x_{\mathrm{bj}} y Q^2} \left[y^2 x_{\mathrm{bj}} F_1^p + \left(1 - y - \frac{x_{\mathrm{bj}} y^2 M^2}{Q^2}\right) F_2^p \mp \left(y - \frac{y^2}{2}\right) x_{\mathrm{bj}} F_3^p \right], \tag{1.30}$$

where the plus (minus) corresponds to negative (positive) lepton scattering. The structure functions can also be related to the individual channels of the DIS process according to:

$$F_{1,2}^p = F_{1,2}^{p,\gamma} - g_V^e \eta_{\gamma Z} F_{1,2}^{p,\gamma Z} + \left[(g_V^e)^2 + (g_A^e)^2\right] \eta_Z F_{1,2}^{p,Z}, \tag{1.31}$$

$$F_3^p = -g_A^e \eta_{\gamma Z} F_3^{p,\gamma Z} + 2 g_V^e g_A^e \eta_Z F_3^{p,Z}. \tag{1.32}$$

In the CC case for which the only channel is $W^{\pm}$, the leptonic tensor is simply given by

$$L_{\mu\nu}^{U,W} = 4(k_\mu k'_\nu + k'_\mu k_\nu - g_{\mu\nu} k \cdot k'). \tag{1.33}$$

The CC structure functions have a similar form to the NC case in terms of their relation to the PDFs:

$$F_1^{p,W^{\pm}} = \sum_q [C_{1,q}^{\mathrm{DIS}} \otimes q + C_{1,g}^{\mathrm{DIS}} \otimes g] + \mathcal{O}\left(\frac{M^2}{Q^2}, \frac{\Lambda_{\mathrm{QCD}}^2}{Q^2}\right), \tag{1.34}$$

$$F_2^{p,W^{\pm}} = 2x \sum_q [C_{2,q}^{\mathrm{DIS}} \otimes q + C_{2,g}^{\mathrm{DIS}} \otimes g] + \mathcal{O}\left(\frac{M^2}{Q^2}, \frac{\Lambda_{\mathrm{QCD}}^2}{Q^2}\right), \tag{1.35}$$

$$F_3^{p,W^{\pm}} = 2 \sum_q [I_q C_{3,q}^{\mathrm{DIS}} \otimes q] + \mathcal{O}\left(\frac{M^2}{Q^2}, \frac{\Lambda_{\mathrm{QCD}}^2}{Q^2}\right), \tag{1.36}$$

however in the CC scenario $q = \bar{u}, d, \bar{s}, c, \bar{t}, b$ for W^+, $q = u, \bar{d}, s, \bar{c}, t, \bar{b}$ for W^- and $I_q = 1$ for quarks and -1 for antiquarks. Repeating the contraction in Eq. (1.17), we obtain

$$\frac{\mathrm{d}^2\sigma^{\mathrm{CC}}}{\mathrm{d}x_{\mathrm{bj}}\mathrm{d}y} = \frac{8\pi\alpha^2}{x_{\mathrm{bj}} y Q^2} \eta_W \left[y^2 x_{\mathrm{bj}} F_1^{p,W^{pm}} + \left(1 - y - \frac{x_{\mathrm{bj}} y^2 M^2}{Q^2}\right) F_2^{p,W^{pm}} \mp \left(y - \frac{y^2}{2}\right) x_{\mathrm{bj}} F_3^{p,W^{pm}} \right]. \tag{1.37}$$

We have outlined here a methodology for computing the cross section for unpolarized inclusive DIS as a demonstration of how PDFs play a fundamental role in allowing us to generate meaningful predictions of the outcomes of hadronic scattering experiments. There are of course many other types of hard scattering processes that exist, however all can be handled through a similar procedure of computing an observable such as a cross section by evolving the PDFs up to the characteristic energy scale of the given process through the DGLAP equations. Now that we have established how one can compare theory and experiment directly, we will now explore the various approaches one can take to determining the PDFs and their associated uncertainties.

Chapter 2

Global Analysis of PDFs

Considerable progress has been made in recent years through several parallel efforts to improve our knowledge of PDFs, both experimentally and theoretically. The inclusion of an increasingly vast and precise dataset and the growth of the theoretical accuracy of QCD analyses have been accompanied by an increasing sophistication of the methodological procedures used to determine the PDFs [38]. These currently include the Hessian [39], Lagrange multiplier [40] and data resampling (DR) methods. The uncertainty estimation method most commonly adopted in the literature has been the Hessian, which has been used by many collaborations in their determination of unpolarized PDFs (see Refs. [41, 42, 43] for a review). The Lagrange multiplier method was used by the DSSV group to determine the polarized proton PDFs [44, 45, 46], while the DR method was used by the NNPDF collaboration [47, 48, 49, 50] in their determination of the unpolarized and polarized proton PDFs (for recent work, see, *e.g.*, Refs. [51, 52]). It was also employed by the JAM collaboration in their determinations of the polarized proton PDFs [53], as well as more recently in the JAM global QCD analyses of unpolarized proton [54, 55] and pion [56, 57] PDFs. Although the NNPDF and JAM approaches share the same method of quantifying uncertainties, they differ in their parametrization of the PDFs. While the JAM collaboration uses a traditional polynomial functional form for the dependence of the PDFs on the quark momentum fraction x, the NNPDF collaboration implements a similar basic parametric form supplemented by a series of trained neural network weights.

We should note that the above methods all attempt to find approximations to the Bayesian posterior distribution for the PDF parameters. An alternative approach is to calculate the posterior exactly using Bayes' theorem. This can then be used to obtain the expectation value and standard deviation of the set of PDF parameters, and of any derived quantity. A popular posterior mapping technique is nested sampling (NS) [58], a Monte Carlo method targeted at the efficient calculation of the Bayesian evidence, which is the average likelihood of a model over its probability space. This method produces posterior inferences as an advantageous byproduct, and contemporary

implementations of NS algorithms, such as `MultiNest` [59, 60] and `PolyChord` [61, 62], are now extensively used for parameter estimation from posterior samples in cosmology (see, *e.g.* Ref. [63]). They have also very recently been used in global QCD analyses of the quark transversity distributions [64] and of the pion distributions [65] by the JAM collaboration.

A popular alternative to nested sampling is the Hamiltonian (or hybrid) Markov chain Monte Carlo (HMC) technique [66, 67], based on the Metropolis-Hastings algorithm [68, 69], which allows for an efficient sampling of posterior inferences, though without the calculation of the evidence which is typically not required in PDF fits. These methods are implemented in xFitter [70], the open-source package for performing global PDF fits, which has been shown to lead to competitive results [71].

Here we outline the statistical underpinnings of Bayesian parameter inference, before reviewing the Hessian, DR, NS and HMC approaches to estimating the Bayesian posterior in the context of a global PDF analysis.

2.1 Bayesian inference

In particle physics, Bayesian inference is widely used not only in fits of PDFs [72, 73, 74], but also in extraction of signs of physics beyond the Standard Model [75, 76, 77, 78, 79], analysis of effective field theories [78, 80, 10], and the determination of the parameters of the neutrino sector [81, 82, 83]. In particle astrophysics, applications range from neutrino astronomy [84, 85, 86], through to cosmic ray studies [87], dark matter searches and phenomenology [88, 89, 90, 91], gravitational wave astronomy [92, 93, 94], and global fits and model comparison [95, 10, 96, 83].

In the framework of Bayesian inference [97, 98, 99] one seeks to determine the probability measure $p([f])$ associated with a space of functions $[f]$. The expectation value $E\{\mathcal{O}[f]\}$ and the variance $V\{\mathcal{O}[f]\}$ of any physical observable $\mathcal{O}$ that depends on a set of PDFs $[f]$ can then be expressed by integration over the space of functions, respectively, of the first and second statistical moments of $\mathcal{O}[f]$, evaluated on each function f and weighted by the probability measure $p([f])$.

The probability measure $p([f])$ is an infinite-dimensional object in the space of functions $[f]$. However, it should be necessarily determined from a finite set of measurements. The problem, which would otherwise be ill-defined, is usually projected from the infinite-dimensional space of functions into a finite-dimensional space of parameters by choosing a suitable parametrization for the set of functions $[f]$. Such a procedure of course requires some assumptions. Some of these, such as a certain degree of smoothness or particular behaviors in certain regions of the functional space, may be physically motivated; nevertheless, they should not bias the result nor affect its statistical interpretation.

Calling $\boldsymbol{a} = \{a_1, \ldots, a_{n_{\rm par}}\}$ the set of $n_{\rm par}$ parameters and $\boldsymbol{m} = \{m_1, \ldots, m_{n_{\rm dat}}\}$ the

set of n_{dat} measurements, the expectation value and variance of an observable $\mathcal{O}$ are defined in the Bayesian framework by

$$E_{\text{Bayes}}\{\mathcal{O}(\boldsymbol{a})\} = \langle \mathcal{O}(\boldsymbol{a}) \rangle = \int \mathrm{d}\boldsymbol{a}\; p(\boldsymbol{a}|\boldsymbol{m})\, \mathcal{O}(\boldsymbol{a})\,, \tag{2.1a}$$

$$V_{\text{Bayes}}\{\mathcal{O}(\boldsymbol{a})\} = \sigma^2_{\mathcal{O}}(\boldsymbol{a}) = \int \mathrm{d}\boldsymbol{a}\; p(\boldsymbol{a}|\boldsymbol{m}) \big(\mathcal{O}(\boldsymbol{a}) - E_{\text{Bayes}}\{\mathcal{O}(\boldsymbol{a})\}\big)^2\,, \tag{2.1b}$$

where $p(\boldsymbol{a}|\boldsymbol{m})$ is the conditional probability of the set of parameters $\boldsymbol{a}$, given the set of measurements $\boldsymbol{m}$. The problem of determining the probability measure in the space of functions $p([f])$ is then reduced to the problem of determining the conditional probability $p(\boldsymbol{a}|\boldsymbol{m})$.

One can readily formulate the problem in terms of Bayes' theorem, which can be expressed as

$$p(\boldsymbol{a}|\boldsymbol{m}) = \frac{1}{\mathcal{Z}}\, p(\boldsymbol{m}|\boldsymbol{a})\, p(\boldsymbol{a})\,, \quad \text{with} \quad \mathcal{Z} = \int \mathrm{d}\boldsymbol{a}\; p(\boldsymbol{m}|\boldsymbol{a})\, p(\boldsymbol{a})\,, \tag{2.2}$$

where $p(\boldsymbol{a})$ is the prior, $p(\boldsymbol{a}|\boldsymbol{m})$ is the posterior, $p(\boldsymbol{m}|\boldsymbol{a}) \equiv \mathcal{L}(\boldsymbol{a};\boldsymbol{m})$ is the likelihood, and $\mathcal{Z}$ is the evidence. The prior is a subjective quantity, and results of a Bayesian analysis are frequently found to be highly dependent on the choice of this function. For problems where the data are sufficiently constraining (as expected in the case of PDF fits), this is less of an issue, and the posterior distribution is mostly determined by the likelihood function. For linear problems, flat priors can be useful: $p(\boldsymbol{a})$ is associated with a uniform distribution, normalized to unit volume, $p(\boldsymbol{a}) = \mathbf{1}$; Eq. (2.2) then implies that

$$p(\boldsymbol{a}|\boldsymbol{m}) \propto \mathcal{L}(\boldsymbol{a};\boldsymbol{m})\,. \tag{2.3}$$

The likelihood $\mathcal{L}(\boldsymbol{a};\boldsymbol{m})$ is seen as a mathematical function of the parameters $\boldsymbol{a}$, given the data $\boldsymbol{m}$. For n_{dat} pairs of independent measurements $\boldsymbol{m} = \{x_i, y_i\}$, $i = 1, \ldots, n_{\text{dat}}$, a common choice is

$$\mathcal{L}(\boldsymbol{a};\boldsymbol{m}) = \mathcal{N} \exp\Big[-\frac{1}{2}\chi^2(\boldsymbol{a},\boldsymbol{m})\Big], \tag{2.4}$$

where $\mathcal{N}$ is an appropriate normalization factor, and the general definition for the χ^2 (often called logarithmic likelihood) is

$$\chi^2(\boldsymbol{a},\boldsymbol{m}) = \sum_{i,j}^{n_{\text{dat}}} \big(T^i(\boldsymbol{a}) - m^i\big)\, C^{\text{dat}}_{ij}\, \big(T^j(\boldsymbol{a}) - m^j\big)\,, \tag{2.5}$$

where $T(\boldsymbol{a})$ represents the predictions of the model for the underlying law given some set of parameters $\boldsymbol{a}$, and C^{dat} defines the covariance matrix in the space of the data. In

the specific case of a global QCD analysis performed with some data from a scattering experiment such as the DIS process in 1.4, the χ^2 can instead be written as

$$\chi^2(\boldsymbol{a}) = \sum_{e,i} \left(\frac{d_{e,i} - \sum_k r_{e,k}\beta_{e,i}^k - T_{e,i}(\boldsymbol{a})/N_e}{\alpha_{e,i}} \right)^2, \tag{2.6}$$

where $d_{e,i}$ is the data point i from experimental dataset e, and $T_{e,i}$ is the corresponding theoretical value for a given set of PDF parameters a. All uncorrelated uncertainties are added in quadrature and labeled by $\alpha_{e,i}$, while $\beta_{e,i}^k$ represents the k^{th} source of point-to-point correlated systematic uncertainties for the i^{th} data point weighted by $r_{e,k}$. Normalization parameters N_e are included for each dataset e, to be fitted alongside the PDF parameters.

If we were only interested in the mean value of the PDFs $E_{\text{Bayes}}\{\mathcal{O}(\boldsymbol{a})\}$ we would have a relatively simple problem to solve. We would maximize the likelihood/minimize the χ^2 with respect to $\boldsymbol{a}$, which would provide us with the optimized PDF parameters corresponding to the best fit. However, the experimental data have associated uncertainties, and we wish to carry these uncertainties through the analysis to characterize the uncertainties on the PDFs $V_{\text{Bayes}}\{\mathcal{O}(\boldsymbol{a})\}$. As such, we require more sophisticated techniques to accurately and efficiently depict the parameter space surrounding the optimum. We will now detail several such approaches to quantifying PDF uncertainties.

2.2 Markov chain Monte Carlo

Monte Carlo algorithms are a set of computational tools which rely on repeated random sampling. Steadily rooted in Bayesian statistics [97, 98, 99], Monte Carlo algorithms consist of using the available data to generate samples of the posterior $p(\boldsymbol{a}|\boldsymbol{m})$ at a discrete set of points. The expectation value and variance of any observable $\mathcal{O}$ can then be expressed as

$$E_{\text{Bayes}}\{\mathcal{O}(\boldsymbol{a})\} = \frac{1}{n}\sum_{k=1}^{n} \mathcal{O}(\boldsymbol{a}_k)\,, \tag{2.7a}$$

$$V_{\text{Bayes}}\{\mathcal{O}(\boldsymbol{a})\} = \frac{1}{n}\sum_{k=1}^{n} \left[\mathcal{O}(\boldsymbol{a}_k) - E_{\text{Bayes}}\{\mathcal{O}(\boldsymbol{a})\}\right]^2, \tag{2.7b}$$

where the parameters $\boldsymbol{a}_k$ are distributed according to the posterior $p(\boldsymbol{a}|\boldsymbol{m})$, and n is the number of parameter sets sampled according to a given algorithm.

Markov chain Monte Carlo (MCMC) methods have existed since the development of Monte Carlo algorithms and rose to prominence in the 1990s. MCMC allows one to bypass the normalization factor in Eq. (2.2) (the evidence $\mathcal{Z}$) entirely, which is usually an extremely difficult integral to calculate. MCMC algorithms make use of the Markov

property in order to perform a stochastic sampling of the posterior, meaning that any future sample will depend only on the current sample. The method has the pitfall that it can still be computationally expensive, even after eliminating the need to calculate $\mathcal{Z}$, to the point of being prohibitive, depending on various elements.

Initially proposed in 1953, an algorithm which improved upon standard MCMC methods came into the statistical realm — the Metropolis-Hastings algorithm [68, 69]. This algorithm operates on the principle that certain iterations of the MCMC are accepted or rejected based on predetermined criteria with the effect of producing more precise results. The Metropolis–Hastings algorithm proceeds as follows. At each Monte Carlo time $t-1$, the next state $\boldsymbol{a}_t$ is chosen by sampling a candidate point $\boldsymbol{a}'$ from a proposal distribution $q(\boldsymbol{a}_{t-1}|\boldsymbol{a}')$. $q(\boldsymbol{a}_{t-1}|\boldsymbol{a}')$ can be any probability distribution function, however a common choice is a Gaussian proposal distribution centred on the previous sample $\boldsymbol{a}_{t-1}$ so that candidate points are more likely to be near the previous point. The candidate point is then accepted with the probability

$$\alpha(\boldsymbol{a}_{t-1}, \boldsymbol{a}') = \min\left[\frac{p(\boldsymbol{a}'|\boldsymbol{m})\, q(\boldsymbol{a}_{t-1}|\boldsymbol{a}')}{p(\boldsymbol{a}_{t-1}|\boldsymbol{m})\, q(\boldsymbol{a}'|\boldsymbol{a}_{t-1})}\right], \tag{2.8}$$

and the Metropolis-Hastings transition kernel is thus

$$T(\boldsymbol{a}_{t-1} \to \boldsymbol{a}') = q(\boldsymbol{a}'|\boldsymbol{a}_{t-1})\, \alpha(\boldsymbol{a}_{t-1}, \boldsymbol{a}')\,. \tag{2.9}$$

If the new set of parameters $\boldsymbol{a}'$ is accepted, the next state of the chain becomes $\boldsymbol{a}_t = \boldsymbol{a}'$. If it is rejected, the chain does not move and the point at the time t is identical to the point at the previous time $t-1$: $\boldsymbol{a}_t = \boldsymbol{a}_{t-1}$.

A special case of the Metropolis-Hastings algorithm is the random walk Metropolis, for which the proposal distribution is chosen to be such that $q(\boldsymbol{a}'|\boldsymbol{a}_{t-1}) = q(|\boldsymbol{a}_{t-1} - \boldsymbol{a}'|)$. The acceptance probability then reduces to $\alpha(\boldsymbol{a}_t, \boldsymbol{a}') = \min[1, p(\boldsymbol{a}'|\boldsymbol{m})/p(\boldsymbol{a}_t|\boldsymbol{m})]$. The efficiency of a Markov chain Monte Carlo algorithm can be optimized by choosing the proposal distribution to be as close as possible to the target distribution or alternatively, in the case of a random walk, by carefully setting the size of the scale parameter [71, 100]. Optimization becomes increasingly delicate if the number of parameters in $\boldsymbol{a}$ is large, as it is in a real fit of PDFs.

2.3 Hybrid Markov chain Monte Carlo

Hamiltonian (or hybrid) Markov chain Monte Carlo algorithms [66] were originally developed in lattice field theory. These algorithms produce candidate proposals for the Metropolis-Hastings algorithm [68, 69] (see Ref. [101] for an extensive review) and can be used to sample the *target* probability density, $p(\boldsymbol{a}|\boldsymbol{m})$. Optimization problems can be circumvented in HMC algorithms by introducing for each set of parameters $\boldsymbol{a}$ a set

of conjugate momenta $\boldsymbol{b}$, and associating a Hamiltonian $H(\boldsymbol{a},\boldsymbol{b}) = \frac{1}{2}\boldsymbol{b}^T M^{-1}\boldsymbol{b} + \mathcal{U}(\boldsymbol{a})$ to this joint state of *position* $\boldsymbol{a}$ and *momentum* $\boldsymbol{b}$, where M is a mass matrix, generally taken to be diagonal, and $\mathcal{U}(\boldsymbol{a})$ an arbitrary potential energy. This allows one to define a joint distribution

$$p(\boldsymbol{a},\boldsymbol{b}) = \mathcal{N}\exp\big[-H(\boldsymbol{a},\boldsymbol{b})\big] = \mathcal{N}\exp\big[-\mathcal{K}(\boldsymbol{b})\big]\exp\big[-\mathcal{U}(\boldsymbol{a})\big]\,, \tag{2.10}$$

where $\mathcal{N}$ is the normalization constant.

A common choice for the potential energy is $\mathcal{U}(\boldsymbol{a}) = -\log[p(\boldsymbol{m}|\boldsymbol{a})\,p(\boldsymbol{a})]$. Starting from a point $\boldsymbol{a}_0$ of the chain, the HMC algorithm selects some initial momenta $\boldsymbol{b}_0$, normally distributed around zero, and proceeds to let the system evolve deterministically for some time according to Hamilton's equations of motion for $H(\boldsymbol{a},\boldsymbol{b})$. The algorithm then reaches a candidate point $(\boldsymbol{a}_1,\boldsymbol{b}_1)$, which is accepted with probability $\alpha(\boldsymbol{a}_1,\boldsymbol{b}_1) = \min(1, e^{-\Delta H})$, where ΔH is the variation of the Hamiltonian with respect to the previous point. In principle, the dynamics conserve energy, *i.e.* $\Delta H = 0$, along a trajectory, so that the acceptance rate is 100%. This acceptance is degraded by numerical errors, since Hamilton's equations are solved numerically. Such errors allow one to explore paths in the parameter space where H is not conserved, and are therefore essential to determine the sampling. The numerical noise can be efficiently exploited by appropriate methods to approximate the solution to a system of differential equations. Among these, the leap-frog method (see Sect. 2.3 in Ref. [101] for details) proved to be particularly convenient — one has to choose only the scale parameter and the number of iterations as free parameters, which are conventionally tuned, to obtain an acceptance rate of 60%.

2.4 Nested sampling

Nested sampling [58] allows for an efficient evaluation of the Bayesian evidence $\mathcal{Z}$ in Eq. (2.2), and also produces a discrete set of posterior inferences. The expectation value and variance of any observable $\mathcal{O}$ can be expressed as

$$E_{\text{Bayes}}\{\mathcal{O}(\boldsymbol{a})\} = \sum_{k=1}^{n} p(\boldsymbol{a}_k|\boldsymbol{m})\mathcal{O}(\boldsymbol{a}_k)\,, \tag{2.11a}$$

$$V_{\text{Bayes}}\{\mathcal{O}(\boldsymbol{a})\} = \sum_{k=1}^{n} p(\boldsymbol{a}_k|\boldsymbol{m})\big[\mathcal{O}(\boldsymbol{a}_k) - E\{\mathcal{O}(\boldsymbol{a})\}\big]^2\,. \tag{2.11b}$$

The key ingredient in NS is to define a *prior volume* X, such that

$$X(\lambda) = \int_{\mathcal{L}(\boldsymbol{a};\boldsymbol{m})>\lambda} \mathrm{d}\boldsymbol{a}\; p(\boldsymbol{a}), \tag{2.12}$$

where the integral extends over the region(s) of parameter space contained within the iso-likelihood contour $\mathcal{L}(\boldsymbol{a}) = \lambda$. The multi-dimensional evidence integral in Eq. (2.2) can then be transformed into a simpler one-dimensional integral,

$$\mathcal{Z} = \int_0^1 \mathrm{d}X \, \mathcal{L}(X), \tag{2.13}$$

where $\mathcal{L}(X)$, the inverse of Eq. (2.12), is a monotonically decreasing function of X. Therefore, if one can evaluate the likelihood $\mathcal{L}_i \equiv \mathcal{L}(X_i)$ over a discrete sequence of decreasing values $0 < X_{n-n_{\rm act}} < \cdots < X_i < \cdots < X_1 < X_0 = 1$, the evidence can be computed numerically using, *e.g.*, standard quadrature methods, as a weighted sum

$$\mathcal{Z} = \sum_{i=1}^{n-n_{\rm act}} \mathcal{L}_i \, w_i \, . \tag{2.14}$$

For instance, for the simple trapezium rule, the weights would be given by $w_i = \frac{1}{2}(X_{i-1} - X_{i+1})$.

The summation in Eq. (2.14) is performed with an appropriate iterative algorithm as follows. At the starting iteration ($i = 0$), a flat prior $p(\boldsymbol{a})$ is assumed so that its volume is $X_0 = 1$, from which $n_{\rm act}$ *active* or *live* samples are drawn. These samples are then ordered according to their increasing likelihood value. The sample corresponding to the smallest likelihood $\mathcal{L}_0$ is removed from the active set, hence becoming *inactive*. It is then replaced by a sample drawn from the prior, requiring that its likelihood value is $\mathcal{L} > \mathcal{L}_0$. The corresponding volume delimited by the iso-likelihood contour $\mathcal{L} > \mathcal{L}_0$ is described, at the first iteration, by a random variable $X_1 = t_1 X_0$, where t_1 is distributed according to the probability for the largest of $n_{\rm act}$ samples drawn uniformly from the interval $[0, 1]$, $\Pr(t) = n_{\rm act} t^{n_{\rm act}-1}$. The procedure is repeated for each subsequent iteration i until the entire prior volume has been traversed through nested likelihood shells; the sample associated with the lowest likelihood $\mathcal{L}_i$ in the active set is removed, then it is replaced with a new sample drawn from the prior with $\mathcal{L} > \mathcal{L}_i$, and finally the corresponding prior volume is reduced to $X_{i+1} = t_{i+1} X_i$. The algorithm stops when the evidence $\mathcal{Z}$ in Eq. (2.14) is determined to some specified precision.

To be precise, an upper limit is set from the remaining active points by assuming that the largest evidence contribution made up by the residual portion of the posterior is $\Delta\mathcal{Z}_i = \mathcal{L}_{\rm max} X_i$, where $\mathcal{L}_{\rm max}$ is the maximum likelihood in the set of active points at the last iteration. The algorithm stops when this quantity no longer changes the final evidence estimate, *i.e.*, $\Delta\mathcal{Z}_i < \epsilon$, with ϵ arbitrarily small. Finally, an increment from the set of $n_{\rm act}$ points is added to the evidence estimate, Eq. (2.14),

$$\mathcal{Z}' = \mathcal{Z} + \Delta\mathcal{Z} \, , \qquad \Delta\mathcal{Z} = \sum_{j=1}^{n_{\rm act}} \mathcal{L}_j w_{n-n_{\rm act}+j} \, , \tag{2.15}$$

where $w_{n-n_{\rm act}+j} = X_{n-n_{\rm act}}/n$ for all j. Once the evidence $\mathcal{Z}$ is found, the posterior can be determined using the full sequence of active and inactive points generated in the nested sampling process as

$$p(\boldsymbol{a}_k|\boldsymbol{m}) = \frac{\mathcal{L}_k w_k}{\mathcal{Z}'}\,, \qquad k = 1, \ldots, n\,. \tag{2.16}$$

The posterior can then be used to compute the expectation value and variance in Eq. (2.11b).

2.5 The Hessian method

The Hessian method was first developed in Refs. [39, 102], and essentially attempts to approximate the posterior distribution as a multi-variate Gaussian distribution in parameter space. To do this, one first finds the set of parameters $\boldsymbol{a}_0$ corresponding to the maximum *a posteriori* (MAP) estimate, which can be obtained by minimizing the χ^2 function assuming flat priors. In a neighborhood of $\boldsymbol{a}_0$, the $\chi^2(\boldsymbol{a})$ function can be expanded up to the quadratic term in its parameters and the posterior approximated by a multi-dimensional Gaussian function,

$$p(\boldsymbol{a}|\boldsymbol{m}) \propto \exp\Big(-\frac{1}{2}\chi^2(\boldsymbol{a}, m)\Big) \propto \exp\Big(-\frac{1}{2}\Delta\boldsymbol{a}^T\, H\, \Delta\boldsymbol{a}\Big). \tag{2.17}$$

In the last term, $\Delta\boldsymbol{a} = \boldsymbol{a} - \boldsymbol{a}_0$, the Hessian matrix elements are given by

$$H_{ij} = \frac{1}{2}\,\frac{\partial^2\chi^2(\boldsymbol{a})}{\partial a^i \partial a^j}\bigg|_{\boldsymbol{a}=\boldsymbol{a}_0}\,, \quad i, j = 1, \ldots n_{\rm par}\,. \tag{2.18}$$

and $\exp\big(-\frac{1}{2}\chi^2(\boldsymbol{a}_0)\big)$ has been absorbed into the normalization of the posterior.

The approximated posterior can be reparametrized in terms of the eigendirections of the Hessian matrix via

$$\boldsymbol{a}(\boldsymbol{t}) = \boldsymbol{a}_0 + \sum_{k=1}^{n_{\rm par}} t_k \frac{\boldsymbol{e}_k}{\sqrt{w_k}}\,, \tag{2.19}$$

where $\boldsymbol{e}_k$ and w_k are the orthonormal eigenvectors and eigenvalues of the Hessian matrix, respectively. The parametrization (2.19) defines a linear change of variables for the posterior, $p(\boldsymbol{a}|\boldsymbol{m}) \to p(\boldsymbol{t}|\boldsymbol{m})$, with t_k parametrizing straight paths in the parameter space that radiate from $\boldsymbol{a}_0$ and run parallel to the corresponding eigenvector. In the new space, the approximated posterior explicitly factorizes along each eigendirection,

$$p(\boldsymbol{t}|\boldsymbol{m}) \;=\; \prod_k p_k(t_k|\boldsymbol{m}) \;\propto\; \prod_k \exp\Big(-\frac{1}{2}t_k^2\Big)\,, \tag{2.20}$$

and allows one to estimate the expectation value of an observable in a simple way,

$$E\{\mathcal{O}(\boldsymbol{a})\} = \int \mathrm{d}^n t \; p(\boldsymbol{t}|\boldsymbol{m})\, \mathcal{O}(\boldsymbol{a}(\boldsymbol{t})) \;\approx\; \mathcal{O}(\boldsymbol{a}_0)\;. \tag{2.21}$$

The observable's variance can, in turn, be calculated as

$$\begin{aligned} V\{\mathcal{O}(\boldsymbol{a})\} &= \int \mathrm{d}^n t \; p(\boldsymbol{t}|\boldsymbol{m}) \big[\mathcal{O}(\boldsymbol{a}(\boldsymbol{t})) - E\{\mathcal{O}(\boldsymbol{a})\}\big]^2 \\ &\approx \prod_k \int \mathrm{d}t_k \;\; p\Big(t_k \frac{\boldsymbol{e}_k}{\sqrt{w_k}} \Big| \boldsymbol{m}\Big) \Big(\sum_l \frac{\partial \mathcal{O}\left(\boldsymbol{a}(\boldsymbol{t})\right)}{\partial t_l}\Big|_{\boldsymbol{a}_0} t_l \Big)^2 \\ &= \sum_k T_k^2 \left(\frac{\partial \mathcal{O}\left(\boldsymbol{a}(\boldsymbol{t})\right)}{\partial t_k}\Big|_{\boldsymbol{a}_0} \right)^2 , \end{aligned} \tag{2.22}$$

where in the second line $\mathcal{O}$ is expanded to first order in $\boldsymbol{t}$, and T_k in the third line is the *tolerance factor*, defined by

$$T_k^2 = \int \mathrm{d}t_k \; p_k(t_k|\boldsymbol{m})\, t_k^2. \tag{2.23}$$

Whenever the posterior is approximately Gaussian, one expects $T_k \approx 1$. Introducing a finite step ξ_k in each eigendirection, the variance (2.22) can then be calculated numerically as

$$V\{\mathcal{O}(\boldsymbol{a})\} \approx \sum_k \frac{T_k^2}{4\xi_k^2} \Big[\mathcal{O}\Big(\boldsymbol{a}_0 + \xi_k \frac{\boldsymbol{e}_k}{\sqrt{w_k}}\Big) - \mathcal{O}\Big(\boldsymbol{a}_0 - \xi_k \frac{\boldsymbol{e}_k}{\sqrt{w_k}}\Big)\Big]^2 . \tag{2.24}$$

The step size ξ_k can be chosen arbitrarily, as long as it is sufficiently small to guarantee the linear approximation in $\boldsymbol{t}$. However, it is convenient to set $\xi_k = T_k$ in order to write the variance as

$$V\{\mathcal{O}(\boldsymbol{a})\} \approx \sum_k \frac{1}{4} \Big[\mathcal{O}\Big(\boldsymbol{a}_0 + T_k \frac{\boldsymbol{e}_k}{\sqrt{w_k}}\Big) - \mathcal{O}\Big(\boldsymbol{a}_0 - T_k \frac{\boldsymbol{e}_k}{\sqrt{w_k}}\Big)\Big]^2 . \tag{2.25}$$

Note that the expressions (2.21) and (2.25) for the statistical estimators have been derived under three assumptions: (i) the quadratic approximation of the χ^2 function, producing a Gaussian posterior; (ii) the ensuing factorization of the posterior along each eigendirection; and (iii) the linear approximation of $\mathcal{O}(\boldsymbol{a}(\boldsymbol{t}))$ in $\boldsymbol{t}$. In practice, however, the Gaussian approximation for the posterior is not always valid in all eigendirections, and one of the common symptoms is that the value of T_k deviates from unity for eigendirections with very small eigenvalues. Such a situation can arise because of a lack of experimental constraints, even when the data set being fitted is statistically

consistent. To address this issue, we can evaluate Eq. (2.23) using the unapproximated posterior as

$$T_k^2 = \int \mathrm{d}t_k \, \frac{1}{Z_k} \exp\left[-\frac{1}{2}\chi^2\big(t_k \boldsymbol{e}_k/\sqrt{w_k}, m\big)\right] t_k^2 \,, \tag{2.26}$$

where the normalization Z_k can be efficiently estimated numerically since it only requires a one dimensional integration over the variable t_k. As long as the posterior remains approximately factorized for t_k values of $O(T_k)$, using the tolerances (2.26) instead of assuming them all equal to 1 provides a faithful representation of an observable's variance, even when the Gaussian approximation for the posterior breaks down.

In actual global QCD analyses of PDFs, one also needs to accommodate statistical inconsistencies among data sets, which cannot be accounted for by the χ^2-based likelihood (2.4). Indeed, such a likelihood is multiplicative with respect to each data set, $\mathcal{L} = \prod_{\text{set}} \mathcal{L}_{\text{set}}$, and does not properly account for a scenario in which two or more data sets are mutually incompatible. To alleviate this problem without adopting a different likelihood, a procedure known as the *tolerance criterion* has been advocated [103, 104, 105, 106], where the tolerance factors in Eq. (2.23) are inflated by demanding that a given fraction of the fitted data falls within the variance of the corresponding observable. Values $T_k \sim 5 - 10$ [43] are typically needed to account for 68% of the data in PDF global analyses.

The tolerances can also be adjusted by requiring that the χ^2 increases above the minimum by a specified $\Delta\chi^2$ value, namely by solving $\chi^2\big(T_k \boldsymbol{e}_k/\sqrt{w_k}, m\big) = \Delta\chi^2$ [107]. With $\Delta\chi^2 = 1$, one should obtain tolerances comparable to Eq. (2.26), and a comparison to the tolerance criterion is possible by setting $\Delta\chi^2 = T_k$. In our analysis we will perform fits to a statistically consistent set of mock data, and therefore will not need to invoke a tolerance criterion. Nonetheless, the tolerance will need to be explicitly calculated according to Eq. (2.26) to accommodate for genuine deviations from the Gaussian behavior and to obtain accurate variance estimates.

2.6 Data resampling

The data resampling method for PDFs was originally pioneered by the NNPDF collaboration some twenty years ago [47, 48] and has been adopted more recently by the JAM collaboration [53, 108, 109]. Data resampling allows one to approximate the Bayesian posterior distribution through the use of frequentist statistics. By contrast to the methods discussed so far, frequentist methods do not make use of Bayes' theorem at all. For a frequentist, there exists a definite true set of parameters $\boldsymbol{a}_{\text{true}}$, and the job of frequentist parameter inference is to obtain best estimates of these unknown true values. The frequentist probability of an observable taking on a certain value is simply

equal to the number of times that value is observed out of a number of repeatable trials. As such, the only information the frequentist has access to is the likelihood $\mathcal{L}(\boldsymbol{a};\boldsymbol{m})$. It is common to choose the estimate of the true parameter values to be the set of parameters that maximizes the likelihood — the so-called maximum likelihood estimate (MLE).

Various algorithms can be used to perform this maximization, however, the main point of interest here is how the uncertainty would be characterized from a frequentist perspective. In the Bayesian approach one would have access to the posterior distribution, and could therefore generate credible intervals which describe the degree of belief about a parameter taking on a certain value. The frequentist approach must instead construct confidence intervals, in which the true observables are expected to fall as a proportion of a total number of trials.

One might wonder why one would bother to introduce alternatives to the Bayesian methods described previously. The answer is that Bayesian techniques tend to require many likelihood evaluations in order to build up sufficient samples to gain a meaningful picture of the posterior, to the point of being ineffective in the context of real PDF data with many free parameters to estimate and large data sets. Since the use of frequentist statistics involves finding the MLE, one often requires far fewer evaluations of the likelihood in order to find a good approximation to the location of the maximum likelihood, depending on the maximization algorithm being used.

Data resampling makes use of the traditional method of generating confidence intervals within frequentist statistics, which involves repeat experimentation. Rather than performing a single likelihood maximization on the mean value of the data, n_{rep} pseudodata replicas of the original measurements $\boldsymbol{m}$ are generated following a multidimensional Gaussian smearing with the mean and uncertainties of the original measurements. This essentially creates n_{rep} new experimental datasets on which likelihood maximization can be performed, resulting in an equal number of replica parameter sets $\boldsymbol{a}_{\text{rep}}$. Distributions of the MLEs in the space of the parameters can then be obtained, from which we can calculate the mean and standard deviation for the corresponding observables. Since the frequentist approach does not have direct access to the posterior, Eqs. (2.1) cannot be applied. Instead, data resampling makes use of the equations

$$E_{\text{freq}}\{\mathcal{O}(\boldsymbol{a})\} = \frac{1}{n_{\text{rep}}}\sum^{n_{\text{rep}}}\mathcal{O}(\boldsymbol{a}_{\text{rep}})\,, \tag{2.27a}$$

$$V_{\text{freq}}\{\mathcal{O}(\boldsymbol{a})\} = \frac{1}{n_{\text{rep}}}\sum^{n_{\text{rep}}}\left[\mathcal{O}(\boldsymbol{a}_{\text{rep}}) - E_{\text{freq}}\{\mathcal{O}(\boldsymbol{a})\}\right]^2. \tag{2.27b}$$

Although these equations closely resemble Eqs. (2.7), the quantities that are being summed over in this case are frequentist MLEs rather than Bayesian MC samples. In JAM global analyses the interpolant is provided by functions of the form $x^\alpha(1-x)^\beta P(x)$, where $P(x)$ is a low order polynomial, while in NNPDF analyses the function

$x^\alpha(1-x)^\beta$ acts as a "preprocessing factor" designed to speed up the minimization and is weighted by $P(x) \to$ a neural network.

Importantly, the distribution of the frequentist maximum likelihood estimators can be shown to coincide with the Bayesian posterior distribution given certain assumptions (namely, flat priors and a Gaussian likelihood), which implies that $E_{\text{Bayes}} \approx E_{\text{freq}}$ and $V_{\text{Bayes}} \approx V_{\text{freq}}$. One such demonstration was performed in Ref. [73], where it was found that distributions in the replicas obtained using the NNPDF4.0 PDFs approximately reproduced the expected Bayesian posterior distribution. As such, it is not necessary to embrace the fully-fledged frequentist methodology in order to accept the data resampling approach. One can argue that the data resampling method is a trick used to calculate the posterior more efficiently, albeit with some caveats. Nevertheless, the data resampling approach can at most provide an approximation to the Bayesian posterior, and so should not be identified fundamentally as a Bayesian technique.

Chapter 3

On the determination of uncertainties in parton densities

It is evident that a reliable and faithful representation of PDF uncertainties is especially critical in applications such as searches for signals of physics beyond the SM, where subpercent precision is often sought. It is often difficult to compare uncertainties between groups using different methodologies in a meaningful and quantitative way [110, 111, 112], especially when there exist tensions between data sets, or when there are strong departures from the Gaussian approximation to probability distributions of parameters.

In this chapter, we compare the existing methods described in Chapter 2 for estimating uncertainties in global QCD analyses of PDFs. Specifically, we use a computationally inexpensive "toy" model of a PDF fit to compare the uncertainty estimates that arise from the DR, Hessian, NS and HMC methods, and explore whether they produce the same results. In each case, we assume that the functional form of the PDF being fitted is the same as that used to generate the toy data, so as not to introduce any inflated uncertainty arising from inaccuracies in the representation of the fitted functional form.

Furthermore, we investigate the case where the known functional form in the PDF fit is replaced by a neural network, and investigate how the uncertainty estimates that result from DR depend on the the cross validation procedure used in the training of the neural network. We also compare the uncertainty estimates from the neural network approach to the other methods in a similar vein to a "closure test" [113], to establish how often the true values of our toy model fall within the uncertainty bands predicted by each method across many fits.

Section 3.1 describes the toy model, which is designed to reproduce the salient features of a real PDF fit, but with a significantly reduced computation time. We then perform a systematic comparison of the results obtained by applying the Hessian, DR, NS and HMC techniques to the toy model analysis. The study of the neural network approach is performed in Sec. 3.2, where we in particular explore the effects of parti-

tioning and cross validation in the training of neural networks on PDF uncertainties. Finally, in Sec. 3.3 we present our conclusions and discuss future applications of this work. The entirety of these results is published in [114].

3.1 Comparing methodologies

In the following we perform a systematic comparison of the uncertainties that result from applying each of the methods discussed in Chapter 2. Our investigation proceeds by fitting a set of pseudodata that has been generated according to a known underlying law and statistical noise. We discuss the generation of the pseudodata in Sec. 3.1.1, and the results of our fits in Sec. 3.1.2.

3.1.1 Construction of toy data

For simplicity, we construct a set of two toy "quark" PDFs with momentum fraction x dependence parametrized by a basic functional form based on Eq. 1.15,

$$q_i(x) = N_i\, x^{\alpha_i}(1-x)^{\beta_i}, \qquad i = 1, 2. \tag{3.1}$$

These functions could, for example, represent the u and d quark PDFs in the proton, with free parameters α_i and β_i. In this case, we can assign true values to the free parameters (which we choose to be $\alpha_1 = 0.5$, $\beta_1 = 2.5$, $\alpha_2 = 0.1$, $\beta_2 = 3.0$), resulting in the functional form illustrated in Fig. 3.1. We can then model toy "cross sections" σ_j using a linear combination of the two PDFs,

$$\sigma_j = \sum_{i=1,2} c_{ji}\, q_i. \tag{3.2}$$

These could, for instance, correspond to inclusive DIS cross sections as in Sec. 1.4 for a proton and a neutron, with the coefficients c_{ij} proportional to the squares of the quark charges, in which case we would assign $c_{11} = 4c_{12} = 4c_{21} = c_{22}$.

The values of x are taken in the range $x = 0.1 - 0.9$ at regular intervals, depending on the number of toy data points we wish to generate. For each x value we calculate the corresponding q_1 and q_2 PDFs, from which the true σ_1 and σ_2 are determined. Toy cross section data points are then generated by drawing randomly from a Gaussian distribution centred on the true cross section values, and uncertainties are assigned to be 0.1 times the magnitude of each toy data point. An example of a toy cross section dataset with 10 points is shown in Fig. 3.1.

3.1.2 Fit results

We perform a fit to a single dataset using each of the methods described in Chapter 2, namely, Hessian, DR, HMC and NS. For DR we follow the traditional $x^{\alpha}(1-x)^{\beta}$

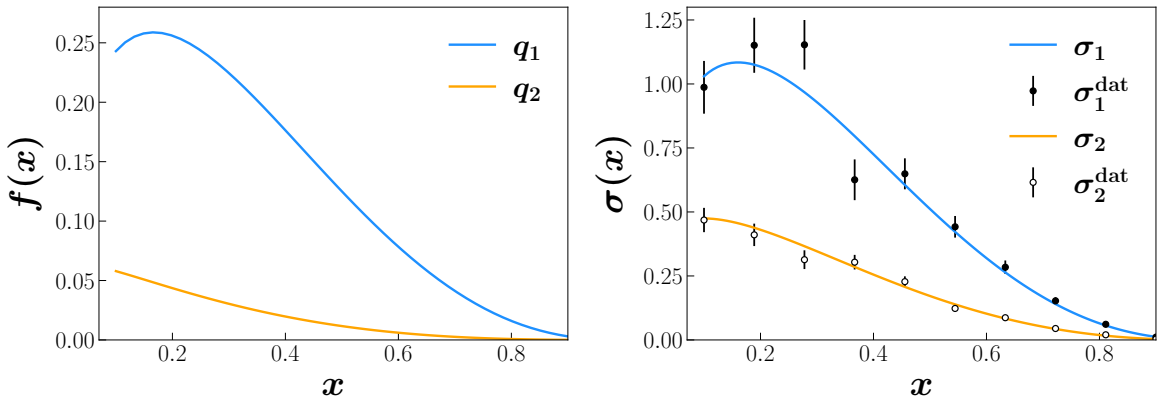


Figure 3.1: *(Left)* Quark distributions q_1 and q_2 in our toy model as a function of x. *(Right)* Example of toy "cross section" datasets σ_1 and σ_2 for the pseudodata (filled and open circles) and cross sections calculated from the input q_1 and q_2 quark distributions (blue and orange curves).

parametrization, as used by the JAM collaboration, and compare the neural network parametrization separately in Sec. 3.2.2 below. All the methods discussed here involve parameter estimation, in which we obtain a set of Monte Carlo samples of the parameters of the model that will ultimately be used to predict the values of the cross sections and the PDFs. The model with which we provide each algorithm is the same as the true model from which the data are drawn (namely, Eq. (3.1)), so that any deviation in the results from the true values is entirely due to the algorithm rather than any underlying inaccuracy in the parametrization.

For all the methods considered, bounds on the parameters were provided as $\alpha_1 \in [-1, 1]$, $\beta_1 \in [0, 5]$, $\alpha_2 \in [-1, 1]$, and $\beta_2 \in [0, 5]$, and all algorithms started in a random location in this parameter space. For the DR method, the `SciPy minimize least_squares` optimizer was used to minimize the χ^2 for each data shuffling step, generating 10^5 replicas.

For the Hessian method, the single χ^2 minimization required to find the MAP estimator was also performed using the `SciPy minimize least_squares` optimizer. The χ^2 itself is given by the sum of square residuals across both the σ_1 and σ_2 cross sections,

$$\chi^2 = \sum_{i=1,2} \sum_{j}^{n_{\text{data}}} \left(\frac{\sigma_i^{\text{data}}(x_j) - \sigma_i^{\text{model}}(x_j, \boldsymbol{p})}{\Delta\sigma_i(x_j)} \right)^2 , \tag{3.3}$$

where i is the index of the two cross sections, j is the index of the data points summing to a total number of points n_{data} in the sample, σ_i^{data} is the toy cross section data with uncertainty $\Delta\sigma_i$, and σ_i^{model} is the model cross section [Eq. (3.2)] for a given set of parameters $\boldsymbol{p}$. The Hessian is diagonalized, and the tolerances T_k are calculated

along each eigendirection according to Eq. (2.26). The resulting set of values $\{T_k\} = \{1.00, 1.00, 1.01, 1.01\}$ slightly deviates from unity for the least constrained parameter combinations, reflecting a sufficiently constraining data set for our simple toy model. On the other hand, we have observed differences of order 10% in the determination of the PDF uncertainties at large x compared to the $T_k = 1$ Gaussian approximation. In real applications with many parton flavors, the deviations from 1 can be substantial in the least constrained corners of parameter space, typically leading to an overestimate of the corresponding PDF uncertainties [107].

For HMC we used the `NUTS` (No U-Turn Sampler) algorithm [115], with the initial step size set to the default value of 0.2. Convergence is achieved when the step size averaged across all iterations approaches the initial step size, which was found to occur with a burn-in of around 10^4 points as well as 10^4 adaptive points. Finally, for the NS method we used `PyMultiNest` [116], with convergence achieved when the ratio of the largest live point contribution to the dead point integral is less than 10^{-3}, which was found to occur with 1000 live points running for around 5×10^4 iterations.

The predicted cross sections and PDFs for each method are shown in Fig. 3.2, along with the ratios of each prediction to the underlying law. The central values for all methods are essentially identical, as are the shapes and sizes of the predicted uncertainty bands. This may intuitively be expected given that the underlying law matches the parametrization utilized in each method, as well as the simplicity of the toy model used to generate the data. We note, however, that the agreement between the different analysis methods is nontrivial, since there are no analytical proofs to demonstrate that the DR or Hessian algorithms agree with Markovian methods such as HMC or NS. The reason why they are in practice equivalent is that the quantity upon which all the algorithms operate is the likelihood function. Provided that one utilizes the same function with all the methods [in this case $\mathcal{L} \sim \exp(-\chi^2/2)$], uses the same parametrization for the PDFs and, of course, analyzes the same data, the results are expected to be equivalent.

Having established that with a parametrization matching the underlying law one obtains the same uncertainty estimates independent of the choice of methodology, we are in a position to make a comparison with a neural network approach. In particular, we will explore the effect of much greater flexibility in the parametric form, and the impact of partitioning and cross validation as a stopping criterion for the underlying iterative minimization algorithm. As we discuss in the following, the resulting PDF uncertainty estimates will be at stark variance with those found in this section.

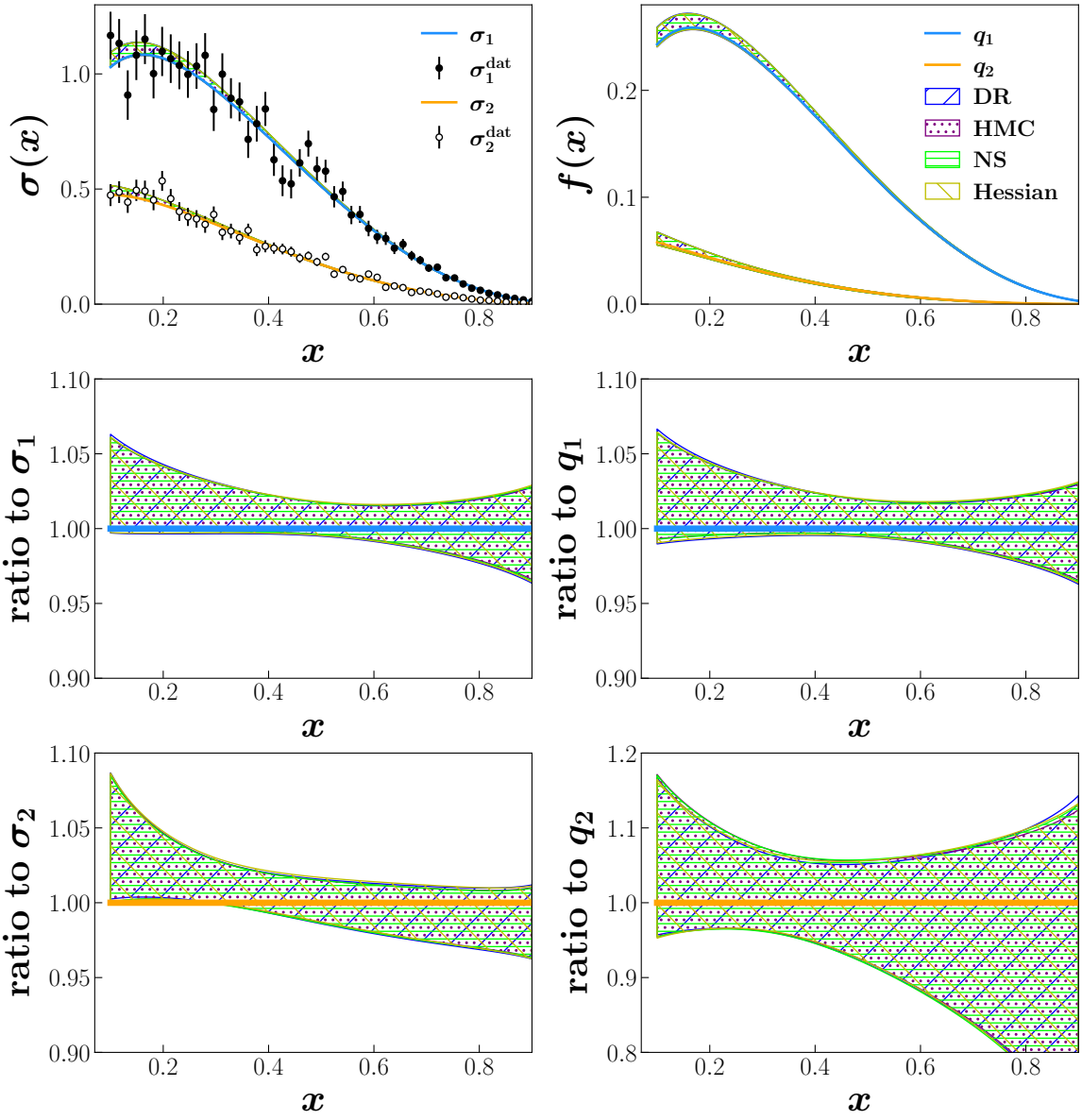


Figure 3.2: Comparison of fit results for the DR, HMC, NS and Hessian methods. The top row contains cross section and PDF distributions, with the middle and bottom rows show ratios to the true values for each distribution.

3.2 Neural networks, partitioning and cross validation

The aim of this section is to explore the effect of partitioning and cross validation on PDF uncertainties in the training of neural networks (NNs). In Sec. 3.2.1 we will describe the role that cross validation plays in fitting data with NNs, before exploring the dependence of predicted uncertainties on the partition fraction in such a framework. In Sec. 3.2.2 we will then compare the NN approach to the methods covered in Sec. 3.1 in terms of potential inflation of uncertainties.

3.2.1 Neural networks with cross validation

Cross validation is often an essential component of the training procedure for NNs if one wants to obtain robust results that generalize to new data sets beyond those used in the training. NN models typically have hundreds of parameters that can produce extremely complicated functional forms. The primary metric of a NN is the loss function, which generally depends in some way on the difference between the predictions of the model and the true values of the function that it is trying to reproduce. Training a network involves finding the values for the weight and bias parameters of the network that minimize the loss function. The danger associated with such a setup is that, given a certain set of data and enough time, an NN will simply end up reproducing the data points with which it is presented. This is referred to as *overfitting*, and reflects the fact that the NN model is no longer describing the underlying law that we ultimately want to learn and use to make predictions.

For our NN fits we make use of the `Tensorflow` platform via the `Keras` interface. The loss function is defined in the same way as the χ^2 function in Eq. (3.3), except that σ_i^{model} is replaced with the predicted value of the cross section according to the NN. Our NN architecture consists of two hidden layers with 10 neurons each, with tanh activation functions and a linear output layer, similar to that used in the NNPDF4.0 analysis [117]. For illustrative purposes, we allow the NN training to run for a long time (10^5 epochs) in order to deliberately overfit the data as a first example.

In this section we only perform NN fits to predict the cross section rather than inferring the PDFs themselves. Our goal is to hone in on the behavior of uncertainties purely due to the introduction of NNs instead of a specific underlying parametrization, and in order to study this it will be sufficiently informative to limit our analysis to the level of the cross section. We also dispense with the preprocessing factor $x^\alpha(1-x)^\beta$ at this stage, although we will demonstrate in Sec. 3.2.2 that including it has no significant effect on our results.

An example of overfitting our toy σ_1 data using an NN with the above settings can be seen in Fig. 3.3. Here, the shape predicted by the NN does not resemble the true

functional form at all, showing drastic and unrealistic variations in order to ensure that the data points match the prediction as closely as possible. Clearly this is a result we wish to avoid when applying an NN to our PDF problem.

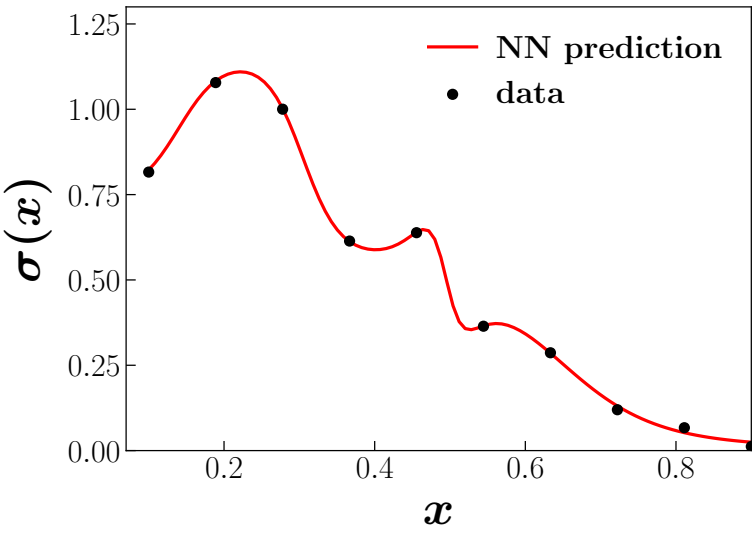


Figure 3.3: An example of an NN overfitting a single set of our toy data.

The solution to the overfitting problem is to only provide a subset of the data to the NN for it to learn, and keep the remaining data for comparison. The proportion of the data given to the training set is called the partition fraction, f, and the remaining fraction $1 - f$ forms the validation set. If the loss function for the validation set does not improve during the learning process, this would be an indication that the NN is overfitting the data and the procedure should then stop. The exact cross validation procedure used in our analysis is slightly different from this traditional early stopping criterion. Instead of employing a stopping criterion, we allow the fit to run for a specific number of epochs (2000) and then make predictions based on the model that minimized the validation loss across all epochs, see Fig. 3.4 for an example (here we would use the model obtained at around 600 epochs). In this way we ensure that the best fit is found, while also avoiding overfitting the training data.

As explained in Sec. 2.6, repeated fitting of resampled data is necessary in order to construct uncertainty bands. By comparing the predicted σ_1 distribution both with and without cross validation, as in Fig. 3.5 (for this exercise $f = 0.4$), the importance of the cross validation procedure becomes clear. Without using cross validation, the NN model simply reproduces the error bars of the data points given, and predicts an oscillatory shape that follows the data. The NN with cross validation produces an overall smooth distribution that approximates much more closely the underlying law that we are trying to learn.

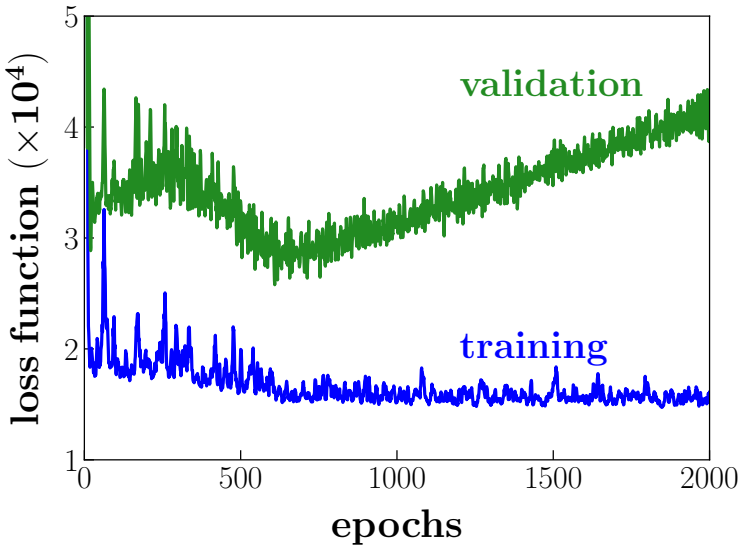


Figure 3.4: Comparison of training (blue) and validation (green) set loss functions for an overfitting example.

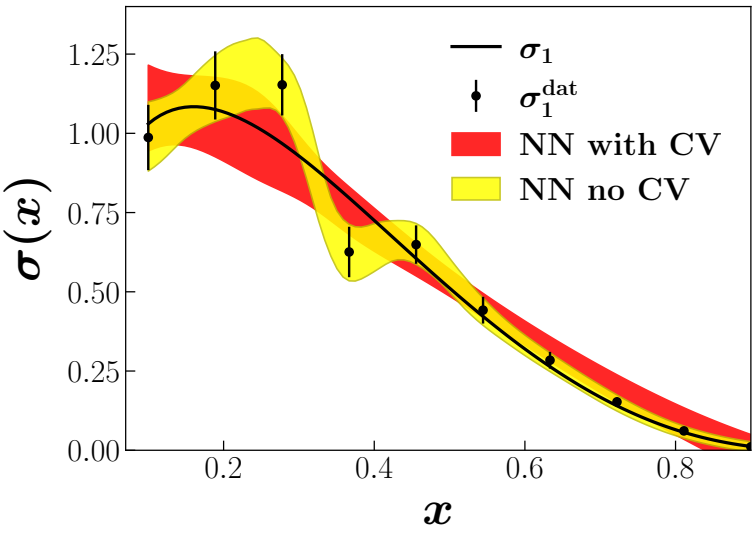


Figure 3.5: Comparison of the σ_1 replica set for the NN fit with (red band) and without (yellow band) cross validation.

One of the few aspects of the cross validation procedure which one can tune is the partition fraction, f, defining the proportion of the data on which the NN is trained. It is interesting to consider what effect a variation in f might have on the eventual uncertainty, particularly below $f = 0.5$, where one might expect very little to change.

In order to study the dependence on f, we performed several full replica fits on the same set of toy data, varying the partition fraction between 0.1 and 0.9. For 50 toy data points, we find that although the mean values of σ_1 are similar, the uncertainty is affected significantly. This can be seen in Fig. 3.6, where the standard deviation is shown for different values of f. We find that there is substantial variation in the uncertainty depending on what value of f is chosen, particularly when the partition fraction is small. In the range from $f = 0.4$ to 0.7, there are only slight differences between the uncertainties, but these are enhanced for the $f = 0.1$, 0.2, 0.3, 0.8 and 0.9 curves, being up to twice as large as for $f = 0.5$ for certain x values.

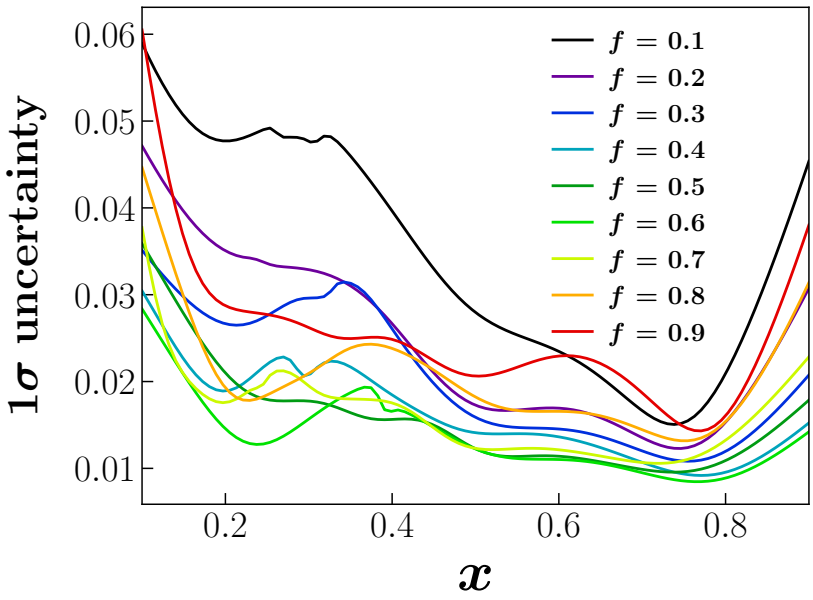


Figure 3.6: Variation of the 1σ uncertainty with x across several f values from 0.1 to 0.9 for a single dataset of 50 points.

Another important question is whether the dependence of an NN fit on f diminishes as the number of data points is increased. In Fig. 3.7 we increase the number of data points up to 500, plotting the total uncertainty for each value of f as a ratio to the $f = 0.5$ line. This allows us to visualize the underlying trend, given that the total uncertainty for all values of f should decrease as the number of points increases. The enhancement for the $f = 0.1$, 0.2 and 0.9 uncertainty seen in Fig. 3.6 remains relatively constant as one increases the number of data points, becoming slightly more pronounced. At the same time, the $f = 0.4 - 0.7$ uncertainties appear consistent for all numbers of data points, with slight enhancements in total uncertainty for $f = 0.3$ and $f = 0.8$. On the strength of these observations, we cannot rule out the importance of f with respect to uncertainty quantification in the NN method, nor can we conclude

that below $f = 0.5$ an increased dataset will remove any potential artificial inflation of uncertainty due to the choice of f.

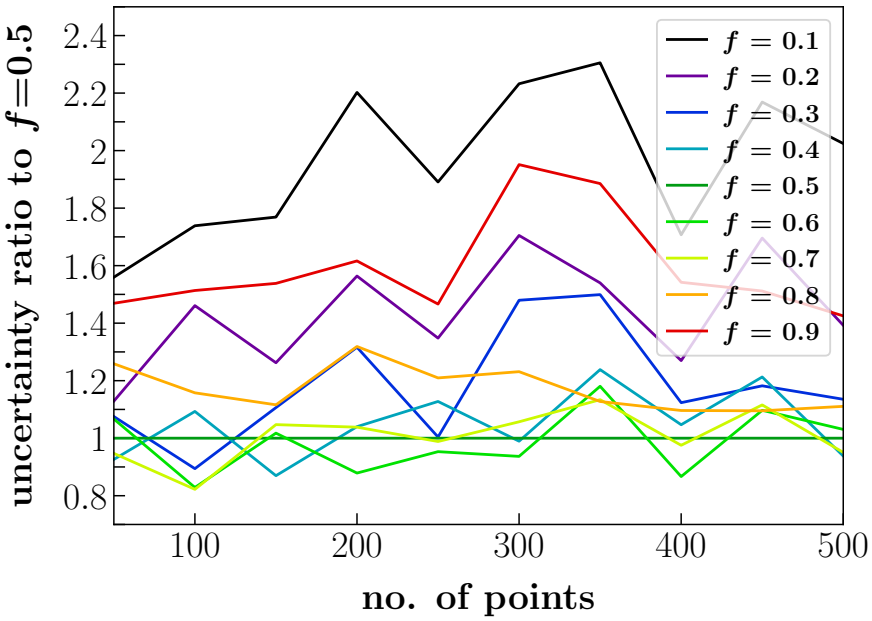


Figure 3.7: Total uncertainty ratio to $f = 0.5$ for different partition fractions as a function of number of data points.

3.2.2 Comparing NN methods to parametric methods

To further comprehend the differences between the NN methodology and that of traditional parametric modelling, it is worth comparing the results of fits to our toy data for all of the methods discussed so far. At this point we also include the preprocessing factor $x^{\alpha}(1-x)^{\beta}$ in the NN fits. This is usually added with the aim of speeding up the minimization procedure. However, it is worth testing whether its inclusion could be playing a greater role in helping the NN to obtain a shape that is more consistent with the underlying law. We apply this factor in the same way as the NNPDF collaboration; specifically, we parametrize the PDFs as

$$q_i(x) = x^{\alpha_i}(1-x)^{\beta_i}\,\mathrm{NN}_i(x), \qquad i = 1, 2, \tag{3.4}$$

where $\mathrm{NN}_i(x)$ represents the non-linear functional form to be described by the trained neural network. We then define effective asymptotic exponents as

$$\alpha_i(x) = \frac{\ln q_i(x)}{\ln(1/x)}, \qquad \beta_i(x) = \frac{\ln q_i(x)}{\ln(1-x)}. \tag{3.5}$$

The ranges for each of the α_i and β_i values are obtained through repeated fitting by taking the envelope of twice the 68% confidence interval for each x value until the preprocessing exponents no longer change. For each replica fit in the actual analysis, the values of α_i and β_i are chosen randomly according to uniform distributions in those ranges.

A full fit for one instance of our toy σ_1 data with 50 points is shown in Fig. 3.8. Here, we compare the distributions and ratio to the true values for the NN ($f = 0.6$) with and without preprocessing, as well as parametric DR (recall that parametric DR was shown to produce the same results as the Hessian, HMC and NS). Both NN methods produce predicted uncertainty bands that are much wider than that of parametric DR, particularly at large x. The fact that there is little difference between the NN predictions with and without preprocessing suggests that preprocessing does not play a significant role in this example.

Such a demonstration only becomes meaningful in the context of a full replica fit for many initial data sets rather than just one. By counting how often a given method predicts an uncertainty band that contains the true underlying law (as inspired by closure tests), we can determine whether the predicted uncertainties are statistically valid. Since the uncertainty bands represent the variance of the cross section around its expected value, we do not expect them to always contain the true values. Rather, a reasonable method will produce uncertainty bands that contain the true values around 68% of the time across a large number of initial data sets.

In Fig. 3.9 we determine the percentage of uncertainty bands across all data sets that contain the true values for all x, corresponding to each method. We plot the NN

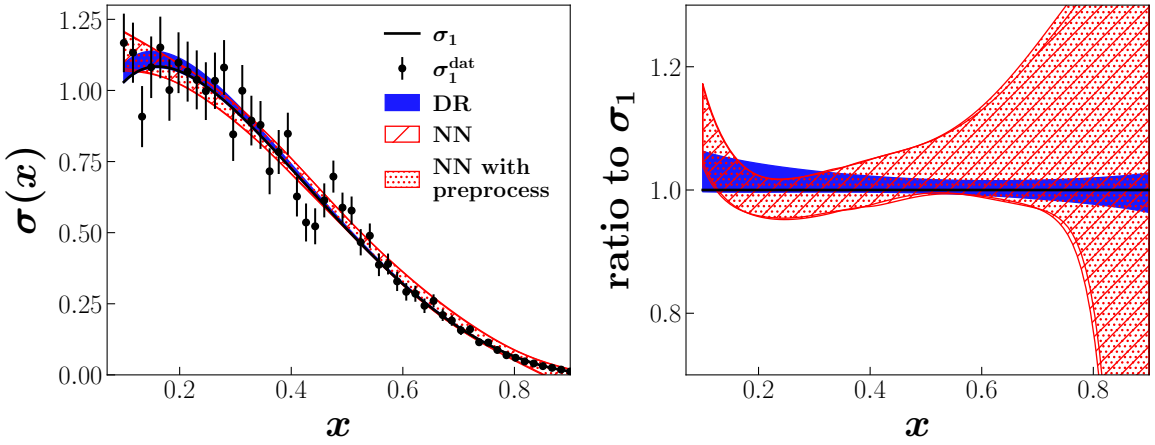


Figure 3.8: Comparison of a single σ_1 replica set between NN and parametric DR methods (left panel), and the ratio of each predicted point to the true underlying law (right panel).

method with $f = 0.2$ and $f = 0.6$ to again demonstrate the importance of the partition fraction, and also show the results for the NN with preprocessing. We find that the NN predicted uncertainty bands contain the true value too often, particularly at high x values. This effect is exacerbated by reducing the partition fraction. By contrast, the parametric DR method consistently hovers around the 68% value that would be expected from accurate confidence intervals.

The inflated uncertainty observed is not entirely due to the usage of neural networks in fitting the toy data. Another contributing factor to this result is the presence of the cross validation procedure itself. In Fig. 3.9 we also include a cross validation method in the context of parametric DR, dividing the data into a training and validation set with $f = 0.6$ for each replica and cutting off the χ^2 minimization when there is no further improvement for the validation set. Comparing DR with and without cross validation, we can clearly see a substantial increase in the width of the uncertainty bands when cross validation is used, up to a similar mean value as that of NNs at the same partition fraction.

These results imply that, for our specific simplified example of PDF data, the uncertainty bands predicted by a NN fit do not realize the 1σ confidence interval that one would naively expect. This may be due in part to a misrepresentation of the underlying law, but also a result of the cross validation procedure that forms an integral part of any NN fit to experimental data.

One may ask how is it that NN fits, as well as partitioning and cross validation applied to simple parametric DR, can so dramatically change the variance of the fitted observable? From our perspective, the algorithms deployed in either case effectively lead to a change in the likelihood function. In the case of cross validation applied to DR,

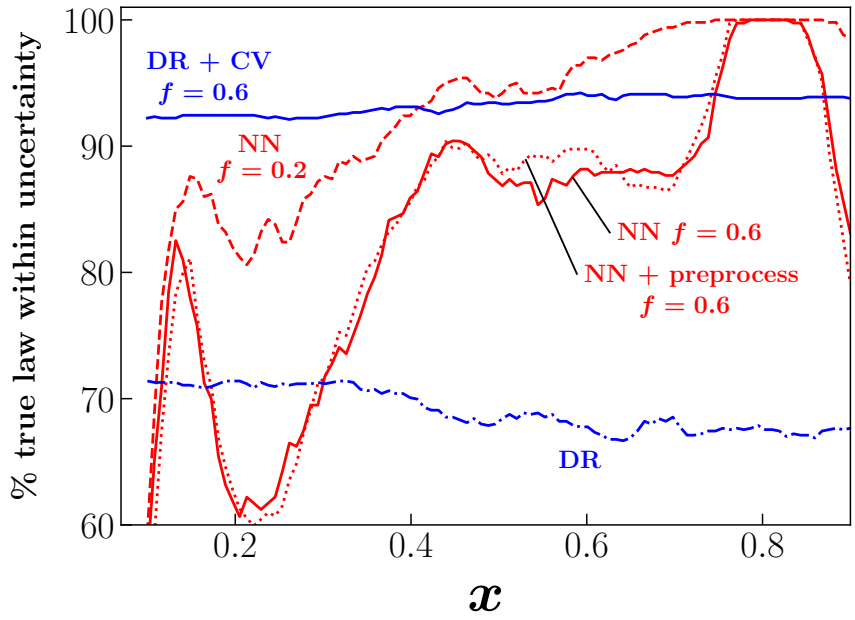


Figure 3.9: Study of confidence interval coverage for the NN (red lines) and parametric DR (blue lines) methods as a function of x.

this can be heuristically understood by viewing the validation χ^2 as a potential that stochastically pulls the parameters away from the true minimum, thereby increasing the variance of any calculated observable, but in a uniform way. This is illustrated with the very similar shapes for the DR and DR with CV results in Fig. 3.9, where one goes from an $\sim 68\%$ confidence level to an $\sim 90\%$ when using partitioning and cross validation, as if the likelihood was rescaled by the adopted stopping criterion. NN methods not only appear to rescale the likelihood, but they also do so to different degrees across the x range, resulting in a behavior that is not immediately amenable to a heuristic analysis.

3.3 Conclusions and outlook

The need for reliable uncertainty quantification in global QCD analysis motivated this study to perform a systematic comparison of different methods for estimating uncertainties on PDFs and similar correlation functions. The test laboratory employed here, in the form of the toy PDF model, although exceedingly simple, allowed us to rigorously test and verify the generation of uncertainties in a controlled context, in which the underlying physical law is known. We were thus able to demonstrate that methods which use a parametrization that matches the underlying law lead to the same

uncertainty estimates — regardless of whether the uncertainties are determined using the traditional Hessian or data resampling methods, or explicitly Bayesian techniques such as nested sampling or HMC. While the result may be intuitively expected, it is by no means trivial to verify in practice that all the approaches are indeed exploring the same likelihood function.

We then focused on the impact that partitioning and cross validation have on uncertainty quantification. We found that these added elements effectively lead to a rescaling, and even a deformation of the likelihood function, which also depends on the partition fraction. Our results suggest that this is inherent in the methodology rather than in the details of the model or the data set analyzed.

Comparing the parametric methods with NN approaches, we find indeed that NNs produce wider uncertainty bands than expected, whether or not one uses a preprocessing factor, and that the dependence on the partition fraction is not reduced as one increases the number of data points to be fitted. The enhanced uncertainty is found to be due, at least in part, to the use of cross validation. While DR+CV exhibits this behavior in a uniform way in observable space, heuristically due to the additional influence that the validation data set has on the fitting of the training data set, the NN's inflation of the cross section uncertainty depends also on x and becomes very large as $x \to 1$. Whether this non-uniform uncertainty inflation is due to the highly flexible PDF parametrizations provided by the NNs or to other elements in the analysis algorithm remains an interesting question for future investigation.

Overall, given that the data set being analyzed by the NN and parametric methods is the same, we can only conclude that NN methods (as well as parametric methods supplemented by cross validation) algorithmically deform the nominal likelihood utilized in the analysis in ways that are not necessarily under control. In view of these results it is natural to ask in what sense are PDF fits based on NN methods compatible with those based on Hessian + tolerance methods, since the likelihood functions that they utilize are technically different. From our perspective, the tolerance criterion effectively is a change in the likelihood function, which can also be cast as a rescaling of the experimental uncertainties to remove potential tensions present among the data. NN-based analyses posit instead that the use of a tolerance criterion is not necessary [47, 48, 49, 50], but the roughly equivalent size of the uncertainties they obtain may simply be coincidental and due to the likelihood deformation and resulting uncertainty inflation that we already observed when analyzing compatible data sets in our toy model. We thus echo the concern expressed in [112] that a meta-analysis, such as PDF4LHC [111], that combines existing PDFs from different groups may not only obscure the fundamental connection between experimental data and theory, but also obfuscate the true meaning of the uncertainties, since these seem to ultimately originate from a different choice of the likelihood function.

Finally, we make a note of caution about the generality of our results, since the

conclusions should be qualified by the fact that we are using a rather simplistic model. While this has been necessary in order to keep the analysis tractable and allow unambiguous conclusions to be drawn from the methodology comparisons, there may be new features present in higher dimensional analyses of more complex data sets, such as in real global QCD analyses.

Chapter 4

High-x Polarized DIS Analysis

The study of the spin structure of the proton has matured significantly in the past decade, with major progress witnessed on several fronts. Ever more precise experimental data from modern accelerator facilities, combined with theoretical developments, including insights from lattice QCD simulations and improvements in analysis methodologies, have furthered our understanding of the decomposition of the proton's spin into its quark and gluon (or parton) constituents over a large range of momentum fractions x carried by the partons. Recently the need for simultaneously analyzing parton distribution functions (PDFs) and fragmentation functions (FFs) [109, 118, 119] has been highlighted with inclusion of semi-inclusive deep-inelastic scattering (DIS) data, which depend on both types of nonperturbative functions, as well as consistently determining unpolarized and polarized PDFs under the same set of conditions and approximations [120]. Such analyses are naturally more complicated than conventional, single-purpose fits, requiring greater computing power and more sophisticated algorithms to provide robust extractions of the various distributions while minimizing reliance on theoretical assumptions.

An illustrative example of this has been the attempts to extract the strange quark helicity distribution, Δs, utilizing semi-inclusive π and K production data, which demonstrated that in the absence of assumptions about SU(3) flavor symmetry, the polarized strange content of the proton is consistent with zero [118, 119]. Removing theoretical constraints due to SU(3) flavor symmetry and assumptions about PDF positivity beyond leading order further revealed that the world data, including the important jet production data from polarized pp collisions at RHIC [121, 122, 123, 124, 125], were not sufficient to determine the sign of the gluon polarization, Δg [126, 127].

In this chapter we present a comprehensive global QCD analysis of helicity-dependent PDFs, combining all available data from inclusive DIS on protons, deuterons, and ^{3}He, semi-inclusive production of pion, kaons, and other hadrons from protons and deuterons, and inclusive W boson and jet production in singly and doubly polarized pp collisions, together with a simultaneous determination of unpolarized PDFs and

FFs [128]. For the DIS data, we study the dependence of the fit results on the cut in the hadron final state invariant mass, W, to determine the extent to which high-x data can provide stable results for the leading twist distributions. This requires taking into account target mass effects and other power corrections to the leading twist formulas, which we determine phenomenologically from the data. The resulting set of PDFs, which we refer to by the moniker `JAM25-polPDF`, are constrained over the range $5 \times 10^{-3} \lesssim x \lesssim 0.7$ for scales down to the charm quark mass, $Q^2 = m_c^2$. We also demonstrate that only when combining the world data with constraints from lattice QCD simulations of pseudo-Ioffe time distributions sensitive to Δg, together with high precision large-x data from Jefferson Lab, does it become possible to eliminate the possibility of negative gluon polarization from the conventional positive solutions in a model independent manner. This result was published in [129].

In the following sections, we begin by reviewing the formulas for observables used in this analysis, including inclusive and semi-inclusive DIS, W-boson and jet production in polarized proton-proton collisions, as well as lattice QCD simulations of pseudo-Ioffe time distributions sensitive to the gluon polarization, Δg. In Sec. 4.2 we outline our methodology, including a discussion of the parametrizations of the polarized PDFs and our treatment of subleading power corrections needed to account for higher twist effects at low Q^2. In Sec. 4.3.1 we present of the results of our global QCD analysis, performing a detailed comparison of data versus theory to assess the quality of the fit. To establish the stability of the fitted results, we also compare the χ^2 values as a function of the cut on the invariant mass of the final state hadrons, from $W^2 = 3.5$ GeV^2 to 10 GeV^2. The results for the helicity distributions are presented in Sec. 4.3.2, with an additional Sec. 4.3.3 dedicated to the sign of the Δg PDF. Residual higher twist contributions to the g_1 and g_2 structure functions are discussed in Sec. 4.3.4, which allow the most robust determination of the d_2 matrix element and the Burkhardt-Cottingham sum rule available to date. Finally, in Sec. 4.4 we summarize our results and discuss future improvements expected in the theoretical description and experimental data from future facilities. For completeness, in Appendix A we outline the derivation of the target mass corrections in collinear factorization. In Appendix B we present the detailed data versus theory comparisons for all the polarized scattering observables used to constrain our analysis.

4.1 Observables

In our analysis we consider a set of observables that can be described using collinear factorization in perturbative QCD. For each observable we utilize short-distance partonic cross sections computed at next-to-leading order (NLO) accuracy, as we describe in the following.

4.1.1 Polarized inclusive DIS

The mechanics of unpolarized DIS were introduced in Sec. 1.4, however there are a number of significant differences in the polarized case. The classical observables measured in polarized lepton DIS from a target T, $\boldsymbol{\ell}\, \boldsymbol{T} \to \ell\, X$, where T represents a proton p, deuteron d, or ^{3}He nucleus, are the longitudinal and transverse double-spin asymmetries [53],

$$A_{\parallel} = \frac{\sigma^{\uparrow\Uparrow} - \sigma^{\downarrow\Uparrow}}{\sigma^{\downarrow\Uparrow} + \sigma^{\uparrow\Uparrow}}, \tag{4.1a}$$

$$A_{\perp} = \frac{\sigma^{\uparrow\Rightarrow} - \sigma^{\downarrow\Rightarrow}}{\sigma^{\downarrow\Rightarrow} + \sigma^{\uparrow\Rightarrow}}, \tag{4.1b}$$

where the superscript $\uparrow$ ($\downarrow$) on the cross section σ denotes the spin of the lepton along (opposite to) the beam direction, $\Uparrow$ denotes the spin of the target nucleon along the beam direction, and $\Rightarrow$ represents the spin of the target perpendicular to the beam direction.

The cross sections are functions of the standard DIS variables x, y and γ^2, defined as

$$x = \frac{Q^2}{2P\cdot q}, \qquad y = \frac{P\cdot q}{P\cdot k}, \qquad \gamma^2 = \frac{4M^2x^2}{Q^2}, \tag{4.2}$$

where k, P and q are the four-momenta of the incident lepton, target nucleon (of mass M), and exchanged virtual photon, respectively, with $Q^2 \equiv -q^2$. In terms of these variables, the longitudinal and transverse spin asymmetries can be written as

$$A_{\parallel} = D\,(A_1 + \eta A_2), \tag{4.3a}$$

$$A_{\perp} = d\,(A_2 - \zeta A_1), \tag{4.3b}$$

where the kinematic variables are given by

$$D = \frac{y\,(2-y)(2+\gamma^2 y)}{2(1+\gamma)^2y^2 + \big[4(1-y) - \gamma^2y^2\big](1+R)}, \tag{4.4a}$$

$$d = \frac{\sqrt{4(1-y) - \gamma^2y^2}}{2-y} D, \tag{4.4b}$$

$$\eta = \gamma \frac{4(1-y) - \gamma^2y^2}{(2-y)(2+\gamma^2y)}, \tag{4.4c}$$

$$\zeta = \gamma \frac{2-y}{2+\gamma^2y}. \tag{4.4d}$$

The virtual photoproduction asymmetries A_1 and A_2 can be written in terms of ratios

of the spin-dependent g_1 and g_2 and spin-averaged F_1 structure functions,

$$A_1 = \frac{(g_1 - \gamma^2 g_2)}{F_1}, \tag{4.5a}$$

$$A_2 = \frac{\gamma(g_1 + g_2)}{F_1}. \tag{4.5b}$$

At large Q^2 the A_1 asymmetry is dominated by the g_1 structure function, while the A_2 asymmetry depends on the sum $g_1 + g_2$. Measurements of A_2 are therefore crucial for determining the g_2 structure function. In Eq. (4.4a) R is the ratio of longitudinal to transverse photoproduction cross sections, and is related to the unpolarized F_1 and F_2 structure functions by

$$R = \frac{\rho^2 F_2 - 2xF_1}{2xF_1}, \tag{4.6}$$

where

$$\rho^2 = 1 + \frac{4M^2x^2}{Q^2}. \tag{4.7}$$

Each of the polarized and unpolarized structure functions is a function of two variables, typically taken to be the Bjorken variable x and four-momentum transfer squared Q^2.

Applying QCD collinear factorization, the spin-dependent g_1 structure function can be written, up to $\mathcal{O}(M^2/Q^2)$ power corrections, in terms of the leading twist spin-dependent quark (Δq) and gluon (Δg) PDFs,

$$g_1^{\rm LT}(x,Q^2) = \frac{1}{2}\sum_q e_q^2 \big[\Delta C_q^{\rm DIS} \otimes \Delta q^+ + 2\Delta C_g^{\rm DIS} \otimes \Delta g\big](x,Q^2), \tag{4.8}$$

where $\Delta q^+ \equiv \Delta q + \Delta\bar{q}$, and the sum runs over all quark flavors q. Here the symbol $\otimes$ denotes the convolution integral,

$$\big[\Delta C \otimes \Delta f\big](x,Q^2) \equiv \int_x^1 \frac{{\rm d}\xi}{\xi}\, \Delta C(\xi,Q^2)\, \Delta f\Big(\frac{x}{\xi},Q^2\Big), \tag{4.9}$$

and we have chosen the renormalization and factorization scales to be $\mu^2 = Q^2$ for the DIS process. The perturbatively calculable polarized DIS hard scattering coefficients, $\Delta C_f^{\rm DIS}$ ($f = q,\, g$), are expanded to NLO in the strong coupling constant α_s,

$$\Delta C_f^{\rm DIS} = \Delta C_f^{\rm DIS(0)} + \frac{\alpha_s}{4\pi}\Delta C_f^{\rm DIS(1)} + \mathcal{O}(\alpha_s^2), \tag{4.10}$$

with the coefficients taken from Ref. [37]. Note that at leading order in α_s the Bjorken scaling variable x in Eq. (4.2) coincides with the parton momentum fraction, while

at NLO and beyond, however, these are different. At leading twist, the g_2 structure function can be related to g_1 through the Wandzura-Wilczek relation [130]

$$g_2^{\rm LT}(x,Q^2) = -g_1^{\rm LT}(x,Q^2) + \int_x^1 \frac{{\rm d}z}{z} g_1^{\rm LT}(z,Q^2)\,. \tag{4.11}$$

For analysis of DIS data at relatively small values of $Q^2 \gtrsim m_c^2$ and invariant mass squared of the unobserved hadronic system X,

$$W^2 = (P+q)^2 = M^2 + \frac{(1-x)}{x} Q^2\,, \tag{4.12}$$

power corrections of the type $\mathcal{O}(M^2/Q^2)$ may not be negligible for some lower energy datasets, and typically a cut in W^2 needs to be made to ensure that the theoretical framework is able to accommodate the data. It is well known that one can extend the reach of a perturbative QCD description of DIS data by taking into account the kinematical target mass corrections (TMCs) [131, 53, 132], which can be applied to the polarized structure functions to incorporate M^2/Q^2 corrections. These corrections scale with the Nachtmann variable $x_{\rm N}$, where [133, 134]

$$x_{\rm N} = \frac{2x}{1+\rho}\,. \tag{4.13}$$

In the collinear factorization framework, the g_1 and g_2 structure function including TMCs can be written as

$$g_1^{\rm TMC}(x,Q^2) = \frac{1}{\rho^2} g_1^{\rm LT}(x_{\rm N},Q^2) + \frac{2(\rho-1)}{\rho^2} \int_{x_{\rm N}}^1 \frac{{\rm d}z}{z} g_1^{\rm LT}(z,Q^2)\,, \tag{4.14a}$$

$$g_2^{\rm TMC}(x,Q^2) = -\frac{1}{\rho^2} g_1^{\rm LT}(x_{\rm N},Q^2) + \frac{2}{(1+\rho)\rho^2} \int_{x_{\rm N}}^1 \frac{{\rm d}z}{z} g_1^{\rm LT}(z,Q^2)\,. \tag{4.14b}$$

In the limit $M^2/Q^2 \to 0$, since $\rho \to 1$ and $x_{\rm N} \to x$, one recovers the leading twist results in Eqs. (4.8) and (4.11), $g_{1,2}^{\rm TMC}(x,Q^2) \to g_{1,2}^{\rm LT}(x,Q^2)$. Note that the lower limit of the integral from the Wandzura-Wilczek relation Eq. (4.11) is now $x_{\rm N}$ rather than x.

Alternative TMC prescriptions have also been used in the literature [130, 135, 136, 137, 138], the most common of which is that based on the operator product expansion (OPE), along the lines of the unpolarized TMC prescription of Georgi and Politzer [139]. The OPE prescription was used in the early JAM spin PDF analysis of inclusive DIS at low Q^2 [53]. For consistency with the application of collinear factorization across other processes considered, such as those considered in the present analysis, we will utilize the TMC results in Eqs. (4.14).

One immediate consequence of Eqs. (4.14), as in other TMC approaches, is that for nonzero M^2/Q^2, as $x \to 1$ one has $x_{\rm N} < 1$, and thus a nonzero structure function at

$x = 1$. The fact that the TMC structure functions do not vanish as $x \to 1$ is known as the "threshold problem" [139, 140, 141] and has been discussed in the literature [142, 143, 132, 144]. This problem is only relevant in the nucleon resonance region, well below values of W^2 where perturbative QCD is applicable, and so in practice is not relevant for this analysis.

From phenomenological analyses it is also well established [131, 53, 145, 146, 147] that in the region where TMCs play a role, namely at low Q^2 and high x, other power corrections, such as those associated with higher twists, or multi-parton correlations, also need to be considered. These are implemented in the standard manner by adding a phenomenological higher twist (HT) term. For the g_1 structure function the HT contribution is power suppressed relative to the leading term, Eq. (4.14a), while for the g_2 structure function the HT correction to the LT term (4.14b) is of the same order as the LT [148, 149],

$$g_1^{\rm HT} = \frac{c_1^{\rm HT}}{Q^2}\,, \qquad\qquad g_2^{\rm HT} = c_2^{\rm HT}, \tag{4.15}$$

with the shapes of the x-dependent functions $c_{1,2}^{\rm HT} \equiv c_{1,2}^{\rm HT}(x)$ fitted to experimental data. The total spin-dependent structure functions in our framework are given by the sum of the leading twist and higher twist contributions,

$$g_i = g_i^{\rm TMC} + g_i^{\rm HT}\,, \qquad i = 1, 2. \tag{4.16}$$

In our numerical analysis both the parameter sets describing the LT and HT functions will be extracted from data.

Integrating the structure functions over x, we can define their lowest moments as

$$\Gamma_i(Q^2) \;=\; \int_0^1 {\rm d}x\, g_i(x, Q^2), \qquad i = 1, 2. \tag{4.17}$$

For the g_1 structure function of the proton, the lowest moment can be written in terms of axial vector charges at NLO as

$$\Gamma_1^{\rm LT}(Q^2) \;=\; \left(1 - \frac{\alpha_s(Q^2)}{\pi}\right)\left(\frac{1}{12}\, g_A + \frac{1}{36}\, a_8 + \frac{1}{9}\,\Delta\Sigma\right), \tag{4.18}$$

where

$$g_A \;=\; \int_0^1 {\rm d}x\, [\Delta u^+ - \Delta d^+](x, Q^2), \tag{4.19a}$$

$$a_8 \;=\; \int_0^1 {\rm d}x\, [\Delta u^+ + \Delta d^+ - 2\Delta s^+](x, Q^2), \tag{4.19b}$$

$$\Delta\Sigma \;=\; \int_0^1 {\rm d}x \sum_q \Delta q^+(x, Q^2), \tag{4.19c}$$

are the quark triple, octet, and singlet axial charges, respectively. Note that both g_A and a_8 are scale invariant, while $\Delta\Sigma = \Delta\Sigma(Q^2)$ explicitly depends on the scale Q^2.

For the lowest moment of the g_2 structure function, in the LT approximation $\Gamma_2^{\rm LT}$ satisfies the Burkhardt-Cottingham (BC) sum rule [150]

$$\Gamma_2^{\rm LT} \;=\; 0. \tag{4.20}$$

Phenomenologically, violation of the BC sum rule would suggest the presence of higher twist contributions to g_2. An additional combination of the spin structure functions which directly reveals higher twist contributions is the d_2 matrix element, defined in terms of the x^2-weighted moments of g_1 and g_2,

$$d_2 \;=\; \int_0^1 {\rm d}x\, x^2 \left[2g_1 + 3g_2\right](x, Q^2). \tag{4.21}$$

From the Wandzura-Wilczek relation (4.11), one can immediately see that the leading twist part of this combination vanishes, $d_2^{\rm LT} = 0$, so that nonzero values can only arise from higher twist or power correction contributions. These contributions can be related to matrix elements describing the nucleon's color polarizability [151, 152, 153] or the transverse color force [154] acting on quarks.

4.1.2 Polarized semi-inclusive DIS

In polarized semi-inclusive DIS (SIDIS), $\boldsymbol{\ell T} \to \ell\, h\, X$, where h is a hadron identified in coincidence with the scattered lepton, the usual observable measured is the longitudinal double-spin asymmetry, A_1^h, which at large Q^2 is given by

$$A_1^h(x, z, Q^2) = \frac{g_1^h(x, z, Q^2)}{F_1^h(x, z, Q^2)}, \tag{4.22}$$

where g_1^h and F_1^h are the polarized and unpolarized semi-inclusive structure functions, respectively [155]. In addition to the inclusive variables x and Q^2, the SIDIS asymmetry also depends on the fraction of the virtual photon's energy carried by the hadron h, defined as

$$z = \frac{P \cdot p_h}{P \cdot q}, \tag{4.23}$$

where p_h is the four-momentum of the produced hadron. In terms of these variables the invariant mass squared of the unobserved hadronic final state in the SIDIS process is given by

$$\begin{aligned} W^2_{\rm SIDIS} &= (P + q - p_h)^2 \\ &= W^2 + m_h^2 - \frac{z}{x}Q^2 - 2p_h \cdot q \\ &= M^2 + m_h^2 + \frac{Q^2(1 - x - z)}{x} + \frac{Q^4 z}{2M^2 x^2}\big(\rho\, \rho_h - 1\big), \end{aligned} \tag{4.24}$$

where in analogy with Eq. (4.7) we define ρ_h by

$$\rho_h = \sqrt{1 - \frac{4M^2x^2\, m_{hT}^2}{z^2Q^4}}, \tag{4.25}$$

with m_{hT} the transverse mass of the produced hadron h,

$$m_{hT}^2 = m_h^2 + p_{hT}^2. \tag{4.26}$$

Integrating over the transverse momentum of the produced hadron, p_{hT}, and expanding in powers of M^2/Q^2 and m_h^2/Q^2, the polarized SIDIS structure function g_1^h can be written in terms of the helicity PDFs Δq, Δg and single-hadron FFs D_f^h,

$$\begin{aligned} g_1^h(x,z,Q^2) \;=\; & \frac{1}{2}\sum_q e_q^2 \big[\Delta C_{qq}^{\rm SIDIS} \otimes \Delta q \otimes D_q^h + \Delta C_{gq}^{\rm SIDIS} \otimes \Delta q \otimes D_g^h \\ & + \Delta C_{qg}^{\rm SIDIS} \otimes \Delta g \otimes D_q^h\big](x,z,Q^2), \end{aligned} \tag{4.27}$$

where the sum q runs over all quarks and antiquarks. Generalizing the convolution in Eq. (4.9), we define the double convolution in Eq. (4.27) by

$$\big[\Delta C \otimes \Delta f \otimes D_{f'}^h\big](x,z) \equiv \int_x^1 \frac{{\rm d}\xi}{\xi} \int_z^1 \frac{{\rm d}\zeta}{\zeta} \Delta C(\xi,\zeta)\, \Delta f\Big(\frac{x}{\xi}\Big)\, D_{f'}^h\Big(\frac{z}{\zeta}\Big). \tag{4.28}$$

The perturbatively calculable polarized SIDIS hard scattering coefficients, $\Delta C_{ff'}^{\rm SIDIS}$ with $ff' = qq,\, gq,\, qg$, are expanded to NLO in the strong coupling constant,

$$\Delta C_{ff'}^{\rm SIDIS} = \Delta C_{ff'}^{\rm SIDIS(0)} + \frac{\alpha_s}{4\pi} \Delta C_{ff'}^{\rm SIDIS(1)} + \mathcal{O}(\alpha_s^2), \tag{4.29}$$

with the coefficients taken from Ref. [156]. The renormalization and factorization scales are taken to be $\mu_R = \mu_F = Q$ for the SIDIS process. For the neutron SIDIS structure function the same proton PDFs are used except with the switch $u \leftrightarrow d$ that is derived from isospin symmetry. The unpolarized SIDIS structure function F_1^h is also given by Eq. (4.27) with the spin-dependent PDFs and hard coefficients replaced by their spin-averaged counterparts.

Since the SIDIS asymmetry and structure function in Eqs. (4.22) and (4.27) are evaluated in the leading twist approximation, we exclude from our analysis of SIDIS data the regions of kinematics where higher twist and hadron mass corrections may contaminate the leading power description [157, 158, 159]. We found that this can be achieved by applying cuts on the invariant mass of the unobserved hadrons in the SIDIS final state, $W_{\rm SIDIS}^2 > 10$ GeV^2, and energy fraction, $z < 0.7$ [160]. A dedicated analysis of SIDIS data in the low-Q^2 region, which includes these corrections, will be given elsewhere [161].

4.1.3 W production in polarized hadron collisions

The longitudinal single-spin asymmetry for W-boson production in proton-proton collisions is defined in terms of the helicity-dependent cross sections as

$$A_L^W = \frac{\mathrm{d}\sigma^+ - \mathrm{d}\sigma^-}{\mathrm{d}\sigma^+ + \mathrm{d}\sigma^-} \equiv \frac{\mathrm{d}\Delta\sigma}{\mathrm{d}\sigma}, \tag{4.30}$$

where $\mathrm{d}\sigma^\pm$ is the W production cross section for scattering a proton with helicity ± 1 from an unpolarized proton. The W boson decays into a positron (or electron) and an antineutrino (or neutrino), which is unobserved, $\boldsymbol{p}\,p \to \ell X$. In collinear factorization the cross section difference in the numerator of Eq. (4.30) can be written in terms of the spin-dependent PDFs Δf_a and spin-averaged PDFs f_b in the polarized and unpolarized protons, respectively, and perturbative hard-scattering partonic cross sections $\mathrm{d}\Delta\hat{\sigma}_{ab}$,

$$\mathrm{d}\Delta\sigma = \sum_{a,b} \int \mathrm{d}x_a \mathrm{d}x_b \, \Delta f_a(x_a, \mu_F) \, f_b(x_b, \mu_F) \, \mathrm{d}\Delta\hat{\sigma}_{ab}(x_a P_A, x_b P_B, p_\ell, \mu_R, \mu_F), \tag{4.31}$$

where x_a and x_b are the momentum fractions of the two partons from the colliding protons, P_A and P_B are the 4-momenta of the polarized and unpolarized protons, p_ℓ is the 4-momentum of the outgoing lepton, and the sum runs over quarks, antiquarks and the gluon. In this analysis we choose the factorization and renormalization scales, μ_F and μ_R, respectively, such that $\mu_R = \mu_F = p_{\ell,T} \equiv p_T$, the transverse momentum of the charged lepton; however, in practice the W-lepton asymmetry in Eq. (4.30) has very little sensitivity to the scales [162].

The experimental measurements of the asymmetry are typically given in terms of p_T and the lepton rapidity η, defined as $\eta \equiv \eta_\ell = \frac{1}{2} \ln \left[(p_\ell^0 + p_{\ell,z}) / (p_\ell^0 - p_{\ell,z}) \right]$, where p_ℓ^0 and $p_{\ell,z}$ are the energy and longitudinal momentum of the lepton. Introducing the Mandelstam variables

$$s \equiv (P_A + P_B)^2 \,, \qquad t \equiv (P_A - p_\ell)^2 \,, \qquad u \equiv (P_B - p_\ell)^2 \,, \tag{4.32}$$

we define the auxiliary variables

$$v \equiv 1 + \frac{t}{s}, \qquad w \equiv \frac{-u}{s+t} \tag{4.33}$$

along with their partonic analogues,

$$\hat{s} \equiv (p_a + p_b)^2, \qquad \hat{t} \equiv (p_a - p_\ell)^2, \qquad \hat{u} \equiv (p_b - p_\ell)^2, \tag{4.34}$$

and

$$\hat{v} \equiv 1 + \frac{\hat{t}}{\hat{s}} \,, \qquad \hat{w} \equiv \frac{-\hat{u}}{\hat{s}+\hat{t}} \,, \tag{4.35}$$

where the partonic momenta are given by $p_a = x_a P_a$ and $p_b = x_b P_b$. In terms of the auxiliary hadronic and partonic variables, one can write the partonic momentum fractions as

$$x_a = \frac{vw}{\hat{v}\hat{w}}, \qquad x_b = \frac{1-v}{1-\hat{v}}, \tag{4.36}$$

and write the hadronic variables v and w in terms of the original p_T and η,

$$v = 1 - \frac{p_T}{\sqrt{s}} e^{-\eta}, \qquad vw = \frac{p_T}{\sqrt{s}} e^{\eta}. \tag{4.37}$$

With these relations between the hadronic and leptonic variables, one can now expressed the cross section in Eq. (4.31) in terms of the lepton kinematics to first order in the strong coupling constant α_s [162],

$$\frac{\mathrm{d}^2\Delta\sigma}{\mathrm{d}p_T \mathrm{d}\eta} = \frac{2}{p_T} \sum_{a,b} \int_{vw}^{v} \mathrm{d}\hat{v} \int_{vw/\hat{v}}^{1} \mathrm{d}\hat{w}\; x_a \Delta f_a(x_a, \mu_F)\; x_b f_b(x_b, \mu_F) \times \left[\frac{\mathrm{d}\Delta\hat{\sigma}^{(0)}_{ab}(\hat{s}, \hat{v})}{\mathrm{d}\hat{v}} \delta(1-\hat{w}) + \frac{\alpha_s(\mu_R)}{2\pi} \frac{\mathrm{d}\Delta\hat{\sigma}^{(1)}_{ab}(\hat{s}, \hat{v}, \hat{w}, \mu_R, \mu_F)}{\mathrm{d}\hat{v}\mathrm{d}\hat{w}} \right], \tag{4.38}$$

where $\mathrm{d}\Delta\hat{\sigma}^{(0)}_{ab}$ and $\mathrm{d}\Delta\hat{\sigma}^{(1)}_{ab}$ are the lowest order and NLO polarized partonic W production cross sections. Similar expressions can be obtained for the unpolarized cross section in the denominator of Eq. (4.30). Integrating over a specified p_T range, Eq. (4.38) can be used to relate the experimental values of the asymmetry A_L^W given in terms of η to the polarized and unpolarized PDFs. The detailed expressions for the partonic cross sections are given shifts are taken from Ringer and Vogelsang [162].

4.1.4 Jet production in polarized hadron collisions

The longitudinal double-spin asymmetry for jet production in polarized proton-proton collisions, $\boldsymbol{p}\,\boldsymbol{p} \to \text{jet} + \text{X}$, is defined as

$$A_{LL}^{\text{jet}} \equiv \frac{\mathrm{d}\sigma^{++} - \mathrm{d}\sigma^{+-}}{\mathrm{d}\sigma^{++} + \mathrm{d}\sigma^{+-}} \equiv \frac{\mathrm{d}\Delta\sigma}{\mathrm{d}\sigma}, \tag{4.39}$$

where $\mathrm{d}\sigma^{+\pm}$ is the cross section for scattering a proton with helicity + from a proton with helicity $\pm$. In analogy with the W-lepton production cross section in Eq. (4.38), one can write the polarized differential cross section at high p_T in terms of the jet

kinematics as [163]

$$\frac{\mathrm{d}^2\Delta\sigma}{\mathrm{d}p_T\mathrm{d}\eta} = \frac{2p_T}{s}\sum_{a,b}\int_{vw}^{v}\frac{\mathrm{d}\hat{v}}{\hat{v}(1-\hat{v})}\int_{vw/\hat{v}}^{1}\frac{\mathrm{d}\hat{w}}{\hat{w}}\;\Delta f_a(x_a,\mu_F)\;\Delta f_b(x_b,\mu_F) \times\left[\frac{\mathrm{d}\Delta\hat{\sigma}^{(0)}_{ab}(\hat{s},\hat{v})}{\mathrm{d}\hat{v}}\delta(1-\hat{w})+\frac{\alpha_s(\mu_R)}{\pi}\frac{\mathrm{d}\Delta\hat{\sigma}^{(1)}_{ab}(\hat{s},\hat{v},\hat{w},\mu_R,\mu_F;R)}{\mathrm{d}\hat{v}\mathrm{d}\hat{w}}\right], \tag{4.40}$$

where p_T and η now refer to the transverse momentum and rapidity of the produced jet, and s is the invariant mass squared of the colliding protons. The hadronic variables v and w, and partonic variables $\hat{s}$, $\hat{v}$, and $\hat{w}$ are as defined in Eqs. (4.33) and (4.35), with the lepton momentum replaced by the jet momentum, and R refers to the radius of the jet. The renormalization and factorization scales are again set to $\mu_R = \mu_F = p_T$.

An analogous expression can be obtained for the unpolarized cross section in the denominator of the asymmetry A_{LL}^{jet} in Eq. (4.39). The expressions for the perturbatively calculated hard-scattering partonic cross sections are taken from Jäger *et al* [163].

4.1.5 Lattice QCD data

In addition to the experimental cross sections and single- and double-spin asymmetries fitted in the global QCD analysis, we also consider the recent high-precision lattice QCD calculations of matrix elements of nonlocal operators that are related to Ioffe time pseudo-distributions (pseudo-ITDs) [164]. Pseudo-ITDs are Lorentz invariant amplitudes that can be matched to the PDFs in the $\overline{\text{MS}}$ scheme when the invariant separation between the field operators, z^2, is sufficiently small [165]. Specifically, the gluon and quark singlet Ioffe time helicity distributions are defined as

$$\mathcal{I}_{\Delta g}(\nu,\mu^2) = \int_0^1 \mathrm{d}x\, x\sin(x\nu)\,\Delta g(x,\mu^2), \tag{4.41a}$$

$$\mathcal{I}_{\Delta\Sigma}(\nu,\mu^2) = \int_0^1 \mathrm{d}x\, x\sin(x\nu)\,\Delta\Sigma(x,\mu^2), \tag{4.41b}$$

where $\nu = p\cdot z$ is the Ioffe time, and μ is the renormalization scale. The gluon pseudo-ITD is a particularly useful observable, as it has direct sensitivity to Δg at leading order in α_s.

Following Karpie *et al.* [127], we include lattice QCD data generated on 1901 configurations of an ensemble with (2+1)-dynamical clover Wilson fermions with stout-link smearing and tree-level tadpole-improved gauge action with a lattice volume $32^3\times 64$. The pion mass for these configurations is $m_\pi = 358(3)$ MeV, with lattice spacing $a = 0.096(1)$ fm determined using the w_0 scale [166]. Uncertainties introduced by

extrapolations to the physical pion mass are expected to be small compared with systematic uncertainties from other sources [167]. In the global fit the lattice QCD data are treated on the same footing as experimental data, as in recent global analyses that have included lattice constraints [64, 168, 169, 170, 171, 127, 172, 173].

4.2 Methodology

Our Bayesian analysis consists of approximating the posterior distribution through data resampling, as described in Sec. 2.6. The likelihood is defined as in Eq. 2.4, with a "reduced" χ^2 similar to Eq. 2.1,

$$\chi^2_{\rm red} \equiv \frac{1}{N_{\rm dat}} \sum_{e,i} \left(\frac{d_{e,i} - {\rm E}\big[\sum_k r_{e,k}\beta^k_{e,i} + T_{e,i}/N_e\big]}{\alpha_{e,i}} \right)^2, \tag{4.42}$$

where $N_{\rm dat}$ is the number of data points under consideration and ${\rm E}[\cdots]$ represents the mean theory as defined in Eq. (2.27b). The $r_{e,k}$ weights are optimized per values of the paramerers via $\partial\chi^2/\partial r_{e,k} = -2r_{e,k}$, which can be solved in a closed form. The prior $\pi(\boldsymbol{a})$ for each replica is given by

$$\pi(\boldsymbol{a}) = \prod_l \Theta\big((a_l - a_l^{\rm min})(a_l^{\rm max} - a_l)\big) \tag{4.43}$$

$$\times \prod_e \prod_k \exp\left\{-\frac{1}{2} r^2_{e,k}\right\} \prod_e \exp\left\{-\frac{1}{2}\left(\frac{1-N_e}{\delta N_e}\right)^2\right\}, \tag{4.44}$$

where $\prod_l$ is a product over the parameters. The step function $\Theta(X)$ forces the parameters to be within the chosen range between $a_l^{\rm min}$ and $a_l^{\rm max}$. A Gaussian penalty controlled by the experimentally quoted normalization uncertainties δN_e is applied when fitting N_e.

Our PDF extraction procedure is based on Bayesian inference using the Monte Carlo techniques developed in previous JAM analyses [53, 108, 109, 118, 119]. This method incorporates the JAM multi-step strategy, whereby data are sequentially added one step at a time, as an efficient way of locating the minimum χ^2 for a given shuffling of the data. The parameterization of the unpolarized PDFs follows the standard form in Eq. 1.15, while for the polarized PDFs and FFs at the input scale $\mu_0 = m_c$ we use the form

$$f(x,\mu_0^2) = \frac{N}{\mathcal{M}}\, x^\alpha (1-x)^\beta (1 + \gamma\sqrt{x} + \eta x), \tag{4.45}$$

where $\boldsymbol{a} = \{N, \alpha, \beta, \gamma, \eta\}$ is the set of parameters to be inferred, and $\mathcal{M} = {\rm B}[\alpha+1, \beta+1] + \gamma{\rm B}[\alpha+\frac{3}{2}, \beta+1] + \eta{\rm B}[\alpha+2, \beta+1]$ normalizes the function to the first moment to

maximally decorrelate the normalization and shape parameters. To characterize the nucleon valence region and discriminate it from the sea components, we parametrize the light-quark and strange PDFs according to

$$\begin{aligned} \Delta u &= \Delta u_v + \Delta \bar{u}, \qquad & \Delta d &= \Delta d_v + \Delta \bar{d}, \\ \Delta \bar{u} &= \Delta S_1 + \Delta \bar{u}_0, \qquad & \Delta \bar{d} &= \Delta S_1 + \Delta \bar{d}_0, \\ \Delta s &= \Delta \bar{s} = \Delta S_1, \end{aligned} \tag{4.46}$$

where the dependence on x and the scale μ_0^2 has been suppressed for convenience. The input quark distributions Δu_v, Δd_v, $\Delta \bar{u}_0$, and $\Delta \bar{d}_0$, as well as the gluon distribution Δg, are parametrized individually as in Eq. (4.45). For the sea quark PDFs, the shared additional function ΔS_1 is also parametrized via Eq. (4.45), and is designed to describe the small x region where it is assumed the sea quarks are all equal. We choose not to parametrize the strange PDFs with a Δs_0, having found no difference in the quality of the fit with or without these extra parameters. Since the data are not sensitive to such an additional strange PDF parametrization, we find it sufficient to set Δs_0 and $\Delta \bar{s}_0$ equal to the sea quark PDFs ΔS_1.

The PDFs are evolved at next-to-leading logarithmic accuracy in the zero-mass variable flavor number scheme with splitting kernels evaluated at $\mathcal{O}(\alpha_s^2)$., and the heavy quark mass thresholds for the evolution of the PDFs and α_s chosen from the PDG values [174] $m_c = 1.28$ GeV and $m_b = 4.18$ GeV in the MS-bar scheme. Along with the DLGAP evolution, we solve the RGE for the strong coupling numerically with the beta function evaluated at two loops with the boundary condition $\alpha_s(M_Z) = 0.118$. The boundary conditions for the RGE are parameterized at the scale μ_0^2 (chosen to be m_c^2 in this analysis) and inferred from data.

Our Monte Carlo analysis consists of a sample of $\approx$ 1000 replicas, utilizing 146 parameters for the leading-twist PDFs and FFs, along with 81 higher twist and off-shell parameters, making a total of 227 free parameters. We set our baseline as the most recent JAM22 analysis of helicity dependent PDFs [120], which performed a simultaneous fit of unpolarized and polarized PDFs along with FFs while employing a $W^2 > 10$ GeV^2 cut on polarized DIS data. The only additional data we will be including in our new analyses beyond this baseline analysis are high-x polarized DIS and LQCD data, which does not affect the unpolarized PDFs or FFs. We therefore choose only to update the helicity PDF parameters, leaving us with only 24 free parameters and holding all other parameters fixed to their baseline value. To adequately describe the data with a $W^2 > 4$ GeV^2 cut, we directly parametrize twist-3 and higher effects by adding $1/Q^2$-dependent functions to g_1 and g_2 consisting of interpolating splines between 5 points, with each point corresponding to an additional free parameter. Separate contributions are fitted for the proton and neutron, giving 4 higher twist functions with 5 parameters each, for a total of 20 extra new parameters. Including the helicity PDF parameters, that gives us 44 total free parameters.

4.3 Global QCD Analysis

In this section we present the results of our global QCD analysis. We begin with reviewing in Sec. 4.3.1 the datasets that are fitted, and the stability of fit, with the resulting χ^2 values for the various high-energy scattering observables. The extracted PDFs are shown in Sec. 4.3.2, with a larger discussion on the Δg PDFs reserved for Sec. 4.3.3 and the details of higher twist corrections and sum rules given in Sec. 4.3.4.

4.3.1 Quality and stability of the fit

In the current analysis we include measurements of the longitudinal DIS double-spin asymmetries $A_\parallel$ and A_1 from fixed target experiments on proton, deuterium, and ^{3}He nuclei from EMC [175], SMC [176, 177], COMPASS [178, 179, 180], SLAC [181, 182, 183, 184, 185, 186], HERMES [187, 188], and Jefferson Lab [189, 190, 191, 192, 193]. We also include measurements of the transverse DIS asymmetries $A_\perp$ and A_2 from SLAC [182, 184, 183, 194, 195], HERMES [196], and Jefferson Lab [189, 197]. For all DIS asymmetries we apply cuts on the four-momentum transfer squared of $Q^2 > m_c^2$ and $W^2 > 4$ GeV^2 on the hadronic final state mass.

For polarized SIDIS, we include measurements of the longitudinal semi-inclusive A_1^h asymmetry on proton and deuterium targets from HERMES [198] and COMPASS [199, 200] for pions, kaons, and unidentified charged hadrons. The cuts $Q^2 > m_c^2$ and $W^2_{\rm SIDIS} > 10$ GeV^2 are made to ensure the applicability of the leading power formalism and enhance separation of the current and target fragmentation regions, and an additional cut on the fragmentation variable $0.2 < z < 0.8$ is applied to avoid large-z threshold corrections [119].

Beyond lepton-hadron reactions, hadron-hadron scattering data include jet production in polarized pp collisions from the STAR [201, 202, 203, 121, 122, 123] and PHENIX [125] Collaborations at RHIC. Specifically, we include the single spin asymmetries A_L^W from STAR [204] and $A_L^{W/Z}$ from PHENIX [205, 206].

Datasets that depend only on unpolarized PDFs are the same as those in the earlier JAM PDF analysis in Ref. [54]. For the single-inclusive annihilation (SIA) data for e^+e^- collisions with production of pions, kaons, and unidentified hadrons, we use the same datasets as in Ref. [119]. The kinematics for all the polarized datasets included in this analysis are displayed in Fig. 4.1, with the scale Q^2 (four-momentum transfer squared for DIS and SIDIS; transverse momentum squared for jet production; vector boson mass squared for W/Z production; e^+e^- invariant mass squared for SIA) plotted versus the relevant scaling variable for each type of dataset (Bjorken x or Feynman x, or the fragmentation variable z).

Also illustrated are the $W^2 = 4$ GeV^2 and 10 GeV^2 contours indicating the region of kinematics that would be excluded for these cuts on the DIS data. The analysis with a $W^2 > 4$ GeV^2 cut on DIS data, for example, would include ≈ 1000 additional data

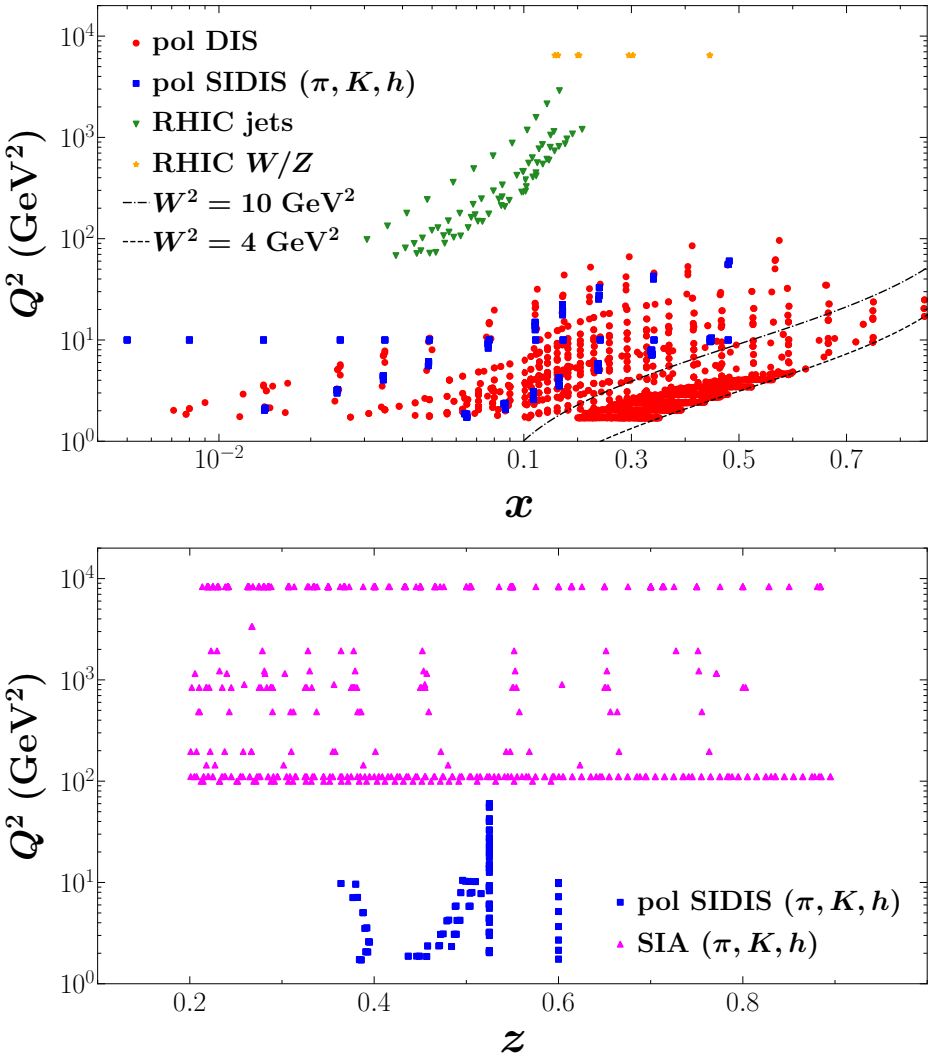


Figure 4.1: Kinematics of polarized datasets included in this analysis with scale Q^2 displayed versus the relevant scaling variable. *(Top panel)* For polarized DIS and SIDIS the four-momentum transfer squared Q^2 is versus the Bjorken scaling variable; for jet production the transverse momentum squared is versus the Feynman variable; and for vector boson production the mass squared of the boson is versus x_F. For DIS, cuts on $W^2 > 4$ GeV^2 (dashed line) and $W^2 > 10$ GeV^2 (dot-dashed line) also shown for reference. *(Bottom panel)* The scale Q^2 is shown versus the fragmentation variable $z = P \cdot p_h / P \cdot q$ for polarized SIDIS, and versus $z = 2\, p_h \cdot q / Q^2$ for single-inclusive annihilation (SIA) of e^+e^- pairs into π, K and unidentified hadrons h.

points than one with a stronger, $W^2 > 10$ GeV^2 cut. To determine the optimal cut,

we performed several independent analyses of the global data (including lattice QCD data) using different DIS data cuts, ranging from the most stringent $W^2 > 10$ GeV^2 cut down to the most relaxed cut of $W^2 > 3.5$ GeV^2.

The resulting reduced global χ^2 values are listed in Table 4.1 for various cut values. The worsening of the fit quality as the cut is lowered is driven primarily by the large number of points with small uncertainties for the JLab Hall B proton data [192], many of which are contained in this kinematic region (see also Fig. B.2 and Table B.1 in Appendix B). While the most stringent cut allows a reduced χ^2 value for the DIS data of ≈ 1, this contains only about 1/3 of the data points accessible with a $W^2 > 3.5$ GeV^2 cut, which, however, gives a reduced χ^2 of ≈ 1.2. As a compromise between these two extreme values for the cuts, for our main global analysis we use a cut of $W^2 > 4$ GeV^2 for all inclusive DIS data.

The stability of the fit is further illustrated in Fig. 4.2, where we plot various moments of the PDFs versus the W^2 cut for $W^2_{\rm min} = 3.5$, 4, 5, 6 and 10 GeV^2. The lowest PDF moments $\Delta f \equiv \int_0^1 {\rm d}x\, \Delta f(x, Q^2)$ for $\Delta f = \Delta u^+, \Delta d^+, \Delta s^+$ and Δg are shown, together with the combinations g_A, a_8, and $\Delta\Sigma$ corresponding to the axial vector charges defined in Eqs. (4.19). Also shown are the d_2 values for the proton and neutron for different $W^2_{\rm min}$. Both g_A and $\Delta\Sigma$ are also stable with respect to variation in the W^2 cut. For the d_2 integral for the proton and neutron, the values are also consistent across the range within the larger error bars plotted, especially for the proton. For both the proton and neutron d_2 value, the result is approximately consistent with zero in each case.

Table 4.1: Dependence of the quality of the global fit (including lattice QCD data) on the minimum value of W^2 used for the inclusive DIS data. The quoted reduced χ^2 values and number of data points $N_{\rm dat}$ for each cut refer only to the inclusive polarized DIS data.

$W^2_{\rm min}$ (GeV^2)	3.5	**4**	5	6	10
$N_{\rm dat}$	2002	**1735**	1287	1008	689
$\chi^2_{\rm red}$	1.17	**1.02**	1.01	0.99	0.99

For the preferred value of the W^2 cut, $W^2_{\rm min} = 4$ GeV^2, a summary of reduced χ^2 values along with Z-scores for all of the types of reactions and observables fitted in this analysis is given in Table 4.2, including polarized DIS, SIDIS, jet, and vector boson production data, along with the lattice QCD data on pseudo Ioffe-time distributions. Shown also are the overall $\chi^2_{\rm red}$ values for the unpolarized lepton and hadron scattering datasets which constrain the unpolarized PDFs, and the e^+e^- SIA data that provide the strongest constraints on the fragmentation functions used in this analysis.

In particular, for the longitudinal DIS polarization asymmetries from various global

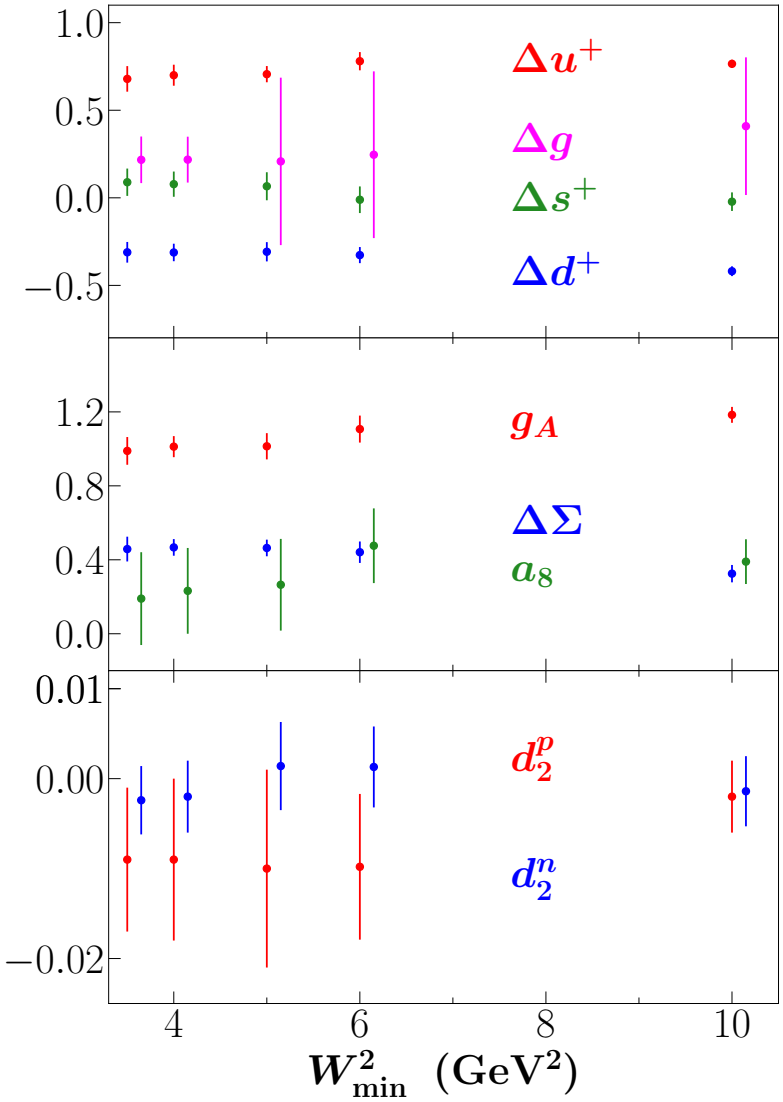


Figure 4.2: Stability of truncated moments with respect to the cut used for the polarized inclusive DIS data, from $W^2_{\rm min} = 3.5$ GeV^2 to 10 GeV^2, at $Q^2 = 10$ GeV^2: (top row) lowest moments of PDFs Δu^+, Δd^+, Δs^+, and Δg; (middle row) triplet g_A, octet a_8, and singlet $\Delta\Sigma$ axial charges; (bottom row) proton and neutron d_2 moments.

experiments, which form about 60% of all the polarized data, we find $\chi^2_{\rm red} = 1.01$, while for the transverse DIS polarization asymmetries, which constitute $\approx 1/4$ of the polarized DIS data, we have $\chi^2_{\rm red} = 1.06$. For the longitudinally polarized SIDIS hadron production asymmetries, the reduced χ^2 value is $\chi^2_{\rm red} = 1.36$, and for the jet and W/Z boson production $\chi^2_{\rm red} = 0.78$. Overall, the total $\chi^2_{\rm red} = 1.02$ for 2070 polarized experimental data points. Tables of reduced χ^2 values for each individual

polarized dataset are displayed in Tables B.1–B.4 in Appendix B. Comparisons of all the polarized data versus theory are shown in Figs. B.1–B.9 in Appendix B. The 48 lattice data points are well fitted with $\chi^2_{\rm red} = 0.58$. For the unpolarized lepton-hadron and hadron-hadron scattering data, which determine the spin-averaged PDFs (5954 data points), and e^+e^- SIA data (564 data points) that constrain the parton to hadron fragmentation functions, the replicas were obtained from the earlier JAM analysis in Ref. [54], giving $\chi^2_{\rm red} = 1.14$ and 0.87, respectively. The combined global χ^2 for the chosen $W^2 = 4$ GeV^2 cut across all unpolarized and polarized experimental and lattice data is $\chi^2_{\rm red} = 1.10$ for $N_{\rm data} = 8633$ data points.

Table 4.2: Summary of $\chi^2_{\rm red}$ values and Z-scores for all reactions included in the global fit and the number of data points $N_{\rm dat}$ fitted. The asterisks on the unpolarized and SIA data indicate that these replicas were obtained from the earlier JAM analysis in Ref. [54].The daggers in the $N_{\rm dat}$ column indicate that those datasets have fewer points for the $W^2 > 10$ GeV^2 cut - specifically, $N_{\rm dat} = 365$ for Polarized DIS, $N_{\rm dat} = 700$ for total Polarized data, and $N_{\rm dat} = 7215$ for total of all data.

		$\chi^2_{\rm red}$ (Z-score)		
Reaction	$N_{\rm dat}$	$W^2 > 10$ GeV^2	$W^2 > 4$ GeV^2	$W^2 > 4$ GeV^2 + HT
Polarized	2070†	0.89 (-2.12)	1.04 (1.22)	1.02 (0.74)
— DIS	1735†	0.95 (-0.66)	1.04 (1.04)	1.02 (0.57)
— SIDIS	231	0.85 (-1.72)	1.16 (1.63)	1.14 (1.51)
— $\pi^\pm$	64	1.05 (0.32)	1.47 (2.39)	1.47 (2.39)
— $K^\pm$	57	0.42 (-3.94)	0.89 (-0.54)	0.88 (-0.60)
— $h^\pm$	110	0.95 (-0.35)	1.11 (0.83)	1.09 (0.69)
— jets	83	0.84 (-1.04)	0.83 (-1.09)	0.83 (-1.09)
— W/Z boson	18	0.60 (-1.30)	0.75 (-0.73)	0.68 (-0.96)
Unpolarized*	5954	1.14 (7.31)	1.14 (7.31)	1.14 (7.31)
SIA*	564	0.87 (-2.27)	0.87 (-2.27)	0.87 (-2.27)
Lattice QCD	48	-	0.58 (-2.38)	0.58 (-2.38)
Total	**8633**†	**1.08 (4.69)**	**1.10 (6.37)**	**1.09 (5.75)**

4.3.2 Helicity-dependent parton distributions

In this section we present the main results for the helicity dependent PDFs inferred from our global QCD analysis. The x dependence of the PDFs $x\Delta f$ is displayed in Fig. 4.3 for each of the quark flavors $\Delta f = \Delta u, \Delta d, \Delta\bar{u}, \Delta\bar{d}$ (and the $\Delta\bar{u} - \Delta\bar{d}$ difference), and Δs^+, along with the quark singlet $\Delta\Sigma$ and gluon Δg PDFs, at a scale $Q^2 = 10$ GeV^2. The results of our full fit, with the cut $W^2 > 4$ GeV^2, are compared with those using the

more restrictive baseline $W^2 > 10$ GeV^2 cut. One of the consequences of the increased W range with the less restrictive cut is that the PDFs are constrained to larger values of x; namely, up to $x_{\rm max} \approx 0.7$, in contrast to $x_{\rm max} \approx 0.54$ for the $W^2 > 10$ GeV^2 cut. In the x region of overlap, there is a marked reduction in magnitude for the lower-W^2 cut, particularly for Δd. For the polarized light antiquarks, $\Delta\bar{u}$ and $\Delta\bar{d}$, the magnitude is more pronounced with the lower-W^2 cut, which leads to a slightly enhanced spin flavor asymmetry $\Delta\bar{u} - \Delta\bar{d}$ at $x \approx 0.01 - 0.1$ compared with the higher-W^2 cut. The effect of the cuts on the polarized strangeness is relatively weak, with a slight preference for nonzero positive values at $x \approx 0.1$ with the lower-W^2 cut, but generally consistent with the no strange quark polarization result indicated by the higher-W^2 cut.

The most dramatic effect, however, is observed for the gluon polarization, Δg, where the $W^2 > 10$ GeV^2 cut was found to allow both positive and negative Δg solutions [126, 127], without additional assumptions about PDF positivity. The introduction of the weaker, $W^2 > 4$ GeV^2 cut, however, which allows inclusion of higher-x DIS data, is able to discriminate between the positive and negative solutions, effectively ruling out the latter (see Sec. 4.3.3 for details). For the quark singlet polarization, since this is dominated by the polarized u-quark distribution, with a smaller d-quark and negligible contributions from other flavors, the impact of the cuts on $\Delta\Sigma$ is not large, and similar to that for Δu.

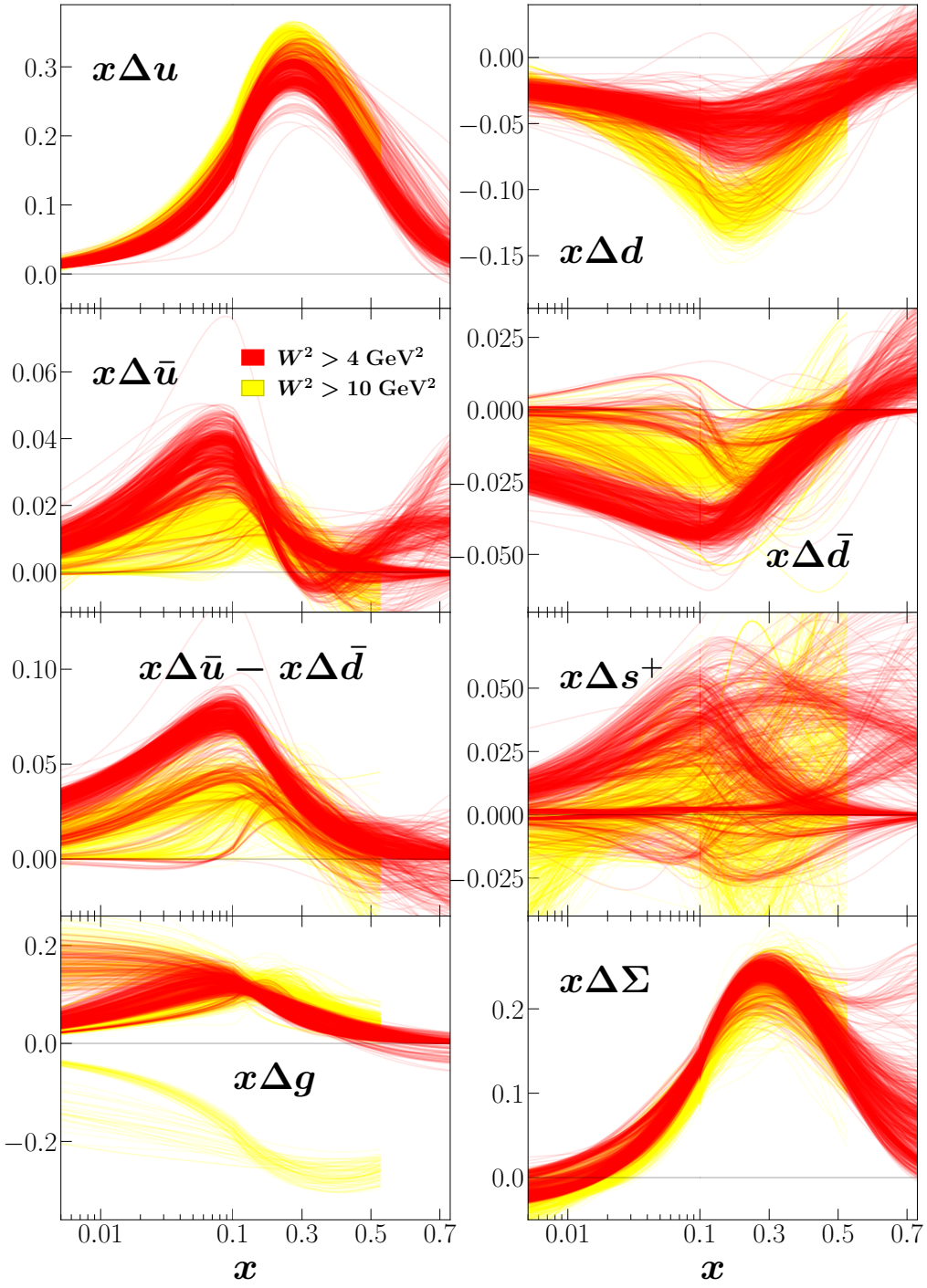


Figure 4.3: Polarized PDFs versus x for various parton flavors with a $W^2 > 4$ GeV^2 cut on the inclusive polarized DIS data (red bands) compared with a stronger $W^2 > 10$ GeV^2 cut (yellow bands) at a scale $Q^2 = 10$ GeV^2. Kinematically the $W^2 > 10$ GeV^2 cut constrains the PDFs up to $x = 0.54$, while the $W^2 > 4$ GeV^2 cut allows PDFs to be constrained up to $x = 0.71$.

Integrating the PDFs in Fig. 4.3 over x, we can obtain the moments of the distributions, which can then in principle be compared with matrix elements of specific operators with a particular twist that can be computed from first principles in lattice QCD. Since our PDFs are determined over a finite range of x, the most direct and experimentally constrained quantity that we can compute is the truncated moment, defined as the integral in the range $x_{\rm min} \leq x \leq x_{\rm max}$, where kinematically $x_{\rm min} = 0.005$ and $x_{\rm max}$ depends on the W^2 cut. These values, computed at a scale $Q^2 = 10$ GeV^2, are listed in Table 4.3. To compute the full moments requires extrapolations into unmeasured regions at small $x < x_{\rm min}$ and large $x > x_{\rm max}$. Lowest truncated and full (extrapolated) moments of the PDFs at a scale $Q^2 = 10$ GeV^2 are listed in Table 4.3.

Table 4.3: Truncated and full moments of PDFs at $Q^2 = 10$ GeV^2. The asterisk * indicates the value is fixed in the fit.

Moment	$[x_{\rm min}, x_{\rm max}]$		
$\int_{x_{\rm min}}^{x_{\rm max}} {\rm d}x$	[0.005, 0.53]	[**0.005**, **0.71**]	[0,1]
Δu^+	0.77(2)	**0.70(6)**	0.74(7)
Δd^+	−0.42(3)	**−0.31(5)**	−0.53(7)
Δs^+	−0.02(5)	**0.08(7)**	0.1(1)
Δg	0.4(4)	**0.2(1)**	0.2(1)
$\Delta\Sigma$	0.33(5)	**0.47(5)**	0.32(9)
g_A	1.18(4)	**1.01(6)**	1.26*
a_8	0.4(1)	**0.2(2)**	0.3(2)
d_2^p	−0.002(4)	**−0.009(9)**	−0.013(16)
d_2^n	−0.001(4)	**−0.002(7)**	−0.005(14)

The polarized PDF ratios in Fig. 4.4 reflect the trend in Fig. 4.3, with a slight reduction in the magnitude of both quantities observed after lowering the W^2 cut to $W^2_{\rm min} = 4$ GeV^2.

In Fig. 4.5 we compare $W^2_{\rm min} = 4$ GeV^2 (which we here label JAM25) with other fitting groups, as well as the previous JAM17 fit. The larger magnitude of $\Delta\bar{u}$ and $\Delta\bar{d}$ of JAM25 is striking when compared to other fits at intermediate x, while the uncertainty bands on Δg are significantly reduced for JAM25 after removing the negative Δg replicas.

4.3.3 Resolving the sign of Δ_g

In order to hone in on the negative Δg solutions we consider several different fits, depending on the datasets included. We label them as follows:

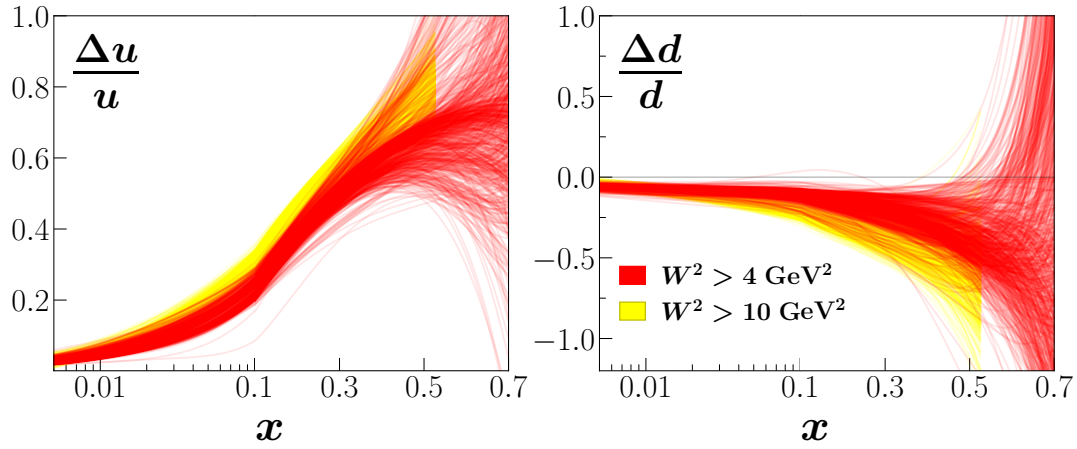


Figure 4.4: Ratios of unpolarized to polarized u and d quarks versus x for various parton flavors with a $W^2 > 4$ GeV^2 cut on the inclusive polarized DIS data (red bands) compared with a stronger $W^2 > 10$ GeV^2 cut (yellow bands) at a scale $Q^2 = 10$ GeV^2.

- **baseline**: a simultaneous fit of unpolarized and polarized PDFs along with FFs from the most recent JAM22 analysis of helicity dependent PDFs [120];
- **+ LQCD**: includes in addition LQCD data, as outlined in Ref. [127], using the "baseline" as the prior;
- **+ high-x DIS**: includes an additional 1370 high-x polarized DIS data points after reducing the W^2 cut from 10 GeV^2 to 4 GeV^2, while retaining the lattice data and using "+ LQCD" replicas as priors.

A comparison of the fit quality between the positive and negative Δg replicas for each of the analyses is summarized in Table 6.1, broken down into specific observables, where we give the reduced χ^2 values, $\chi^2_{\rm red}$, for the different solutions. For the "baseline" analysis, both the positive and negative Δg solutions give good descriptions across all datasets, with an overall $\chi^2_{\rm red} \approx 1.1$ for each case. This is also seen in Fig. 4.6, where a comparison of each replica set with several STAR 2015 [123] and 2013 [124] jet DSA datasets shows excellent descriptions for both positive and negative Δg solutions, highlighting the limited power of the jet data to discriminate the sign of Δg [126]. With the addition of the LQCD data, there is a slight worsening of the $\chi^2_{\rm red}$ for both types of solutions, but no significant inconsistency for the negative Δg replicas in either the lattice or jet DSA data, as previously observed in Ref. [127].

The situation changes markedly when the high-x polarized DIS data are included in the fit. While the positive Δg replicas remain relatively stable with the addition of each successive data set, the negative Δg replicas yield a significantly worse fit, consistently underestimating the data at large p_T values. In particular, for the LQCD data the

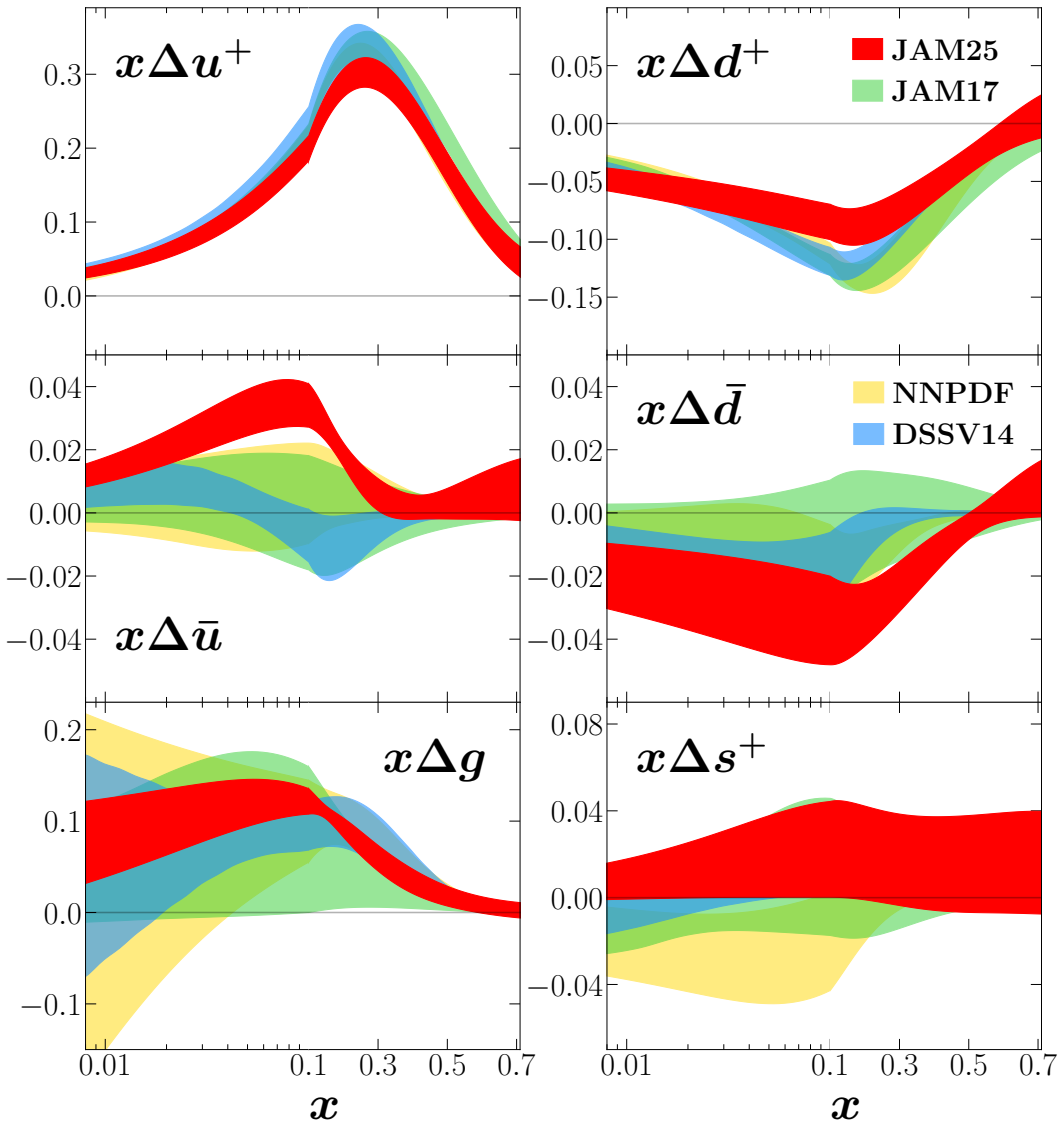


Figure 4.5: Polarized PDFs compared to other PDF fitting groups, computed at $Q^2 = 10$ GeV^2. NNPDF refers to the `NNPDFpol1.1` version.

$\chi^2_{\text{red}} \approx 0.6$ for the positive solutions remains unchanged, but increases from ≈ 1.2 to 3.9 for the negative Δg replicas. A worsening is also seen in the polarized jet data, with χ^2_{red} increasing to ≈ 1.44, and similar effects across all other polarized datasets. We also performed an analysis including the high-x polarized DIS data in the absence of the LQCD data, to check whether the LQCD data were necessary to see this reduction in the quality of the fit. In that case, there was no significant discrepancy in χ^2_{red} between

Table 4.4: Comparison of the $\chi^2_{\rm red}$ values for $\Delta g > 0$ and $\Delta g < 0$ solutions across various datasets. *The "baseline" and "+LQCD" cases have 365 data points in the polarized inclusive DIS dataset, compared with 1735 for the "+ high-x DIS" set.

	$\chi^2_{\rm red}(\Delta g > 0)$			$\chi^2_{\rm red}(\Delta g < 0)$			N
Reaction	baseline	+ LQCD	+ high-x DIS	baseline	+ LQCD	+ high-x DIS	
Polarized	**0.89**	**0.90**	**1.02**	**0.92**	**1.06**	**1.24**	**2067**
DIS	0.95	0.96	1.02	0.98	1.12	1.25	1735*
SIDIS	0.85	0.84	1.14	0.84	0.96	1.11	231
jets	0.84	0.89	0.83	0.88	1.10	1.44	83
$W^{\pm}/Z$	0.60	0.60	0.68	0.83	0.84	1.32	18
Unpolarized	1.14	1.14	1.14	1.15	1.15	1.15	5954
SIA	0.86	0.86	0.89	0.90	0.90	0.92	564
LQCD	—	0.57	0.58	—	1.18	3.92	48
Total	**1.08**	**1.10**	**1.10**	**1.10**	**1.12**	**1.17**	**8633**

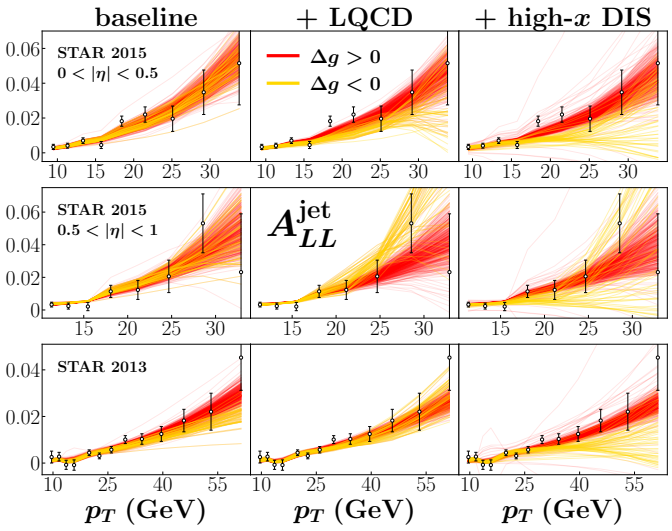


Figure 4.6: Theory versus data comparisons for selected polarized jet DSAs $A_{LL}^{\rm jet}$ from the STAR 2015 ($\sqrt{s} = 200$ GeV) [123] and STAR 2013 ($\sqrt{s} = 510$ GeV) [124] datasets, for positive (red) and negative (yellow) Δg replicas. To the baseline fit (left column) are added lattice data (middle), and further additional high-x polarized DIS data (right).

the positive and negative Δg replicas for any of the datasets. This suggests that to fully rule out the negative Δg solution requires a simultaneous analysis of polarized jets, LQCD, *and* high-x polarized DIS data.

The tension between the LQCD data and the negative Δg solution in the full fit is further illustrated in Fig. 4.7, where we show the contributions to the total χ^2 from each eigendirection of the covariance matrix for the positive and negative solutions. Since for the LQCD data we only have access to the covariance matrix between the 48 lattice points rather than individual uncertainties on each point, we rotate the residuals

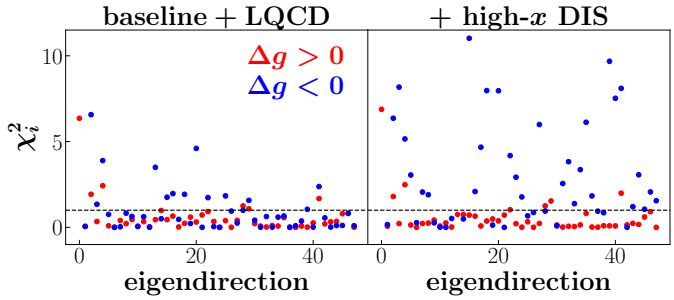


Figure 4.7: Comparison of χ^2 contributions for positive (red dots) and negative (blue dots) Δg mean theory predictions for LQCD data versus eigendirection; χ^2_i represents the i-th lattice data point.

in the eigenspace of the covariance matrix [127]. Compared with the baseline + LQCD scenario, there is clearly a larger disagreement between the theory predictions and the LQCD data for the $\Delta g < 0$ solutions, across a majority of data points after the addition of the high-x DIS data.

Turning to the polarized gluon PDF itself, in Fig. 4.8 we display the positive and negative Δg solutions for each of the replica sets for the baseline, baseline + LQCD, and baseline + LQCD + high-x DIS scenarios, along with the unpolarized gluon PDF for comparison. The negative Δg replicas for the "+ high-x DIS" fit are not shown, having been ruled out by the combination of all the datasets. By reducing the W^2 cut on the polarized DIS data from 10 GeV^2 to 4 GeV^2 in the "+ high-x DIS" case, we are able to accurately characterize the polarized gluon PDF at higher values of x without the need for extrapolation. The uncertainties on Δg are also slightly reduced following the progressive addition of the new data. As a somewhat indirect constraint on the sign of the gluon helicity, de Florian *et al.* [207] observed that the negative Δg replicas corresponding to the JAM22 analysis (our baseline) would lead to violations of PDF positivity at large x, which could lead to unphysical cross sections for Higgs production in polarized pp collisions at RHIC. To leading order the double-helicity asymmetry for Higgs production is dominated by the gluon-gluon channel,

$$A_{LL}^{\mathrm{H}}(\tau) = \frac{[\Delta g \otimes \Delta g]}{[g \otimes g]} + \mathcal{O}(\alpha_s), \tag{4.47}$$

where $\tau = m_H^2/s$ for the Higgs mass m_H and $\sqrt{s} = 510$ GeV is the RHIC center of mass energy. Computing the magnitude of the asymmetry $|A_{LL}^{\mathrm{H}}|$, de Florian *et al.* found this exceeded unity for the negative Δg replicas, especially at large (unphysical) values of m_H.

Repeating the calculation of A_{LL}^{H} at next-to-leading order as in Ref. [207] for the physical Higgs mass, $m_H = 125$ GeV, in Fig. 4.9, we show a histogram of results for negative Δg replicas for the "baseline" and "+ LQCD" scenarios. For the baseline fit,

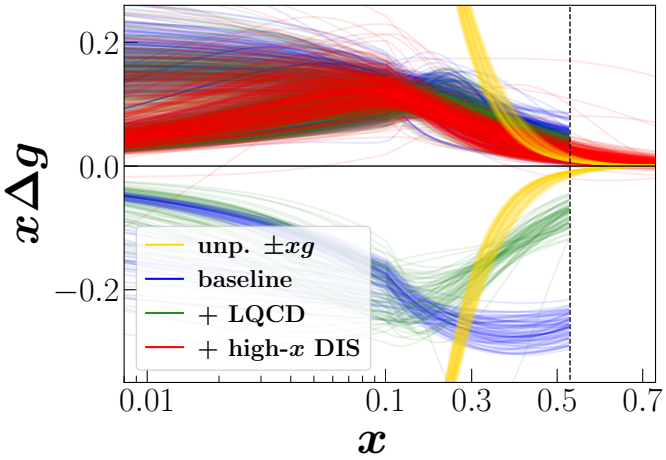


Figure 4.8: Monte Carlo replicas for the polarized gluon PDF $x\Delta g$ and the unpolarized gluon PDF xg at $Q^2 = 10$ GeV^2. The "baseline" and "+ LQCD" cases are cut off at $x \approx 0.53$, corresponding to a $W^2 > 10$ GeV^2 cut for polarized DIS.

we confirm the observations [207] that the asymmetry exceeds unity for most of the replicas. On the other hand, for the "+ LQCD" case, the vast majority of the replicas give rise to $A_{LL}^{\mathrm{H}} < 1$, so that this observable does not rule out negative Δg values. The negative Δg solutions for the "+ LQCD" case shown in Fig. 4.8 are therefore viable solutions which describe all of the data in the baseline fit and respect positivity constraints of the Higgs asymmetry. It is only with the unique addition of the high-x polarized DIS data described above that one is able to conclusively rule out the negative Δg solution from the combination of empirical and lattice data, without the need for artificial constraints on the positivity of PDFs.

4.3.4 Higher Twist Effects

Here we present our findings with respect to higher twist contributions to the polarized structure functions g_1 and g_2. We first look at g_2^{HT}, the parametrized HTs for g_2 over and above the leading twist Wandzura-Wilczek structure functions as described in Eq. 4.16. In Fig. 4.10 the neutron g_2^{HT} is consistent with zero despite a slightly negative shape at high x, while the proton has a slight preference for a negative g_2^{HT} at intermediate x. g_1^{HT} was found to be entirely consistent with zero, meaning that this slightly negative proton g_2^{HT} is the only clear indication we have in our analysis for any significant HT effects. g_2^{HT} can be seen to blow up at low x, however since HTs have no impact at low x the HT parameters are unconstrained in this region.

Turning to the total g_1 and g_2 combining leading twist, TMCs and HT effects, we can see in Fig. 4.11 that there is little difference between the $W^2 > 4$ GeV^2 and $W^2 >$

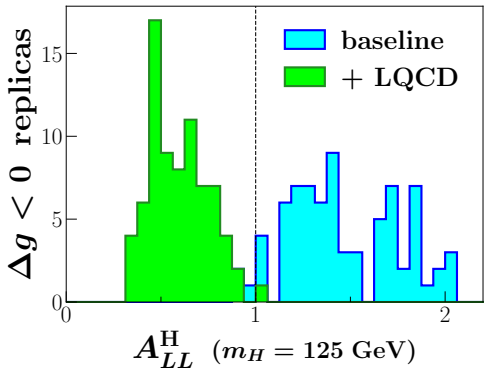


Figure 4.9: Histogram of negative Δg replicas for the Higgs production asymmetry $A_{LL}^{\rm H}$ at RHIC ($\sqrt{s} = 510$ GeV), for the "baseline" analysis and after the addition of LQCD data.

10 GeV^2 cuts. This suggests minimal evidence of any higher twist effects emerging through the inclusion of high-x polarized DIS data, which is further backed up by observing the shape of d_2 in Fig. 4.12, which is once again consistent with zero for the proton and neutron across Q^2. Recall that accoording to Eq. 4.21 the d_2 moment is dependent only on higher twists and power corrections, so the lack of any significant deviation from zero implies that these contributions must be negligible. As a final metric we include the BC sum rule as shown in Fig. 4.13, which depends entirely on HT contributions to g_2 as described in Eq. 4.20. The uncertainty band for the proton is very slightly below zero, reflecting the slightly negative proton $g_2^{\rm HT}$, however this supports the message that any noticeable HT effects seen in our analysis are small at most.

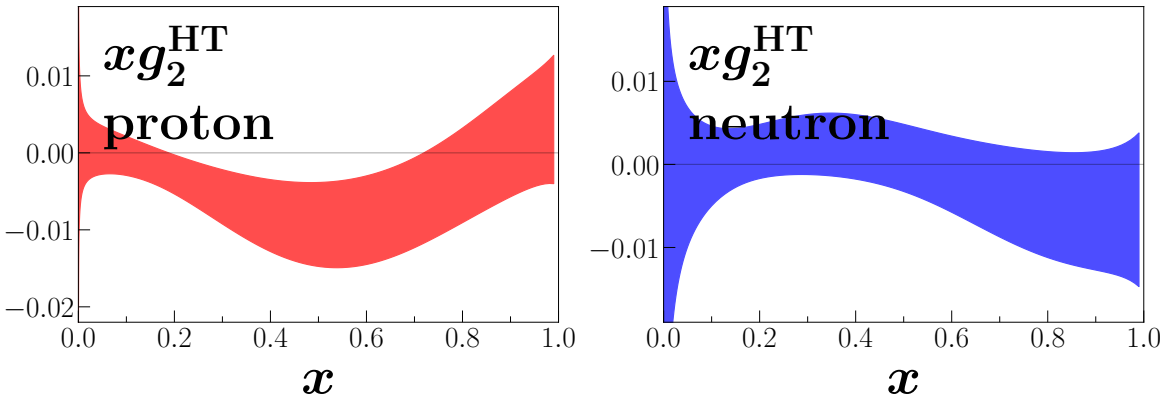


Figure 4.10: $xg_2^{\rm HT}$ for proton and neutron evaluated at $Q^2 = 10$ GeV^2.

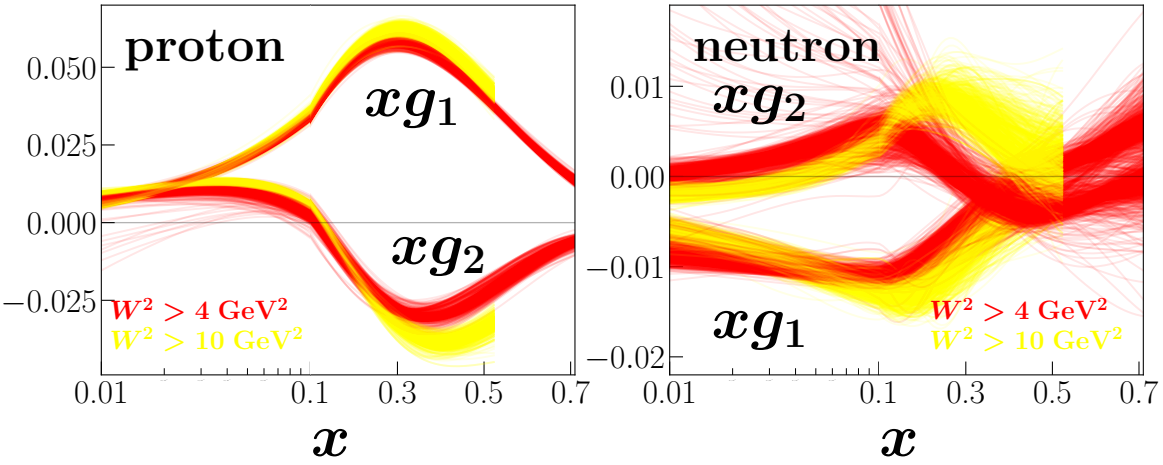


Figure 4.11: Polarized structure functions, for a $W^2 > 4$ GeV^2 cut on the inclusive polarized DIS data (red bands) compared with a stronger $W^2 > 10$ GeV^2 cut (yellow bands) at a scale $Q^2 = 10$ GeV^2.

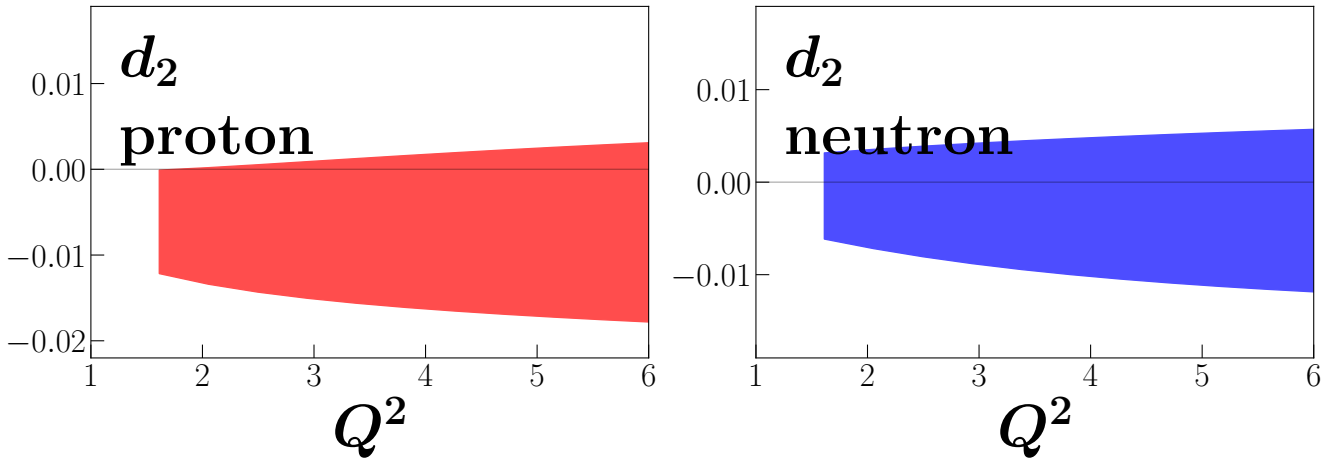


Figure 4.12: d_2 moment for proton and neutron.

4.4 Conclusion

We have performed a new global QCD analysis of spin-dependent parton distributions including all available inclusive DIS data on longitudinal and transverse polarization asymmetries down to $W^2 = 4$ GeV^2, alongside SIDIS, W boson and jet production as well as LQCD data. This marks the most comprehensive JAM analysis to date, once the various unpolarized and SIA datasets are taken into account from our baseline analysis. We observe a sizeable reduction in the magnitude of the Δd PDF at intermediate x,

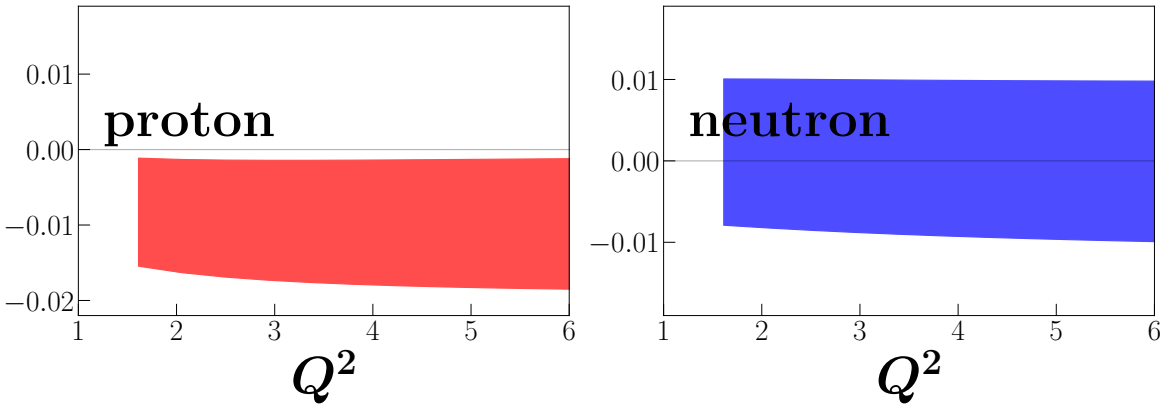


Figure 4.13: Burkhardt-Cottingham (BC) sum rule for proton and neutron.

along with a large increase in spin flavour asymmetry $\Delta\bar{u} - \Delta\bar{d}$. This effect is driven mostly by the inclusion of extra high-x polarized DIS data points, and given that we have found only small higher twist contributions to the polarized structure functions this observation cannot be traced back entirely to higher twist effects. Further work needs to be done in order to isolate the primary cause of these changes to the PDFs within the high-x DIS datasets that have been added here.

We are also able to rule out negative Δg solutions, due to an incompatibility between high-x polarized DIS, jet production and LQCD data for such solutions. Only once all three datasets are included are we able to expose a dramatic worsening of the quality of the fit, which can be entirely avoided by setting Δg to be positive. Additional lattice data on pseudo-ITDs Ioffe for the quark singlet helicity, $\mathcal{I}_{\Delta\Sigma}$, in addition to improved data on the gluon $\mathcal{I}_{\Delta g}$, as well as simulations of their moments [208], would provide important constraints on the sign of Δg. Other observables which have been shown to have sensitivity to Δg include datasets in polarized SIDIS at large p_T [209], which could be studied at higher energies at Jefferson Lab and the Electron-Ion Collider [210].

Chapter 5

Accelerating Markov Chain Monte Carlo sampling with diffusion models

A large number of scientific theories exist as parametric models, in which predictions depend on the values of a set of free parameters that are *a priori* unknown. To progress our understanding of these theories, it is necessary to infer the values of model parameters by comparing theoretical predictions to one or more sets of observed data. Rigorous global statistical fits of theories of interest in particle astrophysics typically involve tens to hundreds of parameters (if one includes nuisance parameters), and frequently require the simulation of a wide range of data in collider physics, flavor physics, astrophysics and cosmology [211, 212]. The computational expense of this process necessitates the development of fast and accurate sampling techniques such as MCMC (see Sec. 2.2) that can reduce the time required to adequately explore large parameter space volumes.

A well-known deficiency of MCMC methods concerns multi-modal posterior functions. A proposal function based on a local proposal from the current point in the Markov Chain tends to lead to one mode being found, but often has low efficiency for jumping between modes. An uninformed, nonlocal proposal distribution, meanwhile, will typically lead to a high rejection rate. Finding more efficient proposal functions for the exploration of multi-modal posteriors reduces the time required to adequately explore the parameter space, hence allowing problems of greater complexity to be tackled.

In this chapter, we present a new algorithm that augments a Metropolis-Hastings sampler with a proposal function based on a diffusion model. A diffusion model is a parametrized Markov chain trained to produce samples matching a set of input data after finite time. Such models have recently been shown to be highly effective in image synthesis, including outperforming previous approaches based on generative adversarial

networks [213, 214, 215, 216, 217, 218, 219, 220]. By creating a diffusion model from the available samples at regular intervals during a Metropolis-Hastings sampling run, one can develop an approximation of the posterior that can be used as an informed global proposal function to increase the efficiency of MCMC sampling. This is not the only way one could perform posterior sampling using diffusion models; one could instead use stochastic differential equations to gradually approximate the posterior via diffusion processes [221]. An alternative approach based on the use of normalizing flows for lattice QCD was previously presented in Ref. [222, 223, 224], and a similar combination of normalizing flows with MCMC sampling was outlined in Ref. [225, 226]. Normalizing flows have in fact been combined with diffusion models, though not in the context of MCMC [227]. We note that there are a number of important differences between diffusion models and normalizing flows:

- A flow-based approach requires the estimation of gradients, while for diffusion models this is unnecessary. This is a highly useful feature when considering, for example, particle astrophysics global fits which often use stochastic simulations where gradient information is impossible to obtain [228, 229, 230].
- A fully-trained normalizing flow has direct access to a probability distribution function that approximately matches the target function [222], while a fully-trained diffusion model can only produce samples that are approximately distributed according to the target function.
- Normalizing flows can be subject to topological limitations, particularly for multi-modal functions [231]. Diffusion models are more successful in matching the shape of sharply-peaked or multi-modal functions.
- Diffusion models can be slow to train compared to normalizing flows given the need to train on many time-discretization steps [227]. We address this concern by fitting parameters that define the variance of the noise added in each time step, rather than training a neural network. This speeds up the training process dramatically.

We benchmark our improved technique against regular Metropolis-Hastings sampling, but one could easily accelerate other MCMC samplers using the same diffusion model technique presented here.

This chapter is structured as follows. In Section 5.1, we provide a brief, self-contained introduction to diffusion models. In Section 5.2, we present our algorithm that marries a diffusion model to an MCMC sampler. In Section 5.3, we perform a series of numerical tests of our new algorithm by attempting to map various analytic test functions. We compare the performance of our algorithm in these tests to other basic MCMC samplers, as well as to a specific normalizing flow algorithm. In Section 5.4,

we extend the numerical tests to a physical example relating to a global fit of parton distribution function (PDF) parameters. We summarize our conclusions in Section 5.5. These results are published in [232], and a code implementation can be found at https://github.com/NickHunt-Smith/MCMC-diffusion.

5.1 A brief introduction to diffusion models

Diffusion models are a form of *generative model* that, after training on a set of data, are able to reproduce examples of that data. Starting with, for example, an image, the key concept in a diffusion model is to systematically decay the image into pure noise through the repeated addition of a noise component (usually chosen to be Gaussian). It is then possible to learn the reverse of this noise addition process and recover the original information. Examples of valid images can then be obtained by feeding pure noise distributions through the learned reverse process.

To put this on a more precise footing, assume that we start with a data vector $\boldsymbol{x}_0$. We can then refer to a data probability distribution $Q(\boldsymbol{x}_0)$, which is the distribution of possible "real" data vectors. Any sample from this distribution is itself a valid data vector, and a sample from this distribution is written as $\boldsymbol{x}_0 \sim Q(x_0)$. We now want to add Gaussian noise for T successive time steps. We can define a *forward diffusion process* $Q(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})$ which adds the noise at time step t. Note that we can formally define this as a Markov chain, since the prediction of the probability density at time t only depends on the immediate predecessor at time $t-1$. In principle, the width of the Gaussian noise that we add at each time step can be different, so we further define a *variance schedule* $0 < \beta_1 < \beta_2 < \ldots < \beta_T < 1$ where each β_t is set to the same value across each parameter, then define the forward diffusion process as:

$$Q(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}) = \mathcal{N}\big(\boldsymbol{x}_t; \sqrt{1-\beta_t}\,\boldsymbol{x}_{t-1}, \beta_t \boldsymbol{I}\big). \tag{5.1}$$

As time progresses, each new data vector at time step t is drawn from a Gaussian distribution with mean $\boldsymbol{\mu}_t = \sqrt{1-\beta_t}\,\boldsymbol{x}_{t-1}$ and variance $\sigma_t^2 = \beta_t$. Starting from the original vector $\boldsymbol{x}_0$, we generate a series of increasingly noisy vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_t, \ldots, \boldsymbol{x}_T$, where the aim is to have $\boldsymbol{x}_T$ be pure Gaussian noise.

We know the forward process very well, since it is easy to apply Gaussian noise. However, what about the reverse process? Let the probability distribution that takes us backwards along the chain of images be $P(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t)$. In general, $P(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t)$ cannot be known analytically, since we would need to know the distribution of all possible images in order to calculate it. Instead, we can introduce an approximation to the reverse transformation $P_\phi(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t)$, with ϕ representing the parameters of a given model for the reverse process. The model is typically provided by a trained neural network, although in principle one could parametrize a model that appropriately describes each reverse step.

For Gaussian diffusion in the continuous limit (*i.e.*, in the limit of small step size, β), the reversal of the diffusion process has the identical functional form as the forward process [213, 233] [1]. The reverse probability distribution can therefore be described by a Gaussian distribution with a mean $\boldsymbol{\mu}_\phi$ and a variance Σ_ϕ:

$$P_\phi\big(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t\big) = \mathcal{N}\big(\boldsymbol{x}_{t-1}; \boldsymbol{\mu}_\phi(\boldsymbol{x}_t, t), \Sigma_\phi(\boldsymbol{x}_t, t)\big). \tag{5.2}$$

The mean and variance are also functions of the time t, since the reverse transformation is slightly different at each time slice.

5.2 A diffusion-model assisted Markov Chain Monte Carlo sampler

For testing the effectiveness of diffusion model accelerated MCMC sampling, we paired the diffusion model with a simple Metropolis-Hastings (MH) MCMC algorithm, as introduced in Section 2.2. Here we simplify the notation from Eq. (2.8) for clarity, relabelling the acceptance probability $\alpha(\boldsymbol{a}_{t-1}, \boldsymbol{a}') = a$, the previous sample $\boldsymbol{a}_{t-1} = \boldsymbol{\theta}$, the candidate point $\boldsymbol{a} = \boldsymbol{\theta}'$ and we drop the $|\mathbf{m}$ from the posterior distribution p:

$$a = \min\left(1, \frac{p(\boldsymbol{\theta}')}{p(\boldsymbol{\theta})} \cdot \frac{q(\boldsymbol{\theta}|\boldsymbol{\theta}')}{q(\boldsymbol{\theta}'|\boldsymbol{\theta})}\right). \tag{5.3}$$

For q we adopt the typical choice of a Gaussian proposal function centred on $\boldsymbol{\theta}$, which as a symmetric function reduces the q term in Eq. (5.3) to unity. We identify this algorithm with a Gaussian q as *pure MH*. To incorporate a diffusion model into this process, we occasionally replace the Gaussian proposal function with a diffusion model, which is periodically re-trained as the MCMC chain accumulates samples. These "diffusion proposal samples" are generated from random noise via the diffusion model as described in Section 5.1, and as such each diffusion proposal sample is completely independent of the previous sample. This is a special case of MH, called *independence MH*, which simplifies the q factors in Eq. (5.3) to $q(\boldsymbol{\theta})/q(\boldsymbol{\theta}')$ [234]. As long as the diffusion proposal samples are run through the MH accept/reject step, the distribution of samples is guaranteed to satisfy detailed balance and converge to the posterior eventually [235, 234]. In addition, as the overall samples get closer to the true posterior, the diffusion model is continually being trained to produce samples that more closely resemble the true posterior. Eventually, the diffusion model will be asymptotically exact — that is, the diffusion model samples will be distributed such that they approximate the target posterior, in which case $p(\boldsymbol{\theta}) \approx q(\boldsymbol{\theta})$ [223]. In the limit of many samples, the acceptance rate of the diffusion samples will be close to 1. One can think of this process as a method of correcting the proposal distribution to match the desired distribution p.

[1] Note that this is also true of binomial diffusion.

Since the diffusion proposal samples are distributed according to our knowledge of the posterior, they will only be symmetric in the cases where the posterior is also symmetric. Therefore, in most situations the q term will not simply reduce to unity when using the diffusion model as a proposal function. We can approximate the probability of proposing a given sample $q(\boldsymbol{\theta})$ with a Gibbs sampling method. For a given proposal vector in D dimensions $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, ..., \theta_D)$, we first draw a number of diffusion proposal samples. We then take a 1D histogram of those samples to find the approximate marginal distribution of the θ_1 parameter. $q(\theta_1)$ is then estimated by dividing the number of samples in the bin that θ_1 falls into by the total number of samples. Following this step, we set θ_1 at its specific value and simulate from the next parameter θ_2, before taking another 1D histogram of the θ_2 parameter to estimate $q(\theta_2|\theta_1)$. We repeat this D times, before multiplying the probabilities together according to Bayes' rule,

$$q(\boldsymbol{\theta}) = q(\theta_1) \times q(\theta_2|\theta_1) \times q(\theta_3|\theta_1, \theta_2) \times \cdots \times q(\theta_D|\theta_1, \theta_2, \theta_3, \ldots, \theta_{D-1}). \tag{5.4}$$

This method was chosen primarily to reduce computation time and avoid having to make use of high-dimensional histograms.

Retaining a mix of diffusion model proposals and Gaussian proposals is necessary because the diffusion model can only mimic the distribution of samples it has been provided with. Hence, at early points in the MCMC run, the diffusion model will not provide an adequate approximation of the final posterior, and therefore samples taken from it will not have a particularly high acceptance rate. Later in the run, however, the diffusion model will provide a more accurate representation of the posterior, and the acceptance rate will improve. Inspired by the algorithmic structure in Refs. [225, 226], we therefore designed Algorithm 1, which uses pure MH some of the time alongside the diffusion proposal samples. We call this algorithm `MCMC-diffusion`.

Algorithm 1 MCMC-diffusion

for $i = 1, ..., n$ **do**
  Generate a random number r from the uniform distribution $\mathcal{U}(0, 1)$
  if $r < p_{\text{diff}}$ **then**
    Metropolis-Hastings with diffusion model proposal
  else
    Pure Metropolis-Hastings with Gaussian proposal
  end if
  if $i \mod \tau = 0$ **then**
    Retrain diffusion model on existing samples
  end if
end for

In `MCMC-diffusion`, n is the total number of samples to be generated, p_{diff} is the probability of using a diffusion proposal sample as opposed to a local Gaussian

proposal sample, and τ is the number of samples to generate before retraining the diffusion model. τ is tunable to the posterior one wishes to sample from, with lower values of τ resulting in better performance at the cost of computation time to retrain the diffusion model.

The primary advantage of using the diffusion model over standard MCMC is the nonlocal nature of its proposed points within a single chain. Where a standard MCMC chain might get stuck in a local mode, the diffusion model can jump to any point in parameter space. Algorithms such as Explore-Exploit MCMC (Ex^{2}MCMC) [236] and normalizing flows make use of this global transition kernel. For multi-modal posteriors, this means one can accurately characterize the relative weights between the modes with only a single chain. Our only requirement occurs if the modes are distanced too far apart for the Gaussian proposal to jump between them with a significant probability. In that case, all the important modes must be located beforehand in order to seed the diffusion model with points at those modes for initial training, allowing transitions between the modes.

The diffusion model can also accelerate the MCMC process even for posteriors with a single mode. The proposed diffusion samples are more likely to be located in strongly peaked regions of the parameter space, meaning that if the MCMC chain strays far away from the mode the diffusion samples will bring it back to the region of interest. This is particularly significant for computationally expensive posteriors, where reducing the number of likelihood evaluations is the highest priority. A further advantage of the diffusion model is that, once the sampling algorithm has converged, the diffusion model has already been trained to produce samples that closely resemble the posterior. We therefore obtain a fast way of generating as many approximate posterior samples as desired. This is particularly useful in applications where a rough initial scan can be used to define a much smarter prior for a subsequent scan. Global fits of parameter spaces which exhibit a degree of fine-tuning are an obvious use-case.

Applying a diffusion model to MCMC samples rather than images is a considerably simpler task in terms of algorithmic complexity. A single high-quality image consists of millions of pixels in several color channels which must be converted into a high-dimensional tensor, and millions of training images must be supplied in order for the diffusion model to learn and then generate an image that looks somewhat like the training images. On the other hand, our MCMC samples are simple low-dimensional arrays with a length equal to the number of dimensions of the problem; rarely more than a few hundred for realistic physics examples. We have therefore constructed a diffusion model that is tailored towards MCMC samples rather than attempting to adapt existing image-based diffusion models.

When we are training the diffusion model, we define the forward Gaussian noise process by adding noise to the samples in the same way as Eq. (5.1). In other words, each vector in the overall set of samples that form the Markov chain is corrupted slightly

at each time step. This forward process will be concluded once the distribution of the vectors has been converted from a shape that resembles the posterior to pure Gaussian noise. This introduces two hyperparameters which can be tuned as appropriate: the number of time steps and the amount of noise progressively added at each time step, β_T. Generally, increasing these values will further reduce the minimum loss, but also increase the computation time needed to find the minimum.

To train the diffusion model, we first perform the forward diffusion process according to Eq. (5.1) until we are left with pure Gaussian noise. On an individual sample level,

$$\boldsymbol{x}_t = \sqrt{1-\beta_t}\boldsymbol{x}_{t-1} + \sqrt{\beta_t}\boldsymbol{\epsilon}, \tag{5.5}$$

where $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \boldsymbol{I})$. We then define a new data vector $\boldsymbol{y}_T$ drawn from pure Gaussian noise with the same length as $\boldsymbol{x}_T$. Each reverse step is then computed using

$$\boldsymbol{y}_{t-1} = \boldsymbol{y}_t - (\boldsymbol{x}_t - \boldsymbol{x}_{t-1})\boldsymbol{\phi}_t. \tag{5.6}$$

Here, the difference in samples between each time step $(\boldsymbol{x}_{t-1} - \boldsymbol{x}_t)$ is multiplied by a free parameter $\boldsymbol{\phi}_t$ to be trained. Repeating this operation T times leads to the denoised samples $\boldsymbol{y}_0$. We can define a loss function L as

$$L = \sum_t \left(\boldsymbol{y}_{t-1}(\boldsymbol{\phi}_t) - \boldsymbol{x}_{t-1}\right)^2, \tag{5.7}$$

summed over the total number of samples. Rather than training a neural network, we therefore perform a series of fits of the $\boldsymbol{\phi}_t$ parameters to minimise the sum of squared differences in each time step between the denoised samples we have generated and the original MCMC samples. There are $T \times D$ parameters to be fit simultaneously for each retrain of the diffusion model. To minimise the loss function we use the `SciPy minimize` function.

We then perform the reverse process by generating another new data vector $\boldsymbol{z}_T$ from pure Gaussian noise, before applying Eq. (5.6) to obtain $\boldsymbol{z}_0$. We can repeat this process indefinitely, generating as many samples as required almost instantaneously. This reverse process is considerably simpler than the typical method of parametrizing reverse Gaussian transformations described in Eq. (5.2). Empirically, we found it to be highly effective at generating unique diffusion samples approximately distributed according to the target distribution.

5.3 Analytic test functions

In the test functions below, we explore the advantages `MCMC-diffusion` has over pure MH, as well as over a normalizing flow MCMC algorithm. The values assigned for each hyperparameter in each function are contained in Appendix C.1. A further test comparing `MCMC-diffusion` and pure MH for the 4D Rosenbrock function can be found in Appendix C.2.

5.3.1 2D Himmelblau function

The Himmelblau function is given by

$$f(\boldsymbol{\theta}) = (\theta_1^2 + \theta_2 - 11)^2 + (\theta_1 + \theta_2^2 - 7)^2, \tag{5.8}$$

which consists of 4 sharply peaked modes of equal height spaced a large distance away from each other, presenting a significant challenge for MCMC samplers. Without multiple chains, we have found that a simple MCMC algorithm like pure MH will either get stuck in one of the modes if the Gaussian proposal is too narrow, or have a very low acceptance rate if the Gaussian proposal is too wide. Augmenting the MH sampler with a diffusion model as in `MCMC-diffusion` allows the chain to jump between these modes, while accurately characterizing the relative weights between the modes. This behavior is shown in Fig. 5.1, which depicts 10,000 samples of the 2D Himmelblau from a single chain using `MCMC-diffusion`. The nonlocal jumps made by the diffusion model are also represented visually by the black lines connecting the 4 modes, of which there are around 1,000 in this case. The jumps between modes ensure that the posterior is very well described, with all of the modes mapped out equally while requiring relatively few function evaluations.

Before the chain began, the diffusion model was seeded by providing 50 points located near each of the modes. Given that the modes are positioned far from each other, it is necessary to do this to allow the diffusion model to propose jumps between modes at the start. These initial points are increasingly unlikely to be proposed as the chain progresses since there are many more samples being generated, and they are removed at the end of the chain.

5.3.2 10D Gaussian mixture

Here we compare our results using `MCMC-diffusion` to results obtained from a normalizing flow MCMC algorithm for a high-dimensional Gaussian mixture function with 2 components. In the same way as in [225], we define a mixture of 2 Gaussians in 10D,

$$f(\theta_i) = w_A \frac{e^{-\frac{1}{2}|\theta_i - \theta_{A,i}|^2}}{(2\pi)^{d/2}} + w_B \frac{e^{-\frac{1}{2}|\theta_i - \theta_{B,i}|^2}}{(2\pi)^{d/2}}, \tag{5.9}$$

with weights $w_A = 2/3$, $w_B = 1/3$ and non-zero means only in the first two dimensions $\theta_{A1,2} = (8, 3)$, $\theta_{B1,2} = (-2, 3)$. We therefore have to sample two modes in the first two dimensions situated significantly far away from one another, with different relative weights. As shown in Fig. 5.2, `MCMC-diffusion` converges to the truth within 5,000 samples, which is very few function evaluations given the high-dimensional nature of this problem. Once again, the diffusion model was initially seeded with 50 points near the two modes.

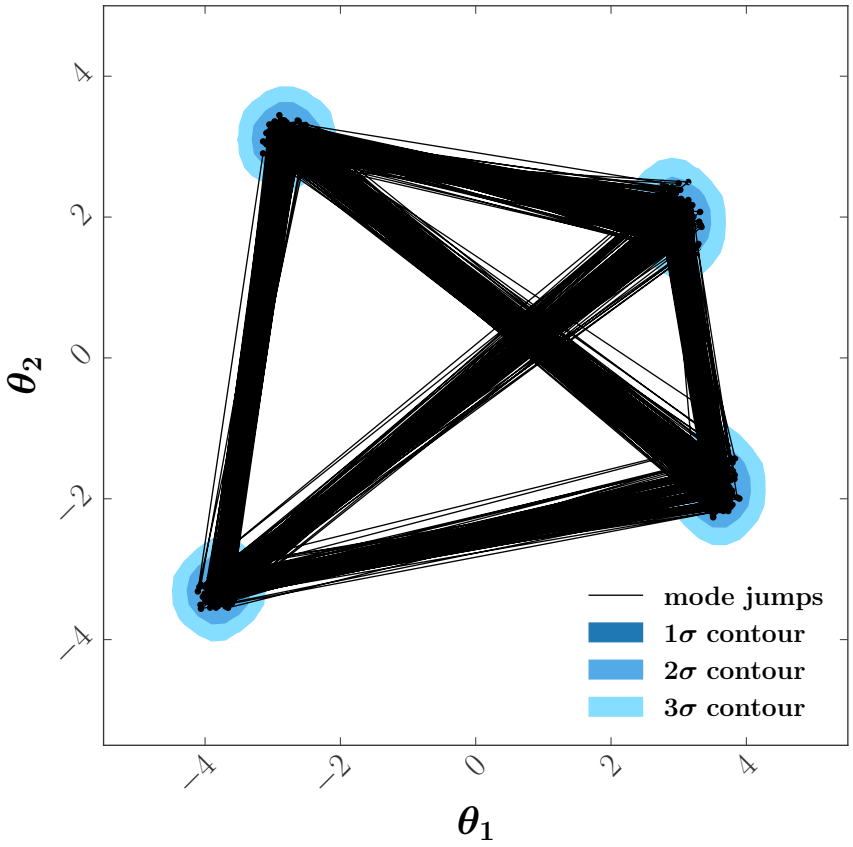


Figure 5.1: 2D Himmelblau function sampled with MCMC-diffusion (10,000 samples). Black lines depict jumps between modes due to the diffusion model.

Importantly, our diffusion model MCMC sampler outperforms a similar normalizing flow MCMC sampler for this same problem by requiring fewer training iterations in order to reach a high acceptance rate. Figure 5.3 shows that `MCMC-diffusion` is able to reach close to 80% acceptance within less than 100 training iterations, whereas the flow-based approach requires at least 500 iterations to reach a comparable acceptance rate [225].

We do however point out that neither the total number of samples nor the number of samples for each training iteration were specified for the normalizing flow algorithm. We further note that the full structure of the normalizing flow algorithm is significantly more complex than our `MCMC-diffusion`. A normalizing flow is an invertible map that takes a base density such as a Gaussian with unit variance, and pushes it forward

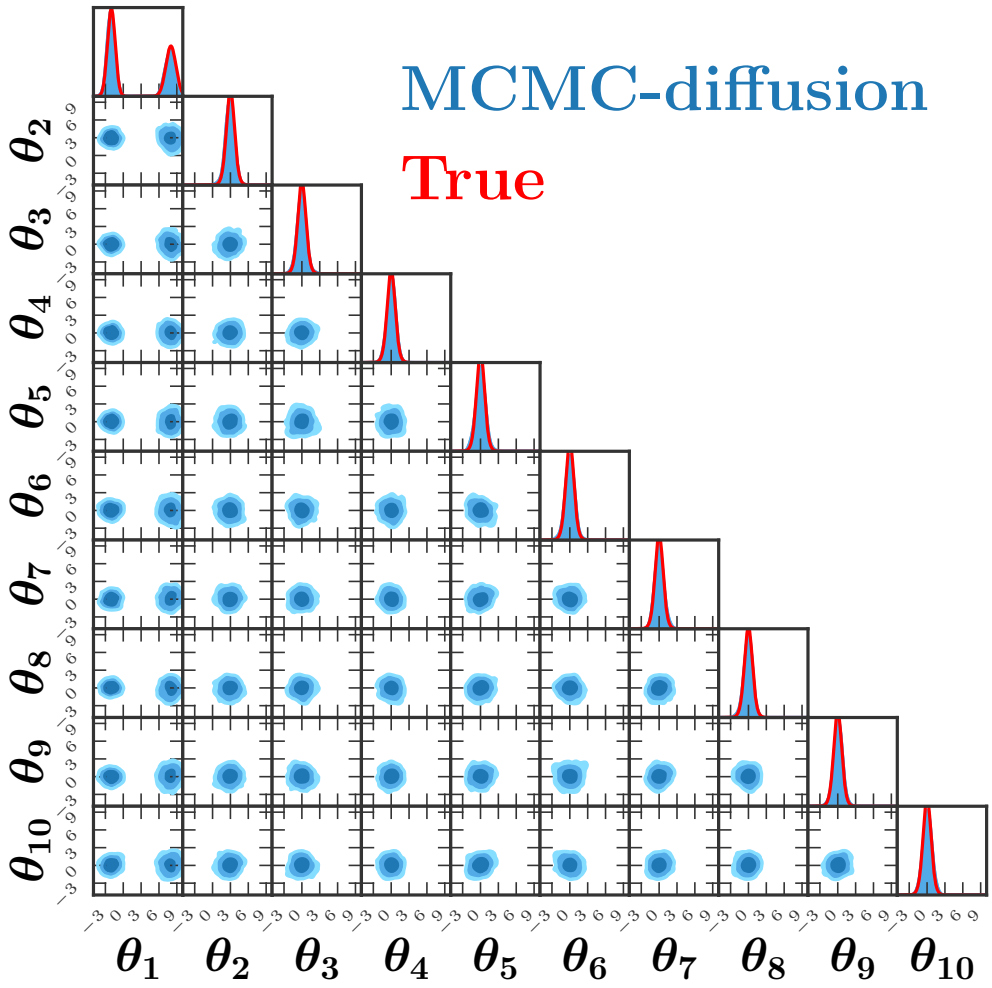


Figure 5.2: Corner plot of MCMC-diffusion for a 10D Gaussian mixture with 2 components (5,000 samples). The true marginalized posteriors are shown in red, blue contours represent sample density.

towards a target density. In [225], neural networks are trained to learn the mapping between the base and target densities by minimizing the Kullback-Leibler divergence. The normalizing flow is therefore supervised at the PDF level rather than the sample level, which may explain the increased performance on the side of the diffusion model.

Where we have paired our diffusion model with a basic local Gaussian transition kernel, the normalizing flow algorithm is combined with a more sophisticated local MALA sampler. Furthermore, the diffusion model algorithm operates in a single chain, whereas hundreds of independent walkers are initialized and split between each mode at the start of the normalizing flow algorithm. We therefore expect our results to be

conservative in performance - a more sophisticated transition kernel and algorithmic structure would likely result in better performance in our case. The acceptance rate comparison is merely suggestive of an advantage the diffusion model may have over normalizing flows for this particular example, an exact comparison is considered beyond the scope of this work.

Figure 5.3 also shows the property of asymptotic exactness shared by independent MH samplers such as these. As the MCMC chain progresses and the diffusion model/normalizing flow is repeatedly retrained, the acceptance rate increases well beyond 70%. This indicates that the proposal diffusion model is closely approximating the target posterior, as explained in Sec. 5.2. It also highlights an important feature of algorithms of this kind - at the conclusion of the MCMC chain, one obtains a very fast approximation to the posterior from which one can draw as many samples as required.

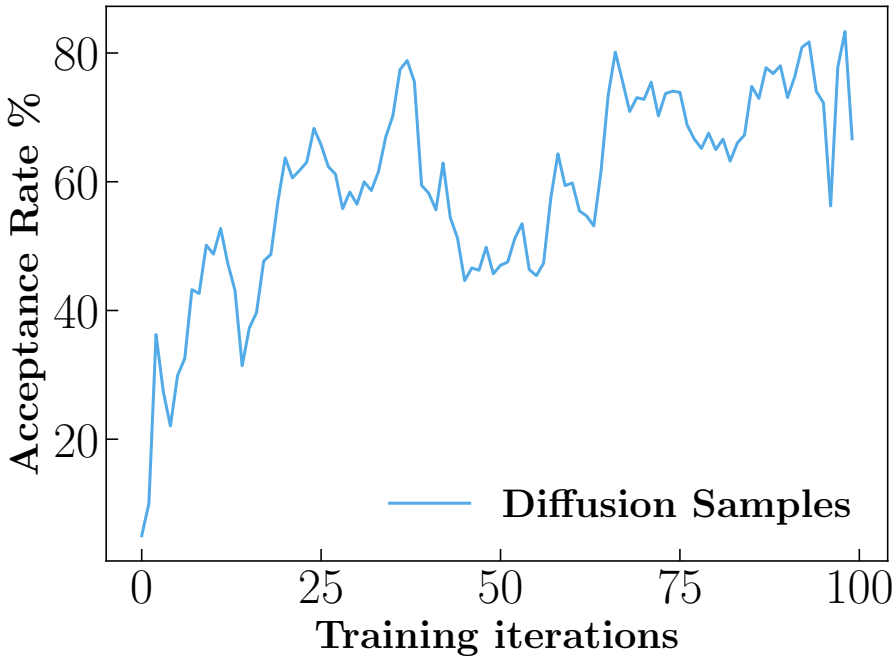


Figure 5.3: Acceptance rate with MCMC-diffusion for a 10D Gaussian mixture with 2 components.

5.3.3 4D EggBox function

The EggBox function is given by

$$f(\boldsymbol{\theta}) = \left(2 + \prod_{i=1}^{D}\left[\cos\left(\frac{\theta_i}{2}\right)\right]\right)^5, \tag{5.10}$$

which consists of many local modes of equal height spread at regular intervals throughout parameter space. The challenge for an MCMC algorithm is to find all of the modes while weighting them equally. Given a wide enough Gaussian proposal, pure MH is eventually able to successfully find all the minima, however the inclusion of the diffusion model accelerates this process dramatically. Figure 5.4 shows a comparison between pure MH and `MCMC-diffusion` for the heavily multi-modal 4D EggBox function over 0.1M samples. For such a low number of samples, pure MH is unable to map a substantial fraction of the total modes (it produced samples in around 55% of modes), particularly at the edges of the parameter space. On the other hand, `MCMC-diffusion` is able to find most of the modes (samples produced in around 80% of modes), with the 1D marginal distributions showing roughly equal peaks across each parameter compared to pure MH. We found that pure MH would have to perform at least 5 times as many function evaluations to obtain similar samples as for `MCMC-diffusion` in Fig. 5.4, demonstrating a significant acceleration of the MCMC sampling process in this case. In this case, the diffusion model was seeded by providing 1,000 random points drawn from a uniform distribution.

It is also worth comparing our diffusion-accelerated MCMC algorithm to an algorithm that makes use of multiple chains, such as the widely-used `emcee` algorithm [237]. After running the default `emcee` algorithm on the EggBox function with 8 chains for a total of 0.1M samples, we created a comparison plot with `MCMC-diffusion`, see Fig. 5.5. We found that there was only a slight improvement over the single chain of pure MH, with around 60% of modes found. This demonstrates that `MCMC-diffusion` can outperform algorithms with multiple chains for heavily multi-modal functions.

5.4 Physics example: global parton distribution function fit

We now apply `MCMC-diffusion` and pure MH to a simple physics example intended to reflect the salient features of a global PDF fit.

We follow the same toy PDF design as described in Sec. 3.1.1, with a likelihood defined in terms of the χ^2 in Eq. 3.3

$$\mathcal{L} = \exp\left(-\frac{1}{2}\chi^2\right). \tag{5.11}$$

Using this likelihood, we can now sample the posterior using `MCMC-diffusion` and pure MH, and generate uncertainty bands for the cross sections and PDFs. The likelihood in this case has a single mode.

Figure 5.6 shows a comparison between `MCMC-diffusion` and pure MH for our 4D PDF physics example. We also include the uncertainty bands generated in the

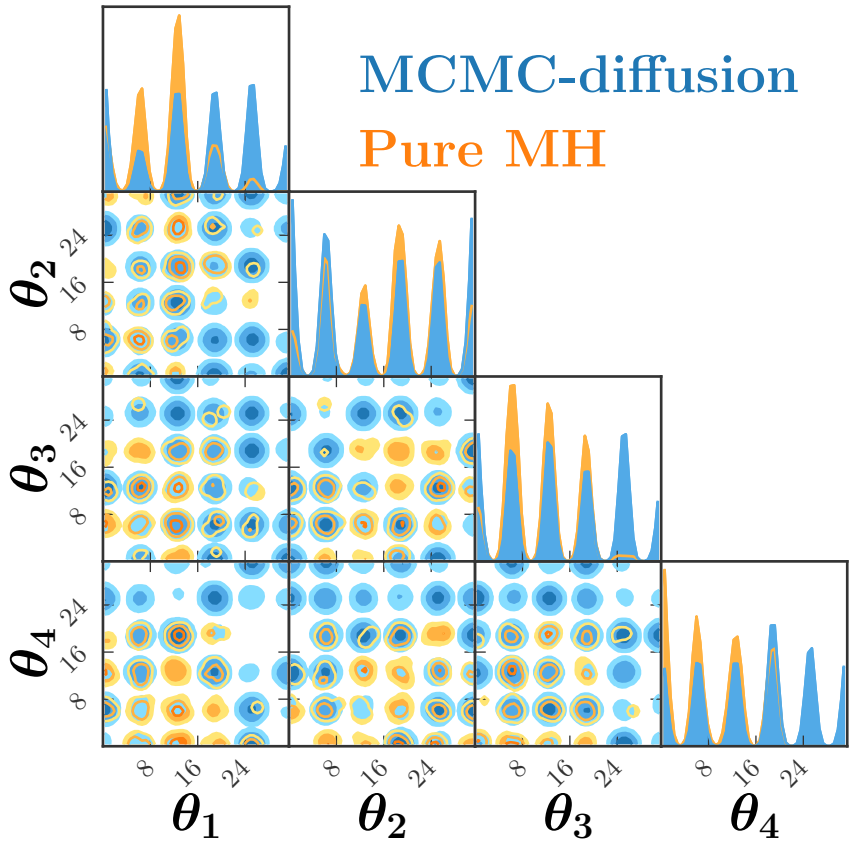


Figure 5.4: Comparison of MCMC-diffusion and pure MH for the 4D EggBox function (0.1M samples). Coloured contours represent sample density for each algorithm, with more identified modes representing greater success at characterising the shape of the true posterior.

limit of many samples, as an estimate of the true uncertainties. We found that both `MCMC-diffusion` and pure MH agree in this limit, meaning that our diffusion-augmented algorithm does converge to the purely Markovian distribution for this example. In general, one ought to be careful when using past samples taken from the existing chain as inputs to generate new samples as we do in `MCMC-diffusion`. For example, adaptive MCMC algorithms require certain conditions to be met in order to ensure ergodicity of the chain, see [238] for further details.

For the 10,000 samples shown here, pure MH heavily overestimates the uncertainty bands. On the other hand, pairing MH with the diffusion model results in samples that approximate the true uncertainty bands very closely. 3 times as many samples would be required for pure MH to produce a fit of comparable quality. One advantage the diffusion model has over pure MH is that it can make nonlocal jumps to key regions

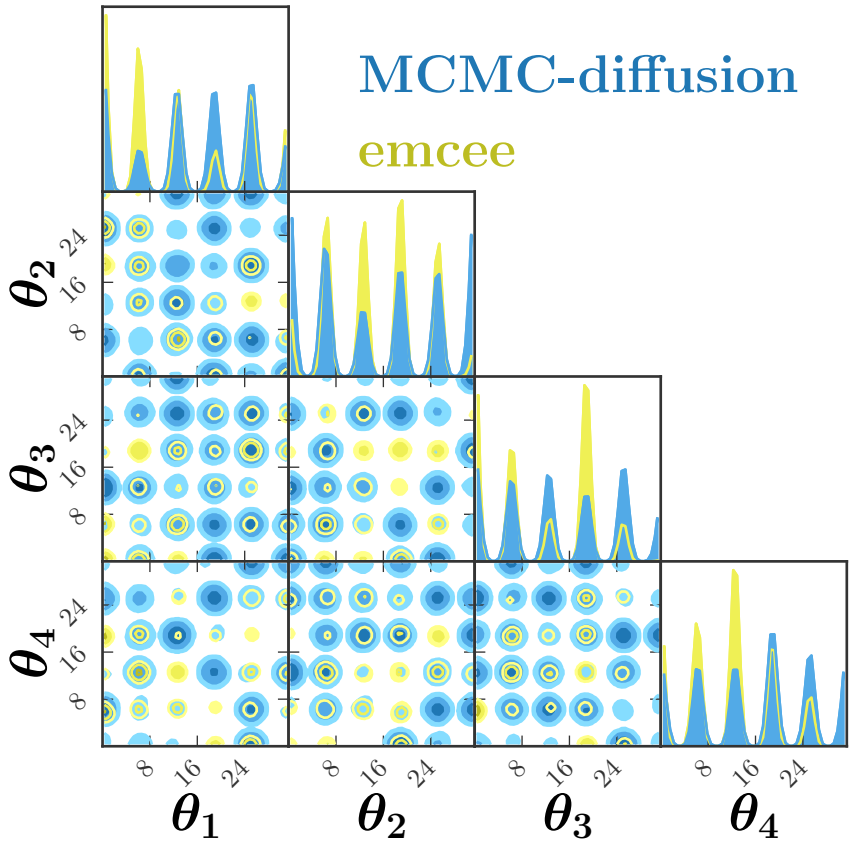


Figure 5.5: Comparison of MCMC-diffusion and the emcee algorithm for the 4D Egg-Box function (0.1M samples). Coloured contours represent sample density for each algorithm, with more identified modes representing greater success at characterising the shape of the true posterior.

of interest and avoid getting stuck in the tails of the distribution, which explains why `MCMC-diffusion` is able to converge to the truth in fewer likelihood evaluations even for a function with a single mode. The diffusion model was initially seeded with 100 points located near the mode.

5.5 Conclusions

We have developed a diffusion model tailored for low-dimensional data, pairing it with a Metropolis-Hastings sampler to create a new accelerated Markov Chain Monte Carlo algorithm. We find that `MCMC-diffusion` is many times faster than pure Metropolis-Hastings at obtaining accurate posterior samples for test functions with any number of

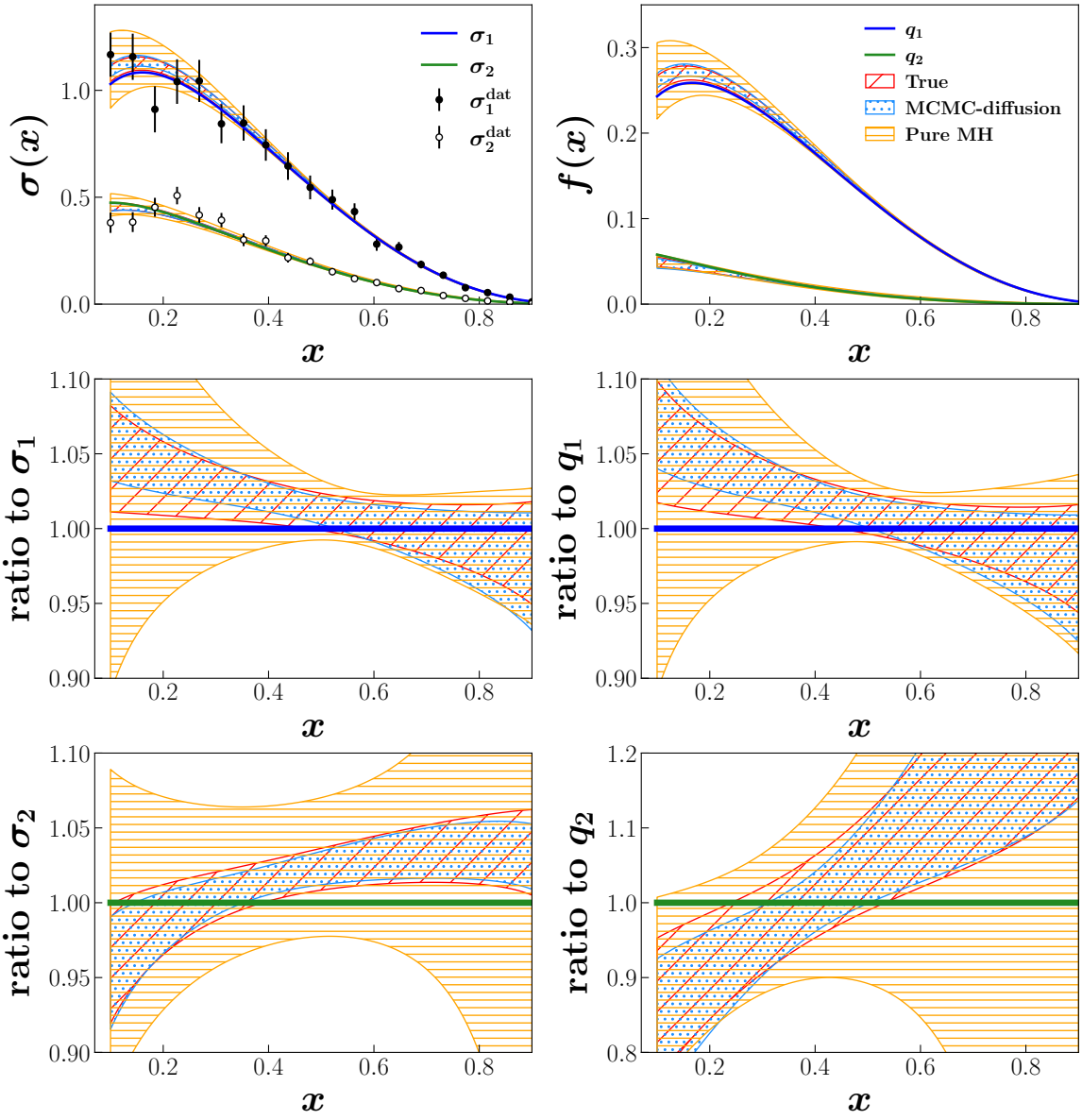


Figure 5.6: Comparison of fit results for MCMC-diffusion and pure MH (10,000 samples) to the many-sample limit (1M samples). The top row contains cross section and parton distributions, while the middle and bottom rows show ratios to the true values for each distribution.

modes, as evidenced by a significant reduction of the number of likelihood evaluations required to adequately explore the Himmelblau, 10D Gaussian mixture, and EggBox functions. `MCMC-diffusion` has also shown significant improvement over a normalizing

flow algorithm for a specific test function. In the context of a real physical example based on a global analysis of parton distribution functions, `MCMC-diffusion` is able to characterize uncertainty bands much more accurately than pure Metropolis-Hastings for a limited number of samples.

Our technique does not require gradient information, and is therefore applicable to a wide range of posteriors encountered in physics applications. Furthermore, our proof of principle indicates that a combination of a diffusion model with a more sophisticated MCMC or other sampler could be expected to yield even better results.

Part II

Beyond the Standard Model

Chapter 6

Global QCD analysis and dark bosons

Despite the enormous success of the Standard Model (SM) of nuclear and particle physics, it remains an incomplete theory, not least because of its inability to explain dark matter. One relatively simple addition that could be made to at least provide a portal to the dark sector would be to introduce a new massive U(1) gauge boson, either through kinetic mixing in the case of the dark photon [239, 240, 241, 242, 243, 244] or anomaly-free $U(1)'$ charges as for heavy Z' bosons [245, 246, 247, 248]. Such models have received considerable attention in connection with hints of possible new physics phenomena, including the value of $g-2$ for the muon [249, 250, 251], the CDF W boson mass anomaly [252, 253, 254, 255], and anomalies in rare kaon and B meson decays [256, 257, 258].

Numerous experimental searches for dark bosons have been undertaken in electron, neutrino, and proton fixed target experiments [259, 260], and at e^+e^- [261, 262] and hadron-hadron colliders [263, 264, 265]. No direct evidence has so far been found. Instead, these experiments have allowed rather strong constraints to be placed on the dark boson parameters. In the case of the dark photon, large regions of parameter space in both light and heavy mass regions have been ruled out [266]. Specifically, the parameter ϵ describes how often the dark photon mixes with the regular SM photon, and is constrained to be $> 10^{-3}$, with a few gaps associated with the production of resonances such as the J/ψ and its excited states. Further competitive constraints have recently been placed on the dark photon from "decay-agnostic" (independent of decay modes or production mechanism) processes, such as the muon $g-2$ anomaly [249, 250], the electroweak precision observables (EWPO) [267, 268, 269], $e^{\pm}p$ deep-inelastic scattering (DIS) [270, 271, 272], parity-violating electron scattering [273, 255], rare kaon and B-meson decays [256, 258], and high-luminosity LHC projections [274]. In the case of the Z' boson there also exist experimental constraints, with the NA64 Collaboration searching in the sub-GeV region [260], while the CMS Collaboration

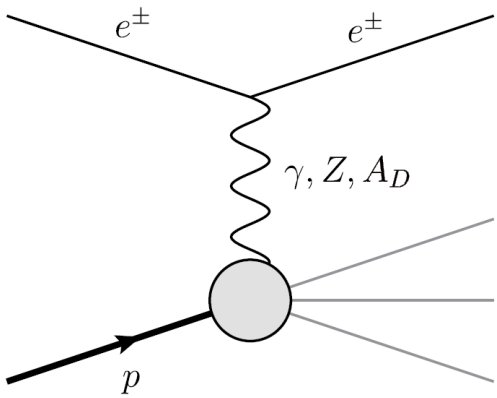


Figure 6.1: Kinetic mixing of the dark photon A_D with SM mediators γ and Z in $e^{\pm}p$ DIS [270].

placed strong constraints on a heavy Z', with mass above 200 GeV [265]. Several planned experiments, including those at the future Electron-Ion Collider (EIC) [275], will also provide opportunities to search for dark bosons [276, 277, 278].

In this chapter, we describe the first global QCD analysis including a dark photon within the JAM next-to-leading order (NLO) analysis framework, as well as the first global QCD analysis including a $U(1)_{B-L}$ gauge boson Z' within the same framework. The dark photon analysis is detailed in Section 6.1 and was published in [279]. The Z' analysis is detailed separately in Section 6.2, and was published in [280].

6.1 Dark photons

The dark photon is here chosen to mix kinetically with the SM hypercharge boson, requiring the additional Lagrangian terms [241],

$$\mathcal{L} \supset -\frac{1}{4}F'_{\mu\nu}F'^{\mu\nu} + \frac{m^2_{A'}}{2}A'_{\mu}A'^{\mu} + \frac{\epsilon}{2\cos\theta_W}F'_{\mu\nu}B^{\mu\nu}\,, \tag{6.1}$$

where A' denotes the unmixed version of the dark photon. The parameter ϵ describes the degree of mixing between the dark photon and the B boson of the standard electroweak theory, θ_W is the Weinberg angle, and $F'_{\mu\nu}$ is the dark photon field strength tensor. After electroweak symmetry breaking and diagonalizing the kinetic terms and gauge boson masses, three physical vectors remain which couple to the SM fermions: the massless photon γ, the massive Z boson, and the physical dark photon, labelled A_D.

The dark photon contributes to DIS processes coherently along with photon and Z boson exchange, as illustrated in Fig. 6.1. It has also been shown to be necessary to simultaneously extract the parton distribution functions (PDFs) from the data when incorporating beyond the SM (BSM) physics into proton structure [281, 282, 283, 284,

285]. An exploratory study [271] included an extraction of the PDFs alongside the dark photon contribution, although that analysis limited itself to a subset of the existing HERA and BCDMS data and employed only a basic leading order (LO) analysis of DIS hard scattering to extract PDFs.

We perform two different analyses in this section. In the first, we assume that the SM is the correct theory of Nature and present an exclusion limit in the plane of the dark photon mass and mixing parameters. In the second, we allow for the possibility that the dark photon may improve the likelihood for experimental observables relative to the SM, and simultaneously determine the optimum set of PDFs, as well as the preferred dark photon parameters. In scanning the parameter space, we consider all possible values for the dark photon mass, and do not make any assumptions about its numerical value. We then perform a hypothesis test against the SM to quantify the improvement at the best fit point. Having identified a preference for the dark photon model, we test the stability of this conclusion against the inclusion of power corrections, an increase in the lower cutoff in Q^2 from m_c^2 to 10 GeV^2, and an estimate of missing higher order uncertainties following Ref. [286].

In Sec. 6.1.1 we begin by reviewing the modifications to the F_2 proton structure function with the addition of a dark photon contribution. Section 6.1.2 outlines the methodological details of the global QCD analysis performed within the JAM framework. Our results are presented in Sec. 6.1.3, first placing exclusion limits on the dark photon parameters, and then exploring possible improvements of the global fit with the inclusion of the dark photon. Following a similar methodology as for the dark photon, we then turn to the Z' boson in Sec. 6.2, placing exclusion limits on the Z' parameters in the $B-L$ model. Finally, in Sec. 6.3 we summarize our conclusions and discuss the implications and future extensions of this work.

6.1.1 Dark photon background

The F_2 and F_3 proton structure functions, including the electroweak and dark photon contributions, are given by [270]

$$\begin{aligned} \widetilde{F}_2 &= \sum_{i,j=\gamma,Z,A_D} \kappa_i \kappa_j F_2^{ij}, \\ \widetilde{F}_3 &= \sum_{i,j=\gamma,Z,A_D} \kappa_i \kappa_j F_3^{ij}, \end{aligned} \tag{6.2}$$

where $\kappa_i = Q^2/(Q^2+m_i^2)$. At LO in the strong coupling α_s, one has

$$\begin{aligned} F_2^{ij} &= \sum_q (C_{i,e}^v C_{j,e}^v + C_{i,e}^a C_{j,e}^a)(C_{i,q}^v C_{j,q}^v + C_{i,q}^a C_{j,q}^a)\, x f_q, \\ F_3^{ij} &= \sum_q (C_{i,e}^v C_{j,e}^a + C_{i,e}^a C_{j,e}^v)(C_{i,q}^v C_{j,q}^a + C_{i,q}^a C_{j,q}^v)\, x f_q, \end{aligned} \tag{6.3}$$

where x is the parton momentum fraction, and f_q is the PDF for quark flavor q in the proton. The vector and axial vector couplings to the electron and quarks for the photon are

$$\{C^v_{\gamma,e}, C^v_{\gamma,u}, C^v_{\gamma,d}\} = \left\{-1, \frac{2}{3}, -\frac{1}{3}\right\}, \quad C^a_\gamma = 0, \tag{6.4}$$

while for the unmixed Z boson the couplings are

$$\overline{C}^v_Z \sin 2\theta_W = T^f_3 - 2q_f \sin^2\theta_W, \quad \overline{C}^a_Z \sin 2\theta_W = T^f_3, \tag{6.5}$$

where T^f_3 is the third component of the weak isospin, and q_f the electric charge. After diagonalizing the mixing term through field redefinitions, the couplings of the physical Z and A_D to SM particles are given by [270]

$$\begin{aligned} C^v_Z &= (\cos\alpha - \epsilon_W \sin\alpha)\,\overline{C}^v_Z + \epsilon_W \sin\alpha \cot\theta_W\, C^v_\gamma, \\ C^a_Z &= (\cos\alpha - \epsilon_W \sin\alpha)\,\overline{C}^a_Z, \end{aligned} \tag{6.6}$$

and

$$\begin{aligned} C^v_{A_D} &= -(\sin\alpha + \epsilon_W \cos\alpha)\,\overline{C}^v_Z + \epsilon_W \cos\alpha \cot\theta_W C^v_\gamma, \\ C^a_{A_D} &= -(\sin\alpha + \epsilon_W \cos\alpha)\,\overline{C}^a_Z. \end{aligned} \tag{6.7}$$

Here α is the $\bar{Z}$–A' mixing angle,

$$\begin{aligned} \tan\alpha = \frac{1}{2\epsilon_W}\Big[1 - \epsilon^2_W - \rho^2 \\ - \text{ sign}\,(1-\rho^2)\sqrt{4\epsilon^2_W + (1-\epsilon^2_W-\rho^2)^2}\Big], \end{aligned} \tag{6.8}$$

with ϵ_W given in terms of the free mixing parameter ϵ,

$$\epsilon_W = \frac{\epsilon \tan\theta_W}{\sqrt{1 - \epsilon^2/\cos^2\theta_W}}, \tag{6.9}$$

and ρ is defined by

$$\rho = \frac{m_{A'}/m_{\bar{Z}}}{\sqrt{1 - \epsilon^2/\cos^2\theta_W}}. \tag{6.10}$$

The physical masses of the Z boson and dark photon then become

$$\begin{aligned} m^2_{Z,A_D} = \frac{m^2_{\bar{Z}}}{2}\Big[1 + \epsilon^2_W + \rho^2 \\ \pm \text{ sign}\,(1-\rho^2)\sqrt{(1+\epsilon^2_W+\rho^2)^2 - 4\rho^2}\Big], \end{aligned} \tag{6.11}$$

with $\bar{Z}$ the unmixed version of the SM neutral weak boson. In our analysis we include the two dark parameters, ϵ and the mass m_{A_D}, amongst the fitting parameters.

6.1.2 Methodology

Our baseline study uses the JAM QCD analysis framework as described in Sec. 4.2. To accurately characterize the PDFs and their uncertainties, as well as the dark photon parameters, the global analysis was performed using 200 replicas from the most recent comprehensive JAM fit [55].

For inclusive DIS as described in Sec. 1.4, QCD factorization theorems allow us to write the structure functions as sums of convolutions of hard scattering functions, non-perturbative quark and gluon PDFs, and higher twist power corrections, see Eqs. 1.26 and 1.27. The hard scattering coefficient functions are computed in fixed order perturbation theory at next-to-leading order accuracy in the QCD coupling, following the "track B" approach of Refs. [287, 288]. At this order, the leading contributions to C_{2j} are of order $\mathcal{O}(\alpha_s^0)$, while C_{Lj} enters at order $\mathcal{O}(\alpha_s)$. We parametrize the unpolarized PDFs at the input scale Q_0^2 using the standard form in Eq. 1.15, in this case allowing the strange PDFs s and $\bar{s}$ to be free so that each individual fit determines 30 free proton PDF parameters. The $C_i^{\rm HT}(x)$ higher twist coefficients can in principle be different for protons and neutrons, and are determined phenomenologically from low-Q^2 data by introducing a further 6 free parameters to be fitted. In this analysis, we assume a multiplicative *ansatz* for the coefficients $C_i^{\rm HT}$, with the higher twist parameters taken to be the same for the longitudinal structure function F_L as for F_2. The definitions of the structure functions in Eqs. (**??**) also contain target mass corrections implemented within the collinear factorization as described in Refs. [289, 290].

In the nuclear impulse approximation at $x \gg 0$, the scattering takes place incoherently from individual (off-shell) nucleon contributions in the nucleus. As such, we can write the nuclear PDFs as a sum of on-shell and off-shell contributions [291, 292, 293],

$$q_A(x,Q^2) = \sum_N q_{N/A}(x,Q^2) = \sum_N \left[q_{N/A}^{(\rm on)} + q_{N/A}^{(\rm off)}\right](x,Q^2), \tag{6.12}$$

where $q_{N/A}$ refers to the PDF of a quark q in a nucleon N, as modified within a nucleus A. We introduce 6 more free parameters to account for these off-shell contributions, which for our chosen datasets only takes the form of small corrections to the deuterium data. Overall, this results in a total of 42 free PDF parameters in our baseline analysis.

The DIS datasets included in this analysis consist of fixed target proton and deuteron DIS data from SLAC [294], BCDMS [295], and NMC [296, 297], and collider data from HERA [298]. For the fixed target DIS data the F_2 structure function is provided, while for the HERA data the reduced cross sections are presented. At lower x and Q^2 values, the longitudinal contribution to the cross section should be taken into account, while in the kinematics of the HERA data included in our analysis the F_L contribution to the reduced cross sections, is negligible. However, we still include the F_L contribution, especially in a precise determination of the dark photon parameters whose effect is also expected to be small. For all DIS datasets, a Q^2 cut of 1.69 GeV^2 and a W^2 cut of

10 GeV^2 were employed, as in Ref. [55]. Other datasets that were added to constrain the PDF parameters include pp and pd Drell-Yan data from the Fermilab NuSea [299] and SeaQuest [300] experiments, Z-boson rapidity data [301, 302], W-boson asymmetry data [303, 304], and jet production data from $p\bar{p}$ collisions at the Tevatron [305, 306].

The modifications required to include the effects of the dark photon were then added to the underlying JAM theory, allowing the two additional dark parameters, m_{A_D} and ϵ, to be fitted alongside the PDF parameters. After repeating the global fits, we then compared the 200 replicas obtained with and without the dark photon. The agreement between the fitted results and the data is once again assessed through the reduced χ^2, see Eq. 4.42.

Because of its particular sensitivity to a dark photon, we included an additional contribution to the total χ^2 corresponding to the value of $g-2$ for the muon, for which there is currently an anomaly [307, 308]. The SM predicts a result for $g-2$ that differs from the experimental measurement by $(251 \pm 59) \times 10^{-11}$ (which has a significance of 4.2σ). While there are indications from lattice QCD [309] for a somewhat higher SM value, there is no consensus on this in the lattice community, nor a successful reconciliation with the dispersion relation method. We therefore use the value reported in Ref. [310] and quoted in Refs. [307, 308]. In the dark photon scenario the correction to $g-2$ scales as $m_\ell^2\,\epsilon^2/m_{A_D}^2$ (with m_ℓ the lepton mass) [249], meaning that the experimental measurement can act as a powerful constraint on the dark photon parameters.

6.1.3 Global QCD analysis results

We first explore the region of parameter space where a χ^2 analysis forbids the existence of the dark photon, assuming that the SM is the true theory. This is accomplished by not allowing the χ^2 to be lower than that obtained with the best fit SM hypothesis, and is similar to the "capped likelihood" treatment followed in Ref. [311]. The excluded region is shown in Fig. 6.2, where we see that, for $m_{A_D} > 20$ GeV, ϵ must be less than ≈ 0.08. In the region around 10 GeV, the mixing parameter ϵ is allowed to be notably larger due to the dark photon fitting anomalies in some datasets, whilst being penalized by others sufficiently to give the same quality of fit as the SM. This weakens the excluded range of ϵ.

We also note that the exclusion limits over the whole mass range for ϵ in Fig. 6.2 are somewhat weaker than those reported in earlier analyses of DIS data [270, 271]. We first observe that the strong constraints derived in Ref. [270] are mainly driven by: (i) the input PDFs were fixed at the values resulting from the best fit results of a HERA analysis, leaving no room for the potential effects of new physics on the extracted PDFs; (ii) the uncertainties in the reduced neutral current (NC) cross section $\sigma_{\rm red}^{\rm NC}$ were assumed to be at most 0.3%–0.4%, which is approximately 4 times smaller

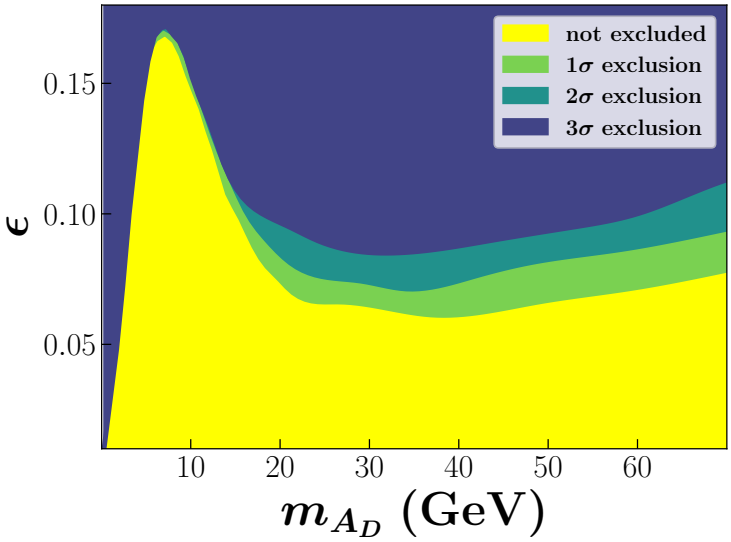


Figure 6.2: The excluded region in the dark photon parameter space, assuming that the SM is the correct theory of Nature.

than those of the raw data [298]. Moreover, the longitudinal structure functions were neglected in Ref. [270], which is a good approximation only when $y = Q^2/xs$ is small. In our work, we have analyzed the original HERA data including both the transverse and longitudinal structure functions, and have allowed the PDFs to adjust to the presence of the dark photon. Compared with Ref. [271], the dataset used here is considerably larger and the restrictions imposed on the PDF parameters in that work have been removed.

Next, we present results of a fit where the dark photon is allowed to improve the fit to the experimental observables. A comparison of the χ^2 per degree of freedom between the baseline average theory and the dark photon average theory for the different datasets included is given in Table 6.1. There is a clear improvement in the reduced χ^2 for the fixed target DIS and HERA datasets with the inclusion of the dark photon. This improvement was essentially independent of the number of replicas considered. All the datasets, other than the Drell-Yan, Z rapidity and jet data, exhibit a slightly lower χ^2 with the dark photon, although no modifications were made to the underlying theory for those observables. In fact, electroweak corrections to the Drell-Yan cross sections are small and generally omitted in global PDF analyses. However, as a check, we did include the dark photon contribution, which yielded a negligible improvement in χ^2 of 0.003 per degree of freedom for that dataset. The sheer number of DIS data points means that the total χ^2 shows a substantial improvement. The best fit including the dark photon also reduces the difference in the anomalous value of $g-2$ between the SM prediction and experiment from 251×10^{-11} to 91×10^{-11}, which in turn reduces

Table 6.1: Comparison of the χ^2 values per degree of freedom, $\chi^2_{\rm dof}$, with ("dark") and without ("base") dark photon modifications for various datasets, taking into account a correction for missing higher order uncertainties (see text).

reaction	$\chi^2_{\rm dof}$(dark)	$\chi^2_{\rm dof}$(base)	$N_{\rm dof}$
fixed target DIS	1.01	1.05	1495
HERA NC	1.02	1.03	1104
HERA CC	1.13	1.18	81
Drell-Yan	1.18	1.16	205
Z rapidity	1.08	1.05	56
W asymmetry	1.04	1.07	97
jets	1.16	1.15	200
total	1.03	1.05	3283

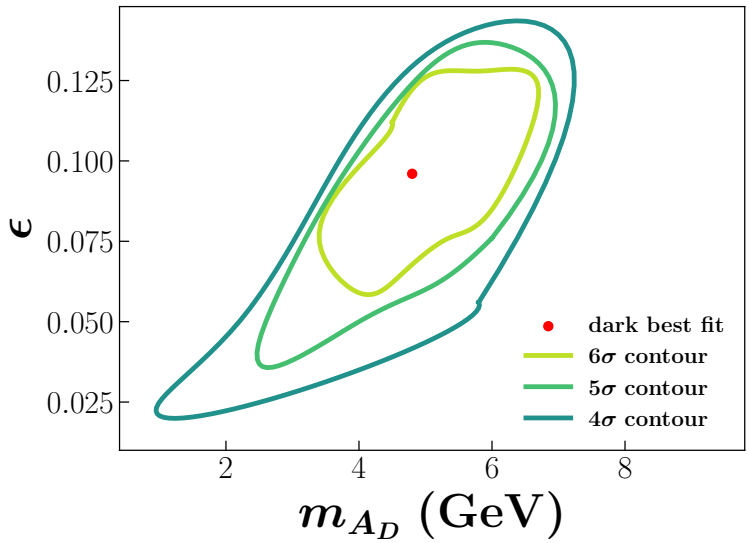


Figure 6.3: Results of an hypothesis test for the likelihood that the SM is the correct theory to describe this data, compared with the case where a dark photon is included. The hypothesis that the SM is the correct theory is excluded at 6.5σ for the best dark photon fit at the red point.

the statistical significance of the discrepancy in that observable from 4.2σ to 1.5σ.

In order to assess the statistical significance of this χ^2 improvement, we performed an hypothesis test of the dark photon hypothesis against the SM-only hypothesis. Since these hypotheses are nested we can make use of the log-likelihood test statistic, defined

as

$$t = -2\left[\ln \mathcal{L}(\hat{\nu}, m_{A_D}, \epsilon) - \ln \mathcal{L}(\hat{\nu}, \hat{m}_{A_D}, \hat{\epsilon})\right], \tag{6.13}$$

where $\mathcal{L}(\hat{\nu}, m_{A_D}, \epsilon)$ is the likelihood for the SM-only hypothesis (or "baseline"), meaning a choice of $m_{A_D} = \infty$ and $\epsilon = 0$, maximized over the nuisance PDF parameters $\hat{\nu}$. The likelihood $\mathcal{L}(\hat{\nu}, \hat{m}_{A_D}, \hat{\epsilon})$ is then maximized over the dark parameters, in addition to the PDF parameters. Using our choice of likelihood function $\mathcal{L} = \exp\left(-\frac{1}{2}\chi^2\right)$, the test statistic reduces to the difference of the χ^2 values between the baseline and dark best fits, $t = \chi^2_{\text{base}} - \chi^2_{\text{dark}}$. The results of the fit are shown in Fig. 6.3, and applying Wilks' theorem to the t test statistic, the difference in total χ^2 for the best fit corresponds to a p-value of $\sim 3.8 \times 10^{-11}$ or $\sim 6.5\sigma$. We checked the validity of applying Wilks' theorem to our results, confirming that we do indeed have a χ^2-distributed likelihood ratio test statistic. This implies that the quoted p-value should be a good approximation.

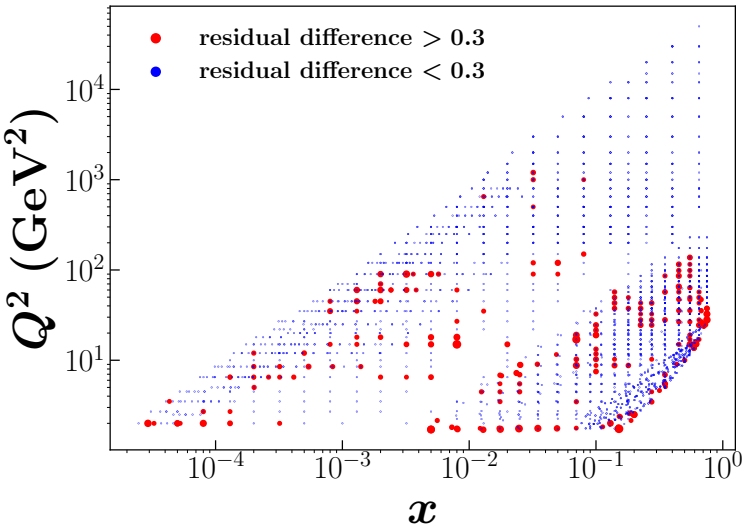


Figure 6.4: Points in DIS kinematics (Q^2, x) exhibiting the greatest improvement in residuals (> 0.3 in red, < 0.3 in blue) between dark and baseline average theory.

There is no localized region of kinematics which contains the greatest difference in residuals, as seen in Fig. 6.4. The marker size of the data points corresponds to an improvement in the absolute value of the residual less than 0.3 set to a very small size and shown in blue, while an improvement larger than this is shown by a larger point in red. The improvement associated with the inclusion of the dark photon covers most of the range of Q^2 values for each given x, as shown by the spread of red points in Fig. 6.4. This indicates that the improvement in χ^2 seen with our dark photon modifications is unlikely to be explained by some missing element in the QCD theory that contributes via a specific region of Q^2 or x.

The inclusion of next-to-next-to-leading order (NNLO) corrections has been shown to improve the fit for HERA data [312] to an extent that could be comparable with that reported here by including the dark photon modifications. As a quantitative test of the potential effects of higher order corrections, or so-called missing higher order uncertainties (MHOU), we follow the procedure developed by the NNPDF Collaboration [286]. That is, we add in quadrature to the experimental uncertainty on each data point the largest variation in the theoretical prediction obtained by varying the Q^2 for that point by a factor of anywhere between 0.5 and 2.0. This reduced the total χ^2 by roughly 250, but still left a very significant improvement when the dark photon was added. The results for the χ^2 per degree of freedom shown in Table 6.1 include this additional uncertainty.

As a test of the effect of power corrections on the improvement in χ^2 associated with the dark photon, we repeated the analysis with power corrections turned off. The total χ^2 increased by around 16, but the reduction associated with the dark photon was unchanged. To check the effects of data at relatively low-Q^2, we also repeated the analysis with a lower bound of 10 GeV^2 instead of 1.69 GeV^2. This reduced the number of data points significantly but the improvement in χ^2 was still such as to favor the existence of the dark photon by 5σ.

In Fig. 6.5 we compare the PDFs from the dark photon and baseline replicas. The dark PDFs overlap with the baseline PDF uncertainty bands most of the time, particularly for the valence quark distributions that are best constrained by existing data. We also found that both the dark and baseline PDFs are similar to the most recent JAM22 unpolarized PDF results [55], up to some small differences in the W^2 cut used in that analysis. This suggests that the dark photon results are indeed consistent, within uncertainties, with existing PDF analyses, albeit with small but potentially interesting shifts in the central values.

Figure 6.6 illustrates the profile likelihood in the ϵ–m_{A_D} plane, with contours indicating 1σ, 2σ, 3σ and 4σ levels. The ranges of the dark parameters favored by the data give a dark photon mass $m_{A_D} = 4.8^{+1.5}_{-1.2}$ GeV and mixing parameter $\epsilon = 0.096^{+0.028}_{-0.028}$ at the 95% confidence level. This value of ϵ appears somewhat high, as $\epsilon > 10^{-3}$ has been largely excluded for a wide range of mass values [266]. On the other hand, there do exist regions of parameter space within this mass range which cannot be excluded because of the presence of other particles, notably the J/ψ and its excited states between ≈ 3 GeV and 4.5 GeV.

The preferred value for the dark photon mass has an uncertainty range that is wide enough to encompass these unconstrained regions, meaning that our fitted values for the dark parameters are not inconsistent with existing experiments. In fact, the improvement in χ^2 is so substantial that if we perform the hypothesis test at the point $\epsilon = 0.05$ and mass $m_{A_D} = 3$ GeV, the SM-only hypothesis is still excluded with a significance above 5σ.

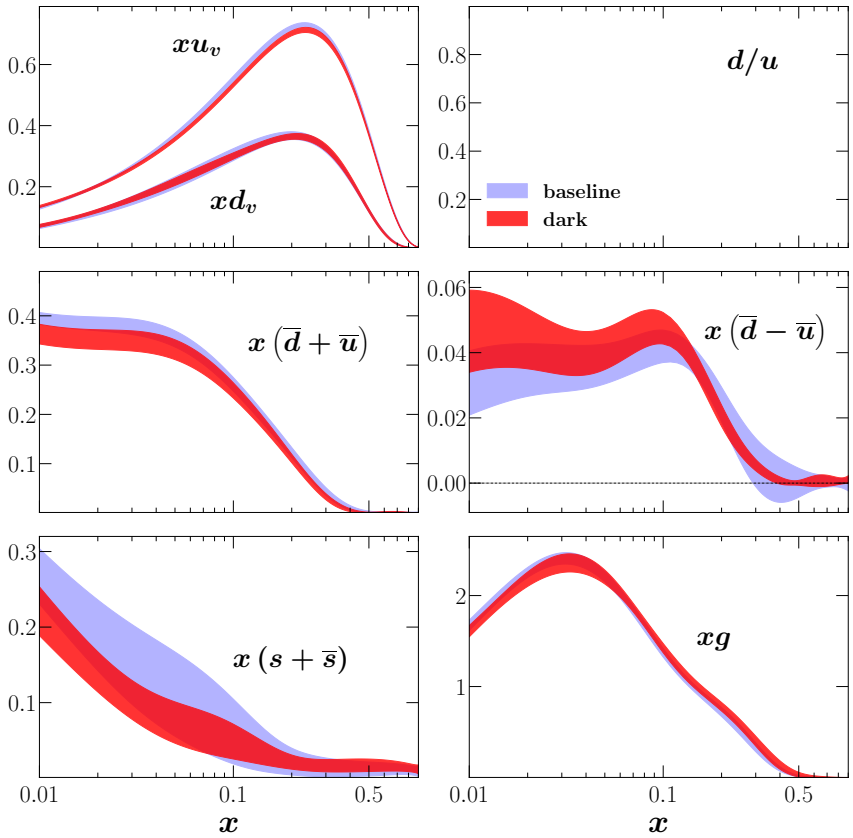


Figure 6.5: A comparison of the PDF replicas with ("dark", red bands) and without ("baseline", blue bands) the dark photon modifications at the input scale, $Q^2 = 1.69\ \mathrm{GeV}^2$.

Finally, it is interesting to comment on the parameters preferred in this analysis in comparison with other existing constraints. Constraints from EWPO [267, 268] typically require a mixing parameter for a dark photon somewhat below our preferred range, in the relevant mass region. These constraints ought to be taken seriously, given the recent resolution of the W-boson mass anomaly from both ATLAS and CMS [313, 314]. In the case of $g-2$ for the electron our results give no significant change to the current satisfactory situation of experiment versus theory [315, 316], with the uncertainty in α_{em} being dominant.

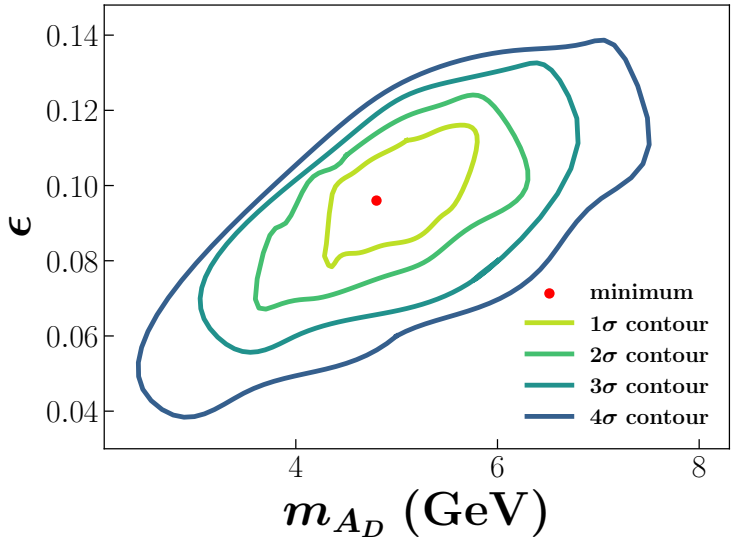


Figure 6.6: Profile likelihood for the dark photon mass m_{A_D} versus the mixing parameter ϵ.

6.2 Z' bosons

Turning now to the Z' boson, we focus on the $B-L$ model. This model is phenomenologically interesting because the gauge boson Z' does not mix with the Standard Model neutral gauge bosons at tree level [247], and is thus less constrained by electroweak precision observables [267, 268, 269]. We proceed in a similar fashion as for the dark photon, aiming to characterize the modifications to the F_2 and F_3 structure functions with the addition of a Z' boson.

In the $U(1)_{B-L}$ model, the $U(1)'$ charges of the Standard Model fermions are proportional to their baryon number B minus their lepton number L. The anomaly free, $B-L$ gauge invariant Lagrangian density describing the interactions between a Z' boson and Standard Model fermions takes the form [247]

$$\mathcal{L}_{\text{int}}^{B-L} = -g_z \sum_{i=1}^{3} \left(\frac{1}{3} \bar{q}_i \gamma^\mu q_i - \bar{\ell}_i \gamma^\mu \ell_i \right) Z'_\mu \,, \tag{6.14}$$

where q_i and ℓ_i denote the i-th generation of quark and lepton fields, respectively, and g_z is the coupling constant. The inclusive spin-averaged DIS cross section is usually parametrized in terms of nucleon structure functions, which, including electroweak and Z' contributions, can be written as

$$\widetilde{F}_n = \sum_{i,j=\gamma,Z,Z'} \kappa_i \kappa_j \, F_n^{ij}, \qquad n = 1,2,3, \tag{6.15}$$

where $\kappa_i = Q^2/(Q^2 + M_i^2)$ and M_i is the boson mass. At leading order (LO) in the strong coupling α_s, the structure functions can be simply written as

$$F_2^{ij} = 2xF_1^{ij} = \sum_q \left(C_{i,e}^v C_{j,e}^v + C_{i,e}^a C_{j,e}^a\right)\left(C_{i,q}^v C_{j,q}^v + C_{i,q}^a C_{j,q}^a\right) x f_q, \tag{6.16a}$$

$$xF_3^{ij} = \sum_q \left(C_{i,e}^v C_{j,e}^a + C_{i,e}^a C_{j,e}^v\right)\left(C_{i,q}^v C_{j,q}^a + C_{i,q}^a C_{j,q}^v\right) x f_q\,, \tag{6.16b}$$

where f_q is the parton distribution function (PDF) for quark flavor $q = u, d, s, \ldots$ in the proton with partonic momentum fraction x (which at LO is equal to the Bjorken scaling variable). The actual QCD analysis is performed at next-to-leading order (NLO) accuracy in α_s. For the photon, the vector and axial-vector couplings to the electron, u-type and d-type quarks are given, respectively, by

$$\left\{C_{\gamma,e}^v, C_{\gamma,u}^v, C_{\gamma,d}^v\right\} = \left\{-1, \frac{2}{3}, -\frac{1}{3}\right\}, \tag{6.17a}$$

$$C_\gamma^a = 0\,, \tag{6.17b}$$

in units of $e = \sqrt{4\pi\alpha}$. The Standard Model couplings of the Z boson are not modified by the Z', so that in this case the vector and axial-vector couplings are

$$\left\{C_{Z,e}^v, C_{Z,u}^v, C_{Z,d}^v\right\} \sin 2\theta_W = \left\{-\frac{1}{2} + 2\sin^2\theta_W, \frac{1}{2} - \frac{4}{3}\sin^2\theta_W, -\frac{1}{2} + \frac{2}{3}\sin^2\theta_W\right\}, \tag{6.18a}$$

$$\left\{C_{Z,e}^a, C_{Z,u}^a, C_{Z,d}^a\right\} \sin 2\theta_W = \left\{-\frac{1}{2}, \frac{1}{2}, -\frac{1}{2}\right\}, \tag{6.18b}$$

where θ_W is the Weinberg or weak mixing angle. The Z' boson only has vector couplings to Standard Model fermions,

$$\sqrt{4\pi\alpha}\left\{C_{Z',e}^v, C_{Z',u}^v, C_{Z',d}^v\right\} = \left\{-g_z, \frac{1}{3}g_z, \frac{1}{3}g_z\right\}. \tag{6.19}$$

The couplings of the dark photon and Z' boson to the Standard Model fermions (electron, u-type quark and d-type quark) are shown in Table 6.2, where ϵ is the kinetic mixing parameter for the dark photon [241]. While the dark photon couplings are proportional to the fermion electric charges, the Z' boson has universal couplings to all quark flavors [247].

For a meaningful comparison with the dark photon model, we maximally keep the same settings and datasets as those in Section 6.1. In contrast to the dark photon analysis, no improvement in χ^2 was found with respect to the baseline result when the Z' boson was included in the global fit. From this result exclusion constraints on the coupling g_z to the Z' boson can be placed at the 95% confidence level (CL) by imposing an upper bound on χ^2 [317],

$$\chi^2_{Z'}(g_z) - \chi^2_{\text{baseline}} = 3.8\,, \tag{6.20}$$

Table 6.2: Couplings of the dark photon and Z' boson to the Standard Model fermions. The second column only shows the vector couplings, as the effect of the axial-vector couplings is negligible.

	A_D coupling	Z' coupling
u-quark	$\frac{2}{3}\, e\, \epsilon$	$\frac{1}{3} g_z$
d-quark	$-\frac{1}{3}\, e\, \epsilon$	$\frac{1}{3} g_z$
electron	$-e\, \epsilon$	$-g_z$

in the case of one degree of freedom (for a given $M_{Z'}$ value).

In Fig. 6.7, we show the 95% CL exclusion limits on g_z for a Z' boson with mass in the range between $M_{Z'} = 2$ GeV and 160 GeV. In this mass region, BaBar gave strong limits on the mixing parameter ϵ for the dark photon with mass below 10 GeV [261, 262]. Those limits can be converted to constraints on the coupling g_z by a rescaling factor $e = \sqrt{4\pi\alpha}$. However, one should bear in mind that these limits could be significantly relaxed in light of the potential couplings of the dark boson to dark matter particles [318]. Neutrino-quark scattering placed limits of order $M_{Z'}/g_z \gtrsim 1$ TeV for $M_{Z'} > 10$ GeV [319], while a more conservative bound of $M_{Z'}/g_z > 0.4$ TeV was adopted in Ref. [320], taking into account systematic errors in the NuTeV measurements [321, 322, 323]. The strongest constraints on the Z' coupling come from Big Bang nucleosynthesis (BBN), albeit within a specific class of E_6 models in which light right-handed neutrinos couple to the Z' boson [324]. Within that framework the limits on g_z are given by $M_{Z'}/g_z > 6.7$ TeV for $M_{Z'} > 10$ GeV and strengthen significantly down to $g_z < 6 \times 10^{-9}$ at $M_{Z'} = 1$ GeV [320]. As Barger *et al.* [324] observe, there are a number of ways to evade these limits, so we do not show them explicitly in Fig. 6.7.

We have also checked the flavored $B_1 - L_1$ model, in which the Z' boson only couples to the first generation of leptons and quarks. We repeated the global fit to the high-energy scattering data, with the muon data playing no role in constraining the Z' parameters. The resulting exclusion limits on g_z at 95% CL become much weaker, with the upper bounds being of $\mathcal{O}(0.1)$.

6.3 Outlook

In summary, we have mapped out the region of dark photon and Z' parameters excluded by existing high energy scattering data. Compared with earlier analyses, the present work constitutes a clear improvement, using all of the available data and allowing for variations in the PDFs associated with the introduction of the dark photon.

On the other hand, introducing the possibility of a dark photon yields a preference for the dark photon model of 6.5σ. The inclusion of the dark photon also leads to a

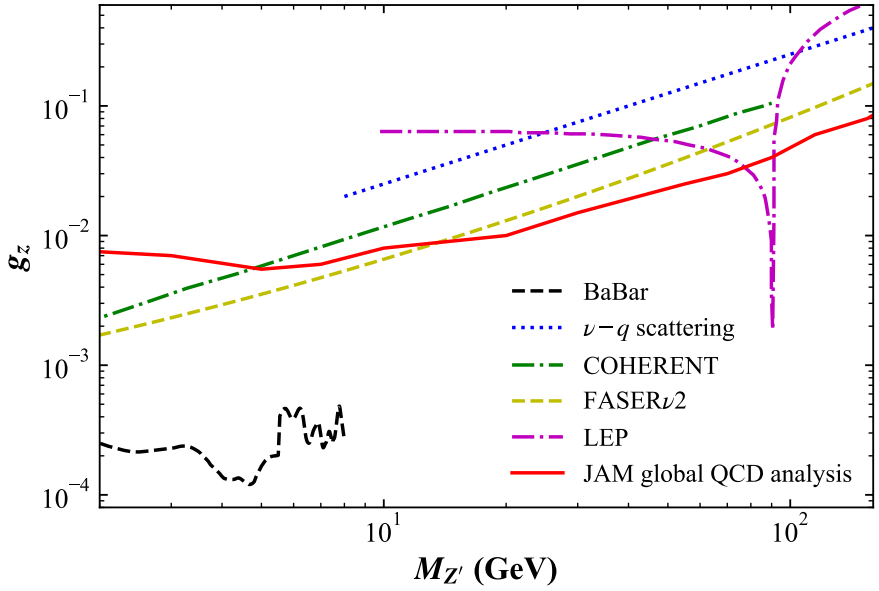


Figure 6.7: Exclusion limits for g_z at the 95% CL (90% CL for BaBar). The BaBar constraints are derived by rescaling the limits of the dark photon mixing parameter ϵ [262]. The upper bound on g_z from neutrino–quark ($\nu - q$) scattering is taken from Ref. [320].

significant reduction in the size of the muon $g-2$ anomaly. We have tested the viability of this hypothesis against possible missing higher order uncertainties, as well as the treatment of power corrections and the lower cutoff applied on Q^2, with the result that none of these qualitatively change the conclusions.

In the future our analysis could be improved by systematically computing all observables at next-to-next-to-leading order (NNLO) instead of the next-to-leading order accuracy employed in the current analysis. We also plan to implement different heavy-quark scheme for the discussion of charm and bottom quark production data. The most important improvement would, however, be direct searches in the mass region suggested by our analysis.

Chapter 7

A comparison of Bayesian sampling algorithms for particle astrophysics applications

Bayesian inference has been a crucial tool underpinning much of the work we have presented in this thesis. We now zoom out to view the landscape of the many Bayesian sampling algorithms out there, with new algorithms and accompanying literature being generated at a consistent rate. For users in the field, it is not easy to determine which technique to use for a given problem, despite the choice having major implications for the accuracy of the final results and the efficiency with which they are obtained. In this chapter, we therefore compare and contrast a wide selection of Markov chain Monte Carlo and nested sampling techniques on analytic test functions and real physics examples, in order to see which approaches work best for which problems. This allows users with physical insight on the expected shape of their posterior to select the most promising technique. Throughout, we aim to compare techniques from the position of the typical user — we do not therefore perform extensive optimisation of the hyperparameters of different algorithms, but instead stick close to default settings in order to compare their relative efficacy under reasonable test conditions.

This study is intended as a Bayesian equivalent to the work in [211], where various optimization algorithms were surveyed in the context of frequentist inference. The real physics examples we explore here are a global fit of the ΛCDM model, as well as a global fit of the Minimal Supersymmetric Standard Model (see Sec. 1.1 for a short introduction to both models). These examples demonstrate the crucial role Bayesian sampling has in BSM physics searches, as well as PDF analyses with both SM and BSM physics as have been considered earlier in this thesis, and we anticipate that our findings may help guide future usage of Bayesian algorithms in such scenarios.

This chapter is structured as follows. In section 7.1, we provide a description of each of the algorithms we use in our study. In section 7.2 we explain the metrics that

we use to compare them, and in section 7.3 we describe the examples that we use for our comparisons and detail the results. Finally, we present conclusions in section 7.4. The appendices include supplementary materials related to the metrics we have used, as well as posterior means and standard deviations for the test functions. The contents of this chapter are currently in preprint [325].

7.1 Sampling algorithms

Due to their current dominance in particle astrophysics and cosmology applications, we consider variants of two classes of Monte Carlo (MC) algorithms to draw samples from the posterior: Markov chain Monte Carlo (see Sec. 2.2) and nested sampling (see Sec. 2.4). Whilst both MCMC and NS allow for parameter inference, their goals and thus their performance characteristics are somewhat different. NS was designed to allow one to make statistical estimates of compression, which appears in Bayesian model selection through the Bayesian evidence and in frequentist computation [326]. MCMC, on the other hand, allows for parameter inference without ever estimating compression. Computing compression ratios is fundamentally more computationally challenging, so it is generally accepted that NS algorithms command a higher computational cost.

MCMC runs are typically divided into two parts: a warm-up phase, that reduces bias, and a sampling phase, that eliminates noise [327]. In the sampling phase, new posterior samples can be continually generated at a fixed computational cost. NS runs, on the other hand, compress from the prior to the bulk of the posterior mass and beyond. Consequently, new posterior samples cannot be generated at a fixed computational cost by running NS for longer and incomplete runs are of limited value, suffering from both bias (a truncation error) and noise. New NS runs, on the other hand, would waste computational effort compressing to the bulk of the posterior mass. Dynamic NS [328] solves this problem by rewinding a completed NS run back to the bulk of the posterior mass and starting an independent run from that point. This amortizes the cost of compressing to the bulk of the posterior.

It should be noted therefore that NS algorithms are generally understood to be considerably more expensive than MCMC algorithms as a method for generating posterior samples, and that very little will beat a well-tuned MCMC. This is because NS algorithms are designed to compute the evidence, which requires compressing from prior through posterior and onto the peak of the likelihood. MCMC algorithms on the other hand cannot compute the evidence alone (as they do not estimate compression), although other methods exist to augment these such as `MCEvidence` [329], `FloZ` [330] or `harmonic` [331]. In practice, NS is used as a posterior sampler when the problem is multimodal, which can cause challenges for MCMC algorithms. In this paper, we demonstrate that modern MCMC algorithms can now be effective on multimodal problems, bringing the efficiency of MCMC to bear on problems that previously only NS

could tackle.

Among our total list of algorithms are two significant variants to the MCMC and NS paradigms. First, we consider an adaptive MCMC algorithm with a diffusion model proposal (`MCMC-diffusion`) that we developed and described in Chapter 5). Second, we consider an importance sampling method, based on an NS sampling scheme.

List of MCMC algorithms

For all MCMC algorithms, we must choose a point at which to start. Ideally, we would start from a draw from the target, though if we could draw from the target, we wouldn't be running MCMC in the first place. A common technique for choosing a starting point is burn-in or warmup. Here, one makes a preliminary choice of starting point and runs the chain for a number of 'burn-in' iterations. The states of the chain in this period are discarded, and the position of the chain at the end of burn-in is used as the starting point. Other possibilities include draws from the prior and the mode of the target.

For a fair comparison of the MCMC techniques used in this paper, we use a common set of starting points for all MCMC chains and the same length of burn-in. The efficiency of an MCMC sampler depends on the auto-correlation time of its chain, which can be reduced by a well-chosen proposal distribution. Thus, the key to an efficient MCMC implementation lies in constructing proposals judiciously. We consider the following MCMC codes:

- `emcee` [332, 333]: affine-invariant proposals are constructed from an ensemble of chains [334]. The affine-invariance makes the algorithm self-tuning and amenable to problems with degeneracies and anisotropies.

- `ptemcee` [335]: builds upon `emcee` by parallel tempering among several ensembles of chains [336, 337, 338]. Tempering involves raising the likelihood $\mathcal{L}^\beta$ to the power of a "temperature" β, in analogy with the Boltzmann factor where $\log \mathcal{L}$ plays the part of the energy, with "hot" being $\beta \to 0$ and "cold" where $\beta \to 1$. An ensemble of cold chains targets the original posterior and ensembles of warm chains target tempered versions of the posterior. The tempered posteriors are more amenable to MCMC because modes are somewhat flattened. A special move swaps states between cold and warm chains. Consequently, the cold chains mix more quickly between modes. Whilst this can help in multimodal problems, abrupt changes in the tempered posterior as we change temperature (phase transitions) could cause problems.

- `zeus` [339]: an ensemble slice sampler [340, 341]. Proposals are based on slices of the posterior, potentially leading to big steps and thus small auto-correlation. Anisotropic and multimodal problems are tackled using the ensemble paradigm:

affine-invariant slices are constructed from an ensemble of walkers and moves between modes are enabled by slices from one mode to another.

- `MCMC-diffusion` [232]: a mixture of local MH moves and global moves from a diffusion model trained on previous states of the chain.

See Table D.13 for a list of the number of walkers used for each test problem. We checked that the number of walkers chosen for each algorithm corresponded to the best available results given the computational constraints for each problem.

List of NS algorithms

We compared several Nested Sampling (NS) algorithms implemented in `Bilby` [94], including `MultiNest` [342, 343], `PolyChord` [61], `Dynesty` [344], `UltraNest/MLFriends` [345, 346], and `Nessai` [347]. We also integrated `JaxNS` [348] and `Nautilus` [349] into `Bilby`. Here is a summary of their methodologies and characteristics:

- `MultiNest`: The original NS algorithm, using ellipsoidal rejection sampling to bound the likelihood contour. We used the Python interface provided by `PyMultiNest`.
- `PolyChord`: Utilizes 1D slice sampling and clustering analysis to handle separable posterior modes. We explored the default settings, focusing on its ability to step out in anisotropic problems and navigate multimodal distributions. We used the Python interface provided by `PyPolychord`.
- `Dynesty`: Offers a variety of rejection and step sampling methods, including dynamic NS. We used the default sampling method, which in our case was random walk sampling away from a current live point since our test problems were all between 10D and 20D. The only exception was the 5D ΛCDM physics example (see section 7.3.2), for which uniform sampling is the default method since this problem is less than 10D.
- `UltraNest/MLFriends`: Implements various region and step sampling methods, along with reactive NS, a generalization of dynamic NS. Both reactive and static nested samplers were explored. The MLFriends version of UltraNest was used in this work.
- `Nessai`: Combines NS with artificial intelligence (AI) by using normalizing flows for sampling from the constrained prior.
- `JaxNS`: A probabilistic programming framework that implements NS in JAX for high performance through compilation. It employs 1D slice sampling without a step out procedure, using exponential contraction from the bounds of the prior domain.

- `Nautilus`: A pure-Python package for posterior and evidence estimation, utilizing importance sampling based on NS and neural networks.

Each algorithm's unique approach and specific features are designed to optimize performance and accuracy in NS, catering to the diverse needs of astrophysical research and beyond. Broadly they divide into two categories [350]: region(/rejection) samplers and path(/step) samplers. Region samplers aim to construct an explicit proxy for the likelihood-constrained prior distribution, and sample from that. They will be highly performant in low dimensions, but suffer from a curse of dimensionality in high dimensions, becoming exponentially inefficient above some problem- and sampler-dependent dimensionality $d_0 \sim \mathcal{O}(10)$. `MultiNest`, `Nessai`, `Nautilus` and `MLFriends` fall into this category. Path samplers run a Markov chain starting from a random live point until it has "forgotten" where it started. These are less efficient in low dimensions, but their inefficiency scales linearly with dimensions, allowing them to operate in principle in hundreds of dimensions. The state-of-the-art is achieved with slice sampling [351], as implemented in `PolyChord` and `JaxNS` (although `JaxNS` simultaneously uses an ellipsoidal element), with `Dynesty` and `UltraNest` switching to slice-sampling modes in high dimensions.

For each NS implementation we explored over a small grid of hyperparameters. In selecting the explored hyperparameters for each implementation we tried to choose similarly to how a typical end-user would in their analysis. This 'end-user mindset' reflects the overall intent of the paper to help the community chose most wisely for their analysis. To this end, we chose to explore the most well-documented hyperparameters specific to each implementation, and used each implementation's default stopping conditions.

Across most implementations there was a concept similar to the number of live points, in which case we explored settings of $n_{\text{live}} = \{25D, 50D, 100D\}$. We note that in the modern era most NS implementations do not have the concept of a static number of live points, but rather treat it as a dynamic quantity (particularly for the purposes of parallelisation). In these cases, it was often possible to set the notion of a minimal number of live points. By far the most impactful hyperparameter for the quality and computation expense of each run is the number of live points.

7.2 Methodology

In this section, we provide a brief outline of how we chose to compare samplers.

7.2.1 Evaluating Nested Sampling implementations

In this study, we aimed to evaluate different NS implementations to find the most computationally efficient solution for various problems, adopting the mindset of the

'typical experimenter'. The goal was to identify an NS implementation that achieves convergence with minimal computational effort, facilitating exploration within limited time frames. Therefore, our approach was to explore each NS implementation within consistent hyperparameter spaces constrained to a fixed computational limit per run, and to identify the most economical converging run.

Since each NS implementation has differing algorithmic structure, ensuring consistent hyperparameter spaces was reduced to a simple scheme of controlling the most well-documented hyperparameters. Specifically, for implementations with the concept of live points, the exploration space was $\{25D, 50D, 100D\}$, except for the Rastrigin problem which also included $\{200D, 500D, 1000D\}$. For implementations using Markov Chains, $5D$ accepted proposals per sample were selected. Where multiple modes of operation existed, e.g. the Static or Reactive mode of `ultranest`, the top recommended mode(s) were used. Static nested sampling was chosen over dynamic to maintain consistency in the number of live points. All other hyperparameters were left at their default, including stopping conditions. Each run was limited to one hour, except for the Rastrigin problem, which had a two-hour limit, reflecting practical constraints faced by non-expert experimenters.

Convergence was assessed via the Wasserstein distance measure, also known as the Earth-Mover's distance from a reference posterior. First, a reference posterior we generated for each problem using a high-resolution run of NS, making sure that this reference posterior closely matched the true posterior as closely as possible for each problem. Specifically, we used `JaxNS` with $2000D$ live points, $10D$ slices per acceptance. Then, an upper bound on the multi-variate Wasserstein-1 distance was used as the distance measure, namely $W_1(P, Q) \leq \sum_i W_1(P_i, Q_i)$ where W_1 is the 1-norm Wasserstein distance, and P_i, Q_i are the respective one-dimensional marginals of the distributions. We then used per-problem convergence thresholds to identify convergence. We opted to compute the Wasserstein distance in the unit-cube scaled parameter spaces, hence the commonly used fiducial threshold of 0.1 was used for all problems except Eggbox and Rastrigin, which assumed 0.3 and 0.2 respectively, due to the more challenging nature of those problems. Plots of the Wasserstein upper bound for the cheapest converged NS runs are shown in Figure D.1. Finally, for each problem and NS implementation, the cheapest run within the convergence threshold was identified as the cheapest converging run. During the course of our experimentation, we extended the hyperparameter space and computational limits for the Rastrigin problem, so that at least one NS implementation converged for each case.

7.2.2 Diagnostics & Metrics

We are using two categories of algorithm to sample from a posterior distribution: MCMC and NS. In each case, we want to assess convergence, diagnose any problems with runs and measure performance. Although under appropriate conditions MCMC

algorithms may be proven to converge asymptotically, at some point we must stop our chain(s). We would like to know whether the samples obtained at that point are representative of the underlying stationary distribution. They will not be representative in a short chain if the chain explores the posterior distribution slowly. This 'slow mixing' may occur, for example, in multi-modal problems.

Unfortunately, assessing the convergence of an MCMC chain is extremely challenging [352, 327] and arguably we at best make checks of pseudo-convergence, since no checks can tell us what would happen if we were to run the chain for longer [353]. We use `arviz` [354] to estimate the Gelman-Rubin $\hat{R}$ parameter [355, 356] for each model parameter. This requires more than one MCMC chain, ideally initialised at different starting points, as it compares the within-chain and between-chain variances of a parameter. See [357, 358] for up-to-date discussions about the limitations, interpretation and developments of the $\hat{R}$ parameter. Recently, $\hat{R} \leq 1.01$ was recommended as a convergence check [357].

Ensemble MCMC samplers are somewhat problematic with regard to diagnostics and metrics. In ensemble samplers, efficient proposals are constructed using an ensemble of walkers. As the proposals are constructed from the ensemble, the walkers are not independent chains. Consequently, the conventional interpretations of and rules of thumb about the $\hat{R}$ parameter (and to some degree the ESS estimates discussed later) are open to question.

NS does not have the same question of convergence to a stationary distribution, since it is not building a Markov chain. This may, however, be seen as a drawback as we cannot efficiently generate more samples by running NS for longer (though see the discussion on dynamic NS). In NS, rather, the key questions are whether we prematurely terminated NS and whether we genuinely drew independent samples from the constrained prior because if we did not, the NS results are faulty. In [359], a new test was introduced to diagnose problems in this aspect of NS runs. The diagnostic, p_{insert}, makes use of the fact that the rank indices of new live points should be uniformly distributed. The result of this diagnostic is a p_{insert} from a one-sample test between the rank indices from an NS run and a uniform distribution. This test has not yet been implemented across all the algorithms used in this paper, so there are some missing p_{insert} in certain test problems. As always, care should be taken in interpreting p_{insert}, though small p_{insert} could indicate a failure to sample from the constrained prior and thus a faulty NS run.[1]

To quantify the performance of an algorithm, we consider what we obtained and for what cost. As we are concerned with sampling from the posterior, we measure what we obtained by the effective number of independent samples drawn from the posterior, n_{eff}.

[1] If we fail to faithfully sample from the constrained prior whilst compressing to the posterior bulk, it might impact the evidence but not the posterior. If, on the other hand, it happens whilst compressing through the bulk of the posterior, posterior inferences may be faulty as well.

Unfortunately, the samples we obtain are not independent draws from the posterior. In Markov chains the samples are correlated as the state of the chain depends on the previous state. In NS, the samples obtained are weighted, as they are weighted draws from the NS procedure rather than draws directly from the posterior. The cost that we are concerned with is the number of likelihood evaluations, $n_{\mathcal{L}}$, such that as a metric we consider

$$\text{Efficiency} = \frac{n_{\text{eff}}}{n_{\mathcal{L}}}. \tag{7.1}$$

This doesn't take into account the fact that different algorithms may be parallelized more efficiently than others. The effective sample size may depend on the parameter or quantity of interest; in such cases we take the minimum among model parameters. See appendix D.1 for a discussion of effective sample size estimates in MCMC and NS. Broadly speaking, n_{eff} estimates may not reliable when $n_{\text{eff}} \lesssim 100$.

7.3 Results

We investigated the algorithms in section 7.1 by running them on several test functions and on examples inspired by global fits in particle physics and cosmology. The test functions were chosen to exhibit pathological features, including high-dimensionality, multi-modality, degeneracy, and phase transitions.

7.3.1 Eggbox

The Eggbox function, given by

$$f(\boldsymbol{\theta}) = \left(2 + \prod_{i=1}^{D}\left[\cos\left(\frac{\theta_i}{2}\right)\right]\right)^5, \tag{7.2}$$

leads to a complicated multimodal posterior distribution, as illustrated in a 10-dimensional corner plot shown in fig. 7.1. The posterior mean obtained through different algorithms is summarised in table D.1, while table 7.1 presents their corresponding performance metrics and diagnostics.[2] The term "Modes Found" refers to the percentage of modes for which an algorithm generates sufficient samples to outline a 1σ contour within the 2D marginal distributions in fig. 7.1. We note that this is a considerably easier problem then enumerating the $5^{10} \approx 10^7$ modes present, and is only marginally interesting for a distribution with a high degree of axis-aligned symmetry such as the Eggbox.

[2] We note that for a highly multi-modal problem, the posterior mean proves to be a statistic which is sensitive to misfitting, but otherwise is not particularly useful in comparison with for example the individual mode means.

Algorithm	Diagnostic	n_{eff}	$n_{\mathcal{L}}$	Efficiency	Modes Found
MCMC: `emcee`	$\hat{R} = 1.07$	31	1.50M	2.06×10^{-5}	69%
MCMC: `ptemcee`	$\hat{R} = 1.00$	110k	1.50M	0.0733	91%
MCMC: `zeus`	$\hat{R} = 1.10$	25	1.50M	8.80×10^{-6}	53%
MCMC: `MCMC-diffusion`	$\hat{R} = 1.00$	120k	1.08M	0.111	93%
NS: `dynesty`	$p_{\text{insert}} = 0.23$	2.04k	26.8M	7.59×10^{-5}	96%

Table 7.1: Algorithm performance metrics for the $10D$ Eggbox test function.

Scrutinising these results shows that `MCMC-diffusion` and `ptemcee` demonstrate superior performance in locating most of the modes efficiently compared to NS methods. Specifically, `dynesty` excels in identifying the highest number of modes among the algorithms, although it requires significantly more likelihood evaluations than MCMC approaches to achieve this outcome. `MCMC-diffusion`'s efficacy in navigating multimodal distributions, characterised by its ability to transition between modes efficiently, is evident in this scenario. While `ptemcee` performs well, it struggles with modes situated at the parameter space's edges, leading to narrower posterior standard deviations.

Notably, other traditional MCMC algorithms fail to converge within the specified computational constraints. Regarding `MCMC-diffusion`, multiple chains are executed to identify various modes efficiently, as outlined in the original paper [232]. This allows the algorithm to overcome the traditional limitations of MCMC (e.g., fixed step size, local optima trapping). It should be noted that in a system of high symmetry (namely equal mode weighting and separation) multi-start MCMC produces the correct result. However, if modes are asymmetric, multi-start MCMC will initially populate the modes in proportion to the size of their burn-in basin of attraction, which may not be the same as their posterior weight. Only after running for long enough to transition between modes will the correct posterior be sampled, which may be after the computational budget has been exhausted, and often long after a convergence criterion has been triggered. We shall see this for the spike-slab function discussed in section 7.3.1.

Rosenbrock

The Rosenbrock function [360], defined as

$$f(\boldsymbol{\theta}) = \sum_{i=1}^{D-1} [100 \times (\theta_{i+1} - \theta_i)^2 + (1 - \theta_i)^2], \tag{7.3}$$

displays a banana-shaped degeneracy in two dimensions. A 10-dimensional corner plot illustrating this function is depicted in fig. C.1, accompanied by a posterior mean

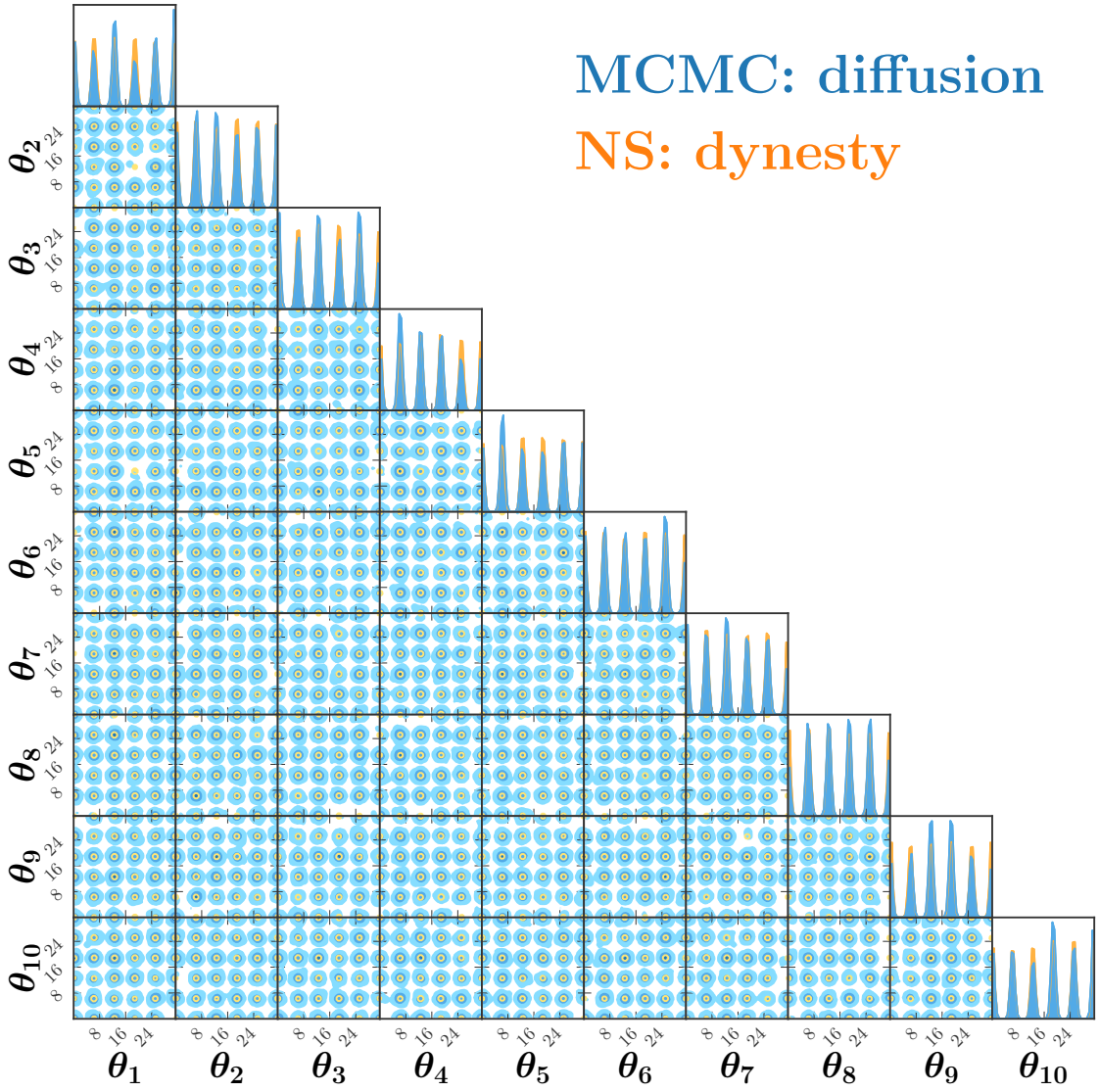


Figure 7.1: Corner plot of the $10D$ Eggbox function in eq. (7.2). The plot shows the posteriors from the best performing MCMC and NS implementations.

detailed in table D.2 and summary metrics provided in table 7.2.

Analyzing the outcomes, the following key observations arise. Among the NS algorithms, `dynesty` appears to perform more accurately compared to others, as indicated by the posterior mean. The other NS algorithms exhibit a slight underestimation in parameters $\theta_4 - \theta_{10}$, with narrower standard deviations. Further examination of the posterior plots reveals that the remaining NS algorithms fail to explore the entire distribution width effectively. Notably, the second peak in parameters $\theta_6 - \theta_8$ is notably

shifted to the left, suggesting that these algorithms are unable to locate the broader peak positioned further to the right.

Algorithms such as `emcee` and `zeus` demonstrate insufficient convergence within the specified number of likelihood evaluations, as evidenced by the $\hat{R}$ metric for these methods. With `dynesty` requiring significantly more likelihood evaluations than `ptemcee` or `MCMC-diffusion`, the results imply that Markov chain Monte Carlo methods are once again demonstrating greater efficiency compared to NS in this context.

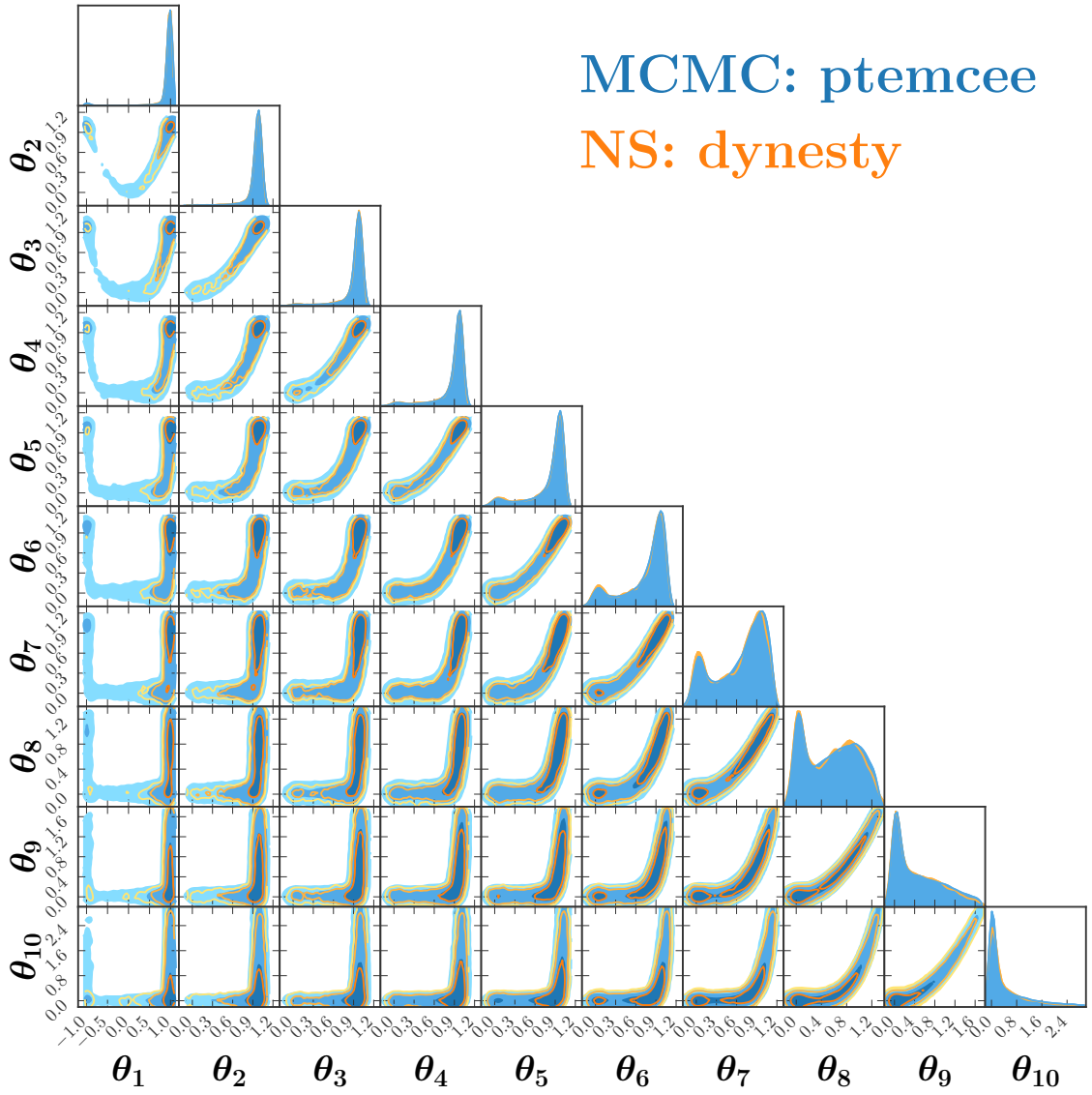


Figure 7.2: Corner plot of the $10D$ Rosenbrock function.

Algorithm	Diagnostic	n_{eff}	$n_{\mathcal{L}}$	Efficiency
MCMC: `emcee`	$\hat{R} = 1.35$	7.10k	4.00M	0.00180
MCMC: `ptemcee`	$\hat{R} = 1.00$	4.55k	550k	0.00828
MCMC: `zeus`	$\hat{R} = 1.09$	26.5k	2.00M	0.0131
MCMC: `MCMC-diffusion`	$\hat{R} = 1.01$	2.12k	700k	0.00303
NS: `Dynesty`	$p_{\text{insert}} = 0.38$	2.6k	5.1M	5.1×10^{-4}
NS: `Multinest`	$p_{\text{insert}} = 0.1$	2.6k	167k	0.02
NS: `Polychord`	$p_{\text{insert}} = 0.69$	3k	2.9M	0.001
NS: `Nessai`	$p_{\text{insert}} = 5.13\times10^{-5}$	3.9k	203k	0.02
NS: `JaxNS`		1.6k	3.9M	4.07×10^{-4}
NS: `Nautilus`		10k	62k	0.16

Table 7.2: Summary of algorithm performance for $10D$ Rosenbrock test function.

Rastrigin

The Rastrigin function [361] is given by

$$f(\boldsymbol{\theta}) = 10D + \sum_{i=1}^{D} \left[\theta_i^2 - 10\cos(2\pi\theta_i)\right], \tag{7.4}$$

where D is the number of dimensions of θ.

The Rastrigin function is also multimodal. A 10-dimensional corner plot depicting this function is presented in fig. 7.3, complemented by the computed posterior mean documented in table D.3 and summarised metrics detailed in table 7.3.

Key observations from the analysis of these results include the following. The `ptemcee` algorithm demonstrates a notable proximity to the true distribution compared to other methods, requiring a significantly lower number of likelihood evaluations (400,000) compared to `Dynesty` (11M) to achieve this result. This efficiency indicates a compelling performance advantage for `ptemcee` in capturing the underlying distribution accurately within a more feasible computational budget. The contrast in performance metrics highlights the varying computational demands and effectiveness of different algorithms when applied to this multimodal function.

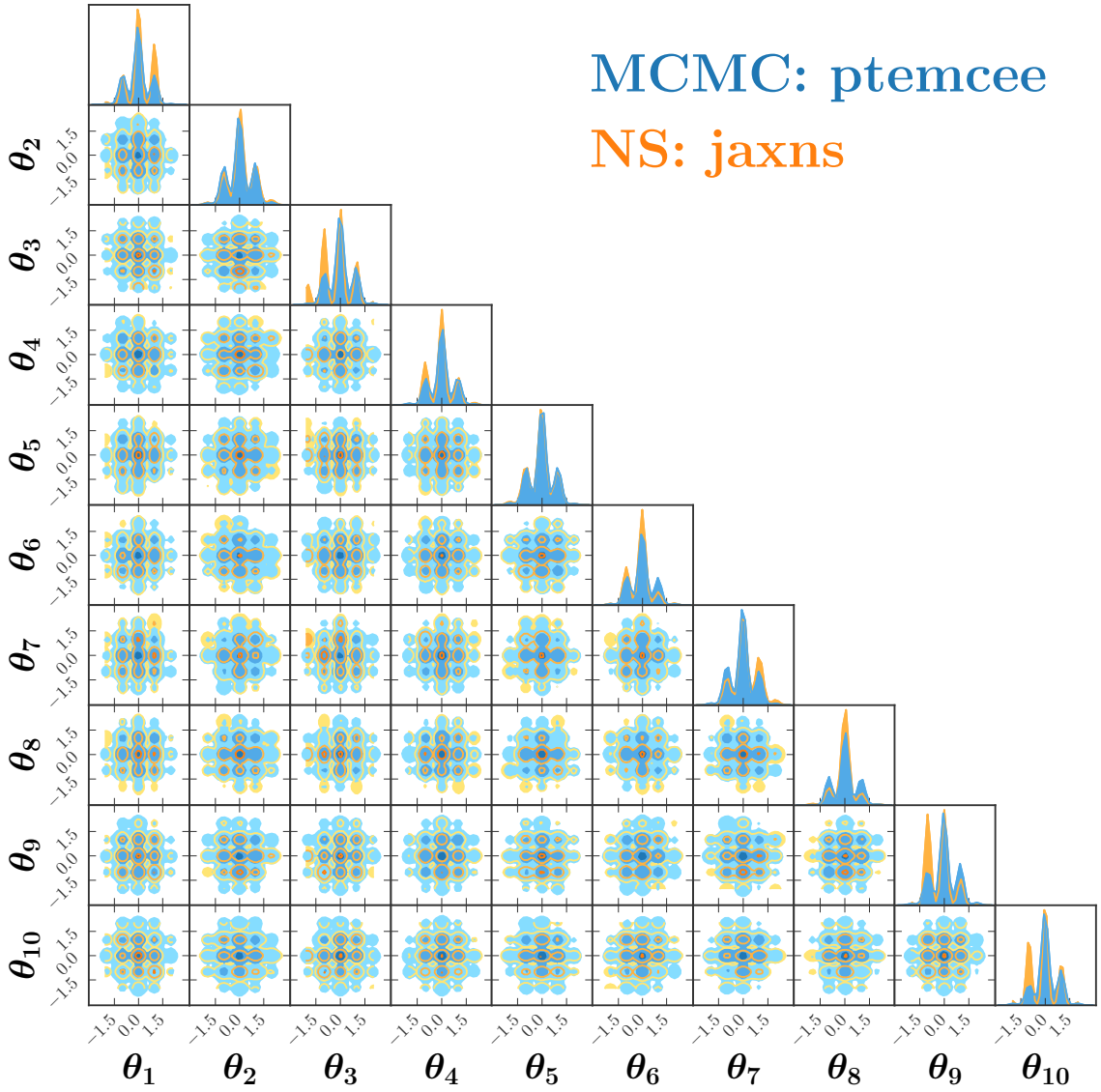


Figure 7.3: Corner plot of the $10D$ Rastrigin function.

Spike-slab function

The spike-slab function is commonly used in statistical modeling, particularly in the context of variable selection, as a Bayesian prior distribution. It combines a point-like likelihood mass (spike) with a continuous distribution (slab). This distribution is typically used to encourage sparsity in model parameters, allowing for automatic variable selection during model estimation.

The spike-slab function may have a spike that is similar in volume to the slab which

Algorithm	Diagnostic	n_{eff}	$n_{\mathcal{L}}$	Efficiency
MCMC: `emcee`	$\hat{R} = 1.12$	25	400k	6.28×10^{-5}
MCMC: `ptemcee`	$\hat{R} = 1.00$	14.1k	400k	0.0353
MCMC: `zeus`	$\hat{R} = 1.12$	27	400k	2.67×10^{-5}
MCMC: `MCMC-diffusion`	$\hat{R} = 1.01$	11.2k	400k	0.0275
NS: `JaxNS`		54k	104M	5.2×10^{-4}

Table 7.3: Summary of algorithm performance for $10D$ Rastrigin test function.

can pose challenges for Markov chain Monte Carlo algorithms. These challenges include poor mixing between modes, leading to inefficient exploration, and convergence issues due to sharp transitions and narrow bridges.

Our spike-slab function is the sum of a narrow $2D$ Gaussian and a broad multivariate Gaussian. The narrow Gaussian has a mean of 6 for each parameter and a diagonal covariance matrix with a standard deviation of 0.08 for each parameter. Similarly, the broader Gaussian has a mean of 2.5 and a standard deviation of 0.8. A corner plot of this function in $10D$ is shown in fig. 7.4, along with the posterior mean and error in table D.4 and summary metrics in table 7.4.

Most MCMC algorithms, except for `MCMC-diffusion`, struggle to correctly balance the weights assigned to the spike and slab components. Despite indicating convergence based on the $\hat{R}$ metric, these algorithms converge to incorrect estimates, underscoring the complexity and potential confusion posed by the spike-slab function for MCMC methods. Even `ptemcee` demonstrates a tendency to incorrectly estimate the relative weight of the spike and slab components, as can be seen by the underestimation of the posterior mean in parameters θ_1 and θ_2 (see table D.4), indicating `ptemcee` has focused too heavily on the slab. This further highlights the difficulty in achieving accurate balance when dealing with the spike-slab function within Bayesian inference settings.

Nested sampling can also show similar difficulties. Whilst region samplers with a well-trained proxy will capture mode weighting up to some statistical error, chain based samplers on their own will mis-estimate modes that they cannot ergodically mix [61]. Samplers with semi-independent mode tracking such as `PolyChord` and `GGNS` [362] mitigate this (up to a statistical error which they track). This is backed up by both the posterior mean and Wasserstein distances, since `PolyChord` was the only NS algorithm that got close to the true distribution in this test problem.

These observations emphasise the intricate nature of algorithmic performance, and are important to bear in mind when considering the challenges associated with complex

model priors.

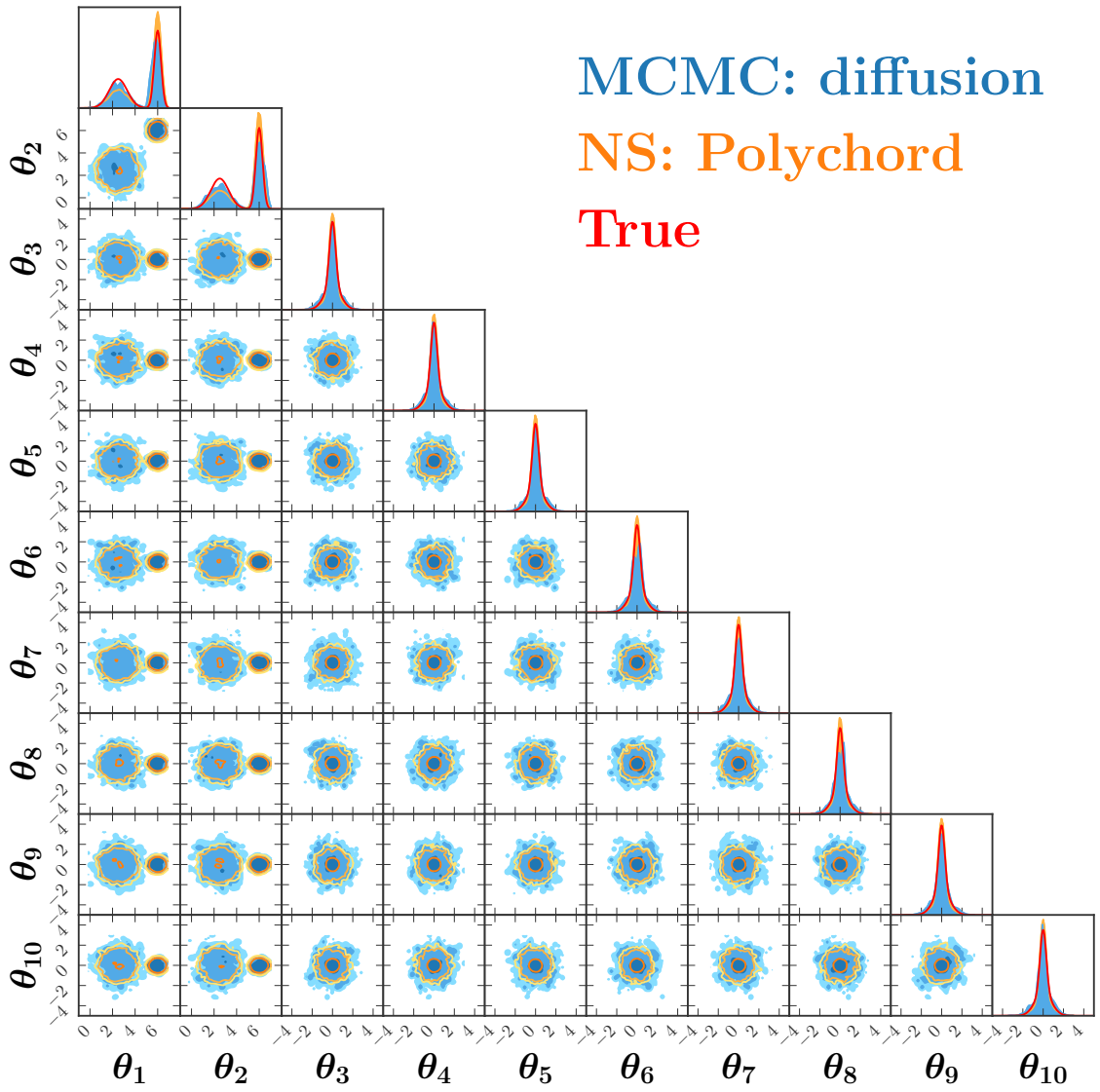


Figure 7.4: Corner plot of the $10D$ spike-slab function.

7.3.2 Cosmology — Planck

The Lambda Cold Dark Matter (ΛCDM) model stands as the prevailing cosmological paradigm, aiming to describe the universe's composition, structure, and evolution. At its core, the ΛCDM model is anchored in the Big Bang Theory and inflation, further

Algorithm	Diagnostic	n_{eff}	$n_{\mathcal{L}}$	Efficiency
MCMC: `emcee`	$\hat{R} = 1.01$	1.44k	250k	0.00600
MCMC: `ptemcee`	$\hat{R} = 1.00$	22k	250k	0.0880
MCMC: `zeus`	$\hat{R} = 1.00$	8.79k	250k	0.0293
MCMC: `MCMC-diffusion`	$\hat{R} = 1.00$	9.66k	90k	0.107
NS: `Polychord`	$p_{\text{insert}} = 0.32$	15k	7.6M	0.002
NS: `JaxNS`		5.6k	4M	0.001

Table 7.4: Summary of algorithm performance for $10D$ spike-slab test function

augmented by the cosmological constant (Λ) and cold dark matter (CDM) components to address observations not fully explained by traditional cosmological frameworks.

The model's elegance lies in its foundation upon a mere six parameters, each playing a pivotal role in shaping our understanding of our Universe. Namely, the angular size of the sound horizon at the surface of last scattering (θ_s), the baryon density ($\Omega_b h^2$), the dark matter density ($\Omega_c h^2$), the amplitude of primordial fluctuations (A_s), the scalar spectral index (n_s) and the optical depth to reionization (τ).

Bayesian global fits of the ΛCDM model are the state of the art for extracting knowledge of these fundamental cosmological parameters. For any given parameter point, the Boltzmann code CLASS [363] can be used to compute various power spectra of the fluctuations in the cosmic microwave background, which can then be compared to current observations. The spectra are expressed in terms of the multipole moment ℓ, related to angular scales on the sky, with the notation C_ℓ used for the angular power spectrum. This spectrum reflects the variance of the temperature fluctuations as a function of the angular scale. Depending on the nature of the fluctuations, we can categorize them into different types, primarily TT, TE, and EE, each offering unique insights into the early universe. The TT (Temperature-Temperature) spectrum represents the correlation of temperature fluctuations, providing a direct measurement of the cosmic microwave background's anisotropies. The TE (Temperature-E-mode Polarization) spectrum captures the cross-correlation between temperature fluctuations and the E-mode polarization pattern, offering a complementary view that helps break degeneracies in cosmological parameter estimation. Lastly, the EE (E-mode Polarization-E-mode Polarization) spectrum focuses on the correlation within the E-mode polarization itself, sensitive to the early universe's density fluctuations and the epoch of reionization. We denote them as C_ℓ^{TT}, C_ℓ^{TE} and C_ℓ^{EE}.

With those we can compute the likelihood function (see Eq. 3.7 in [364])

$$\ln p(\hat{C}(\theta)|C) = -\frac{1}{2}\sum_{\ell}(2\ell+1)\left[\frac{D}{|C|} + \ln\frac{C}{\hat{C}} - 2\right], \tag{7.5}$$

where $D = C_\ell^{\mathrm{TT}}\hat{C}_\ell^{\mathrm{EE}} + \hat{C}_\ell^{\mathrm{TT}}C_\ell^{\mathrm{EE}} - 2C_\ell^{\mathrm{TE}}\hat{C}_\ell^{\mathrm{TE}}$, and we shift the likelihood such that

$$\ln p\left(C = \hat{C}(\theta_0)|C\right) = 0, \tag{7.6}$$

where we choose for our fiducial cosmology[3]

$$\theta_0 = \left(\Omega_b h^2, \Omega_{\mathrm{cdm}} h^2, 100\theta_s, \ln\left(10^{10}A_s\right), n_s, \tau\right) = (0.0224, 0.12, 1.0411, 3.0753, 0.965, 0.054). \tag{7.7}$$

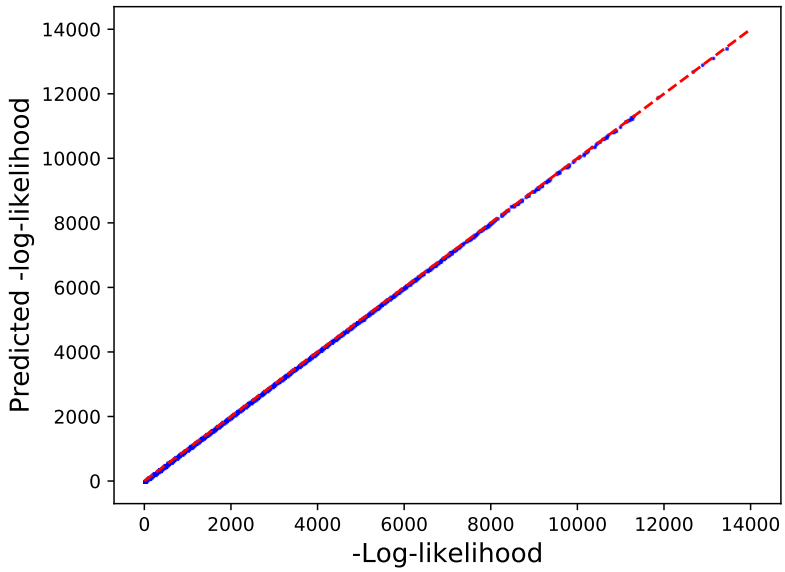


Figure 7.5: Validation plot of the network, showing the true log likelihood on the x-axis and the predicted value on the y-axis.

The computational expense of evaluating the likelihood function represents a major challenge in inferring the parameters of cosmological models such as the six-dimensional ΛCDM model. To overcome this challenge, we have used a deep neural network to create a direct and efficient mapping from the cosmological parameter space to the log-likelihood defined in eq. (7.5). By training the network with many simulations, we

[3]For details see [364].

bypass the need for computationally intensive direct simulations during the global fit itself. This strategy significantly speeds up the parameter estimation process, as has been demonstrated by previous work such as `CosmoPower` [365]

To this end, we have generated a large dataset of 1.2 million data points, each representing a unique configuration within the ΛCDM parameter space, based on a flat prior distribution as shown in table 7.5. We used 80% of this dataset for training the neural network and split the remaining 20% evenly between validation and testing. Both the input and the output data were normalised to a Gaussian distribution. The architecture of our neural network is explicitly tailored to map ΛCDM parameters to their log-likelihood counterparts.

Parameter	Prior Range
$100\theta_s$	[1.03, 1.05]
$\Omega_b h^2$	[0.020, 0.024]
$\Omega_c h^2$	[0.10, 0.14]
n_s	[0.9, 1.0]
$\ln(10^{10} A_s)$	[2.98, 3.16]
τ	[0.010, 0.097]

Table 7.5: Prior ranges for the Lambda CDM model parameters.

The neural network model was developed with `TensorFlow` and our model consists of a sequential arrangement of six hidden layers, each with 100 neurons. These layers use the Scaled Exponential Linear Unit (SELU) [366] activation function, which was chosen for its self-normalising properties that significantly improve training efficiency. The initial weights of these layers were set using the LeCun Normal initialiser to optimise the initial conditions for training the network. The output layer consists of a single neuron with a linear activation function for our regression task. Finally, the Mean Absolute Error was selected as the loss function.

A multi-layered training optimisation strategy was used to refine the performance of the model. This included the use of the Adam optimiser [367] with an initial learning rate of 0.001. To further improve the accuracy of the model, the learning rate was halved with each complete run of the dataset. To minimise the risk of overfitting, training was also terminated prematurely if no improvement in validation loss was observed over a period of ten epochs.

The performance of our neural network on the test data set is visualised in fig. 7.5, demonstrating the ability of the network to accurately predict the log-likelihood values.

A corner plot is shown in fig. 7.6, along with the posterior mean and error in table D.5 and summary metrics in table 7.6.

All algorithms converge effectively to the correct answer except for `nautilus`, as evidenced by the excellent agreement in the posterior mean, which is expected for a

Algorithm	Diagnostic	n_{eff}	$n_{\mathcal{L}}$	Efficiency
MCMC: `emcee`	$\hat{R} = 1.01$	265	62.0k	0.00875
MCMC: `ptemcee`	$\hat{R} = 1.00$	2.68k	62.0k	0.0432
MCMC: `zeus`	$\hat{R} = 1.00$	2.14k	62.0k	0.0345
MCMC: `MCMC-diffusion`	$\hat{R} = 1.00$	2.09k	62.0k	0.0337
NS: `Dynesty`	$p_{\text{insert}} = 0.72$	2k	359k	5.6×10^{-3}
NS: `Multinest`	$p_{\text{insert}} = 6.04\times10^{-7}$	861	8.2k	0.1
NS: `Polychord`	$p_{\text{insert}} = 0.96$	2k	872k	2.3×10^{-3}
NS: `Nessai`	$p_{\text{insert}} = 0.18$	1.37k	25.6k	0.054
NS: `JaxNS`		604	633k	9.5×10^{-4}
NS: `Nautilus`		10k	24.6k	0.41
NS: `Ultranest`		958	63k	0.015

Table 7.6: Summary of algorithm performance for CMB likelihood

well-behaved unimodal function. Whilst neural network approximations can reliably reduce the computational load [368] of historically expensive portions of the calculation, as we move further into the era of precision cosmology with more and more systematics to be modelled it will prove crucial to study which algorithms are the most efficient in terms of computational cost and performance.

Ignoring Nautilus due to the inaccurate posterior, the most efficient algorithm was found to be `Multinest`, which reached a satisfactory estimate of the posterior with only 8200 likelihood evaluations in total. The best MCMC algorithm was found to be `ptemcee`, although `zeus` and `MCMC-diffusion` showed very similar performance. This matches the intuition that, for an unimodal posterior, the benefit of algorithms that allow for better mixing of modes is reduced. However, the evidence of this test shows that they still allow for a more efficient exploration of the posterior relative to more basic techniques. It is worth noting that the overall efficiency is relatively low for all the algorithms.

Most importantly, this work shows that a CMB ΛCDM cosmological example is not a stringent test of a sampler's efficacy in comparison with the other examples considered in this paper, as the posterior is well-constrained in all parameters and approximately gaussian. This is not true for beyond ΛCDM models, or for non CMB data as is the case on both counts for the recent DESI analysis [369].

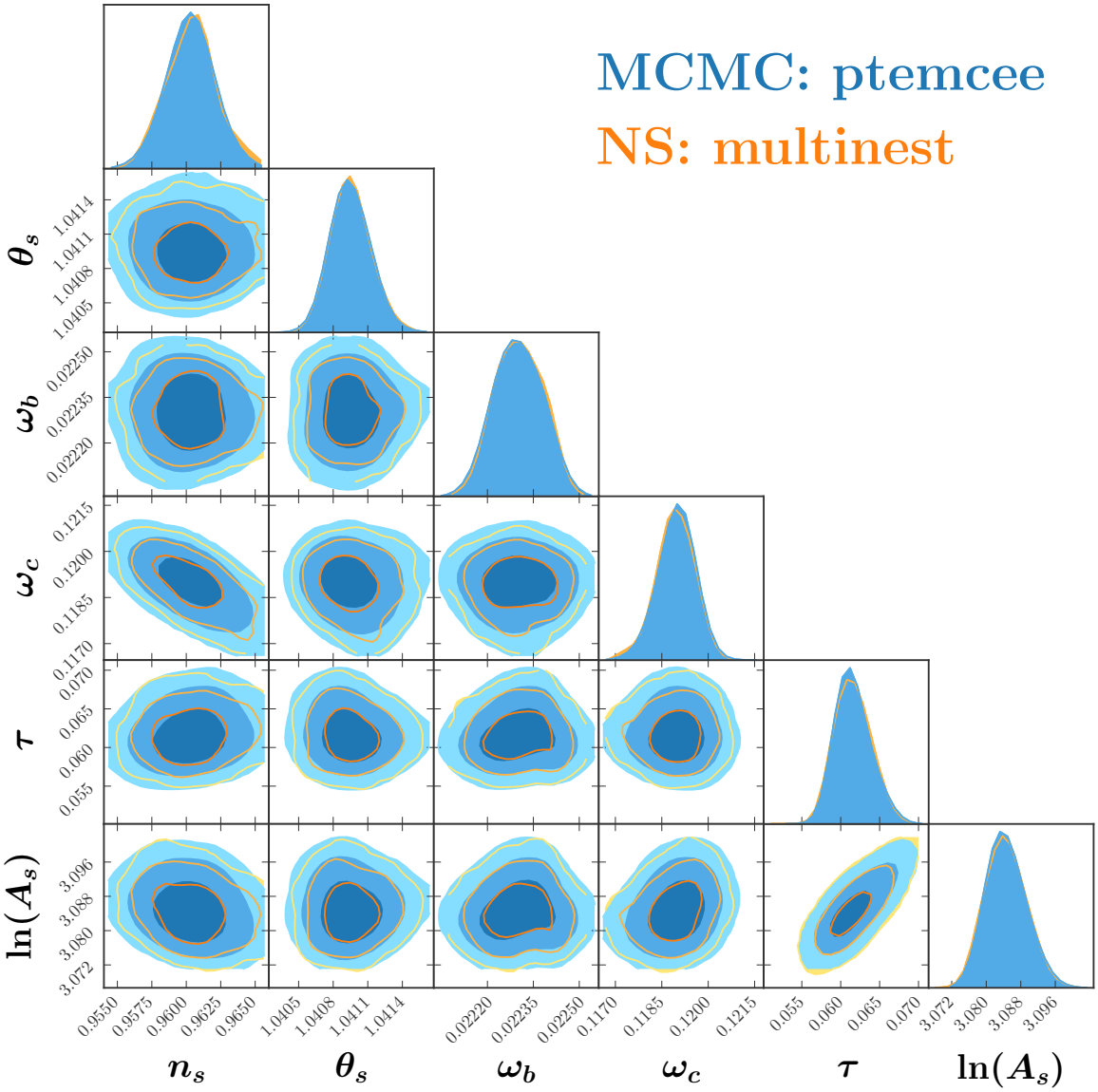


Figure 7.6: Corner plot for the CMB Likelihood.

7.3.3 Particle physics — LHC

An additional example of the use of global fit techniques in particle astrophysics is the attempt to explore models of beyond-Standard Model particle physics. To compare our Bayesian sampling techniques on a realistic example, we use the results of a previous global fit of the Minimal Supersymmetric Standard Model, performed by the GAMBIT collaboration [370]. The 7 parameters of interest were the soft mass terms M_2, $m^2_{\tilde{f}}$, $m^2_{H_u}$, $m^2_{H_d}$ (defined at the scale $Q = 1$ TeV), the trilinear couplings

A_{u_3}, A_{d_3} and $\tan\beta$ (defined at the scale m_Z). The fit also varied a number of nuisance parameters, including the top quark mass, the strong coupling constant, the nuclear matrix elements for WIMP-nucleon scattering and the local dark matter density, leading to 12 parameters in total. The fit included an extensive range of experimental constraints, including results from particle colliders, flavour physics, measurements of the cosmic microwave background and null observations from direct and indirect dark matter search experiments. Evaluating the likelihood of any particular parameter point was computationally expensive, and the total fit required extensive CPU hours.

To ensure that we can compare our Bayesian sampling techniques with relatively minor computational effort, we interpolated the likelihood obtained in [370] using a deep neural network, as originally proposed in [371, 372]. The network was used in a previous DarkMachines challenge, and details of the network architecture and validation can be found in Ref. [211].

A corner plot is shown in fig. 7.7, along with the posterior mean and error in table D.6 and summary metrics in table 7.7.

The results of this test show some distinct patterns of behaviour among the algorithms:

- The posterior mean is broadly consistent between the `ptemcee`, `MCMC-diffusion`, `Dynesty`, `JaxNS`, `Nautilus`, `Ultranest` and `PolyChord` algorithms, with the `emcee` and `zeus` values differing due to apparent lack of convergence. The likelihood surface for the MSSM7 is very complicated, with several disparate regions of high likelihood separated in the large volume of the parameter space, and this leads to a more complex posterior than that explored in any of the previous examples.

- Even among the algorithms that broadly agree with each other, the `JaxNS` algorithm gives discrepant mean values for the parameters m_f^2, $m_{H_u}^2$, $m_{H_d}^2$, A_u, and $\tan(\beta)$, and has a large Wasserstein distance in parameters m_f^2 and A_u. It does not look wise to rely on the low-resolution results of `JaxNS` in this case.

- In terms of efficiency, `Nautilus` is the most efficient algorithm, able to map the posterior satisfactorily with only 11,000 likelihood evaluations. For a likelihood that is as expensive as the original GAMBIT MSSM7 likelihood (which could take tens of minutes to hours per evaluation due to the need to simulate LHC proton-proton collisions), this efficiency is a crucial feature. The most efficient MCMC algorithm is `ptemcee`, though again we see that the `MCMC-diffusion` algorithm has a similar efficiency.

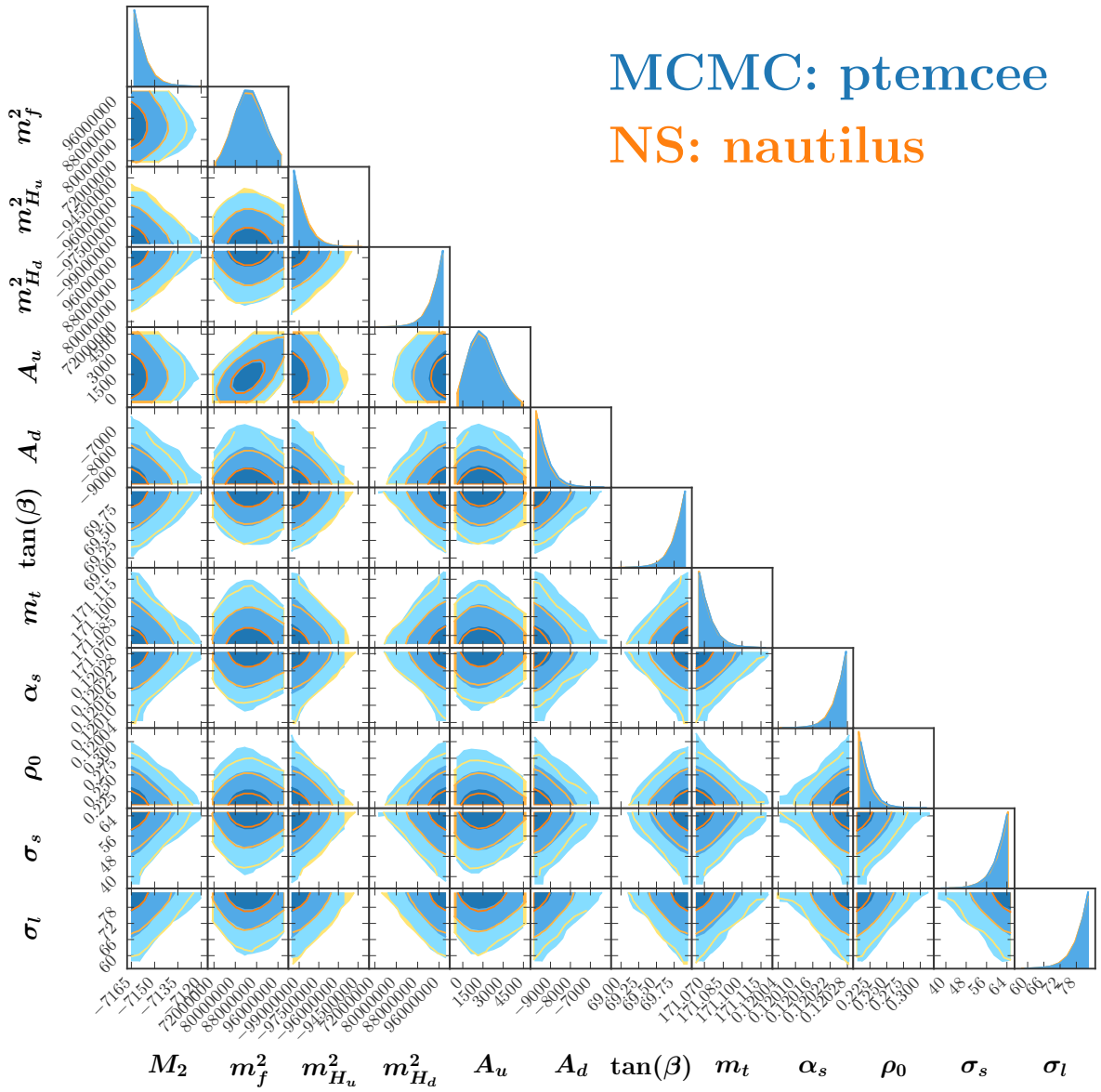


Figure 7.7: Corner plot for the MSSM7 Likelihood.

7.4 Conclusions

We have performed a detailed comparison of a variety of Bayesian sampling algorithms in an attempt to determine to what extent each of these algorithms can accurately and efficiently describe complicated high-dimensional test functions and real-world physics likelihoods. We have approached this question by studying algorithmic performance out-of-the-box with minimal hyperparameter tuning and low computation time, so as

Algorithm	Diagnostic	n_{eff}	$n_{\mathcal{L}}$	Efficiency
MCMC: `emcee`	$\hat{R} = 1.76$	43	111k	4.18×10^{-4}
MCMC: `ptemcee`	$\hat{R} = 1.00$	6.1k	123k	0.0495
MCMC: `zeus`	$\hat{R} = 1.21$	141	111k	1.17×10^{-3}
MCMC: `MCMC-diffusion`	$\hat{R} = 1.00$	4.3k	111k	0.0387
`Dynesty`	$p_{\text{insert}} = 0.15$	3.5k	969k	3.7×10^{-3}
`JaxNS`		1.9k	3.9M	4.9×10^{-4}
`Nautilus`		10k	11k	0.09
`Ultranest`		2.8k	428k	6.4×10^{-3}

Table 7.7: Summary of algorithm performance for MSSM7 likelihood

to simulate the typical performance one might expect to quickly obtain as a physicist unfamiliar with the inner workings of each algorithm. On this basis we can reach a number of conclusions:

- Modern MCMC tools such as `MCMC-diffusion` and `ptemcee` can tackle multimodal problems, bringing MCMC efficiency to a realm previously only solvable with Nested Sampling. These modern MCMC algorithms tend to be more efficient than their NS counterparts across all the test functions studied here.
- The `emcee` and `zeus` algorithms are severely outclassed by the other algorithms present in this paper in terms of efficiency, and are rarely able to obtain reasonable results within the computational limits we have imposed.
- No NS algorithm has clearly emerged as the "best" in our results, rather certain algorithms appear more well suited to certain types of functions. For example, `Polychord` appears to be the most effective for a problem with differently weighted modes like the spike-slab, while `dynesty` appears to be the most effective for a heavily multimodal problem like the Eggbox.
- Subtlety exists even within NS when considering region-based sampling versus chain-based sampling with/without semi-independent mode tracking.
- The CMB+ΛCDM physics likelihood was accurately and efficiently sampled by all the algorithms tested here, so cosmologists dealing with similar unimodal problems need not be picky about which specific algorithm they choose among the options we have explored in this paper.

For both the analytic functions and the two physics likelihoods explored in this chapter, we have observed that many of the algorithms used here show promising performance. This certainly motivates their use in further real-world BSM physics applications.

Summary and Outlook

This thesis has demonstrated the power of global QCD analysis when coupled with advancements in Bayesian methodology. These techniques have not only enabled us to accurately characterise various quantities such as PDFs within the SM, but also have allowed us to push forward and use our precise knowledge of the fundamental building blocks of the Universe to seacrh for physics beyond the SM. We anticipate that the existing approaches to global analysis along with the new methods we have developed here will continue to have important implications in a wide variety of fields.

Having outlined in Chapter 1 how exactly PDFs fit into the overall picture of strong force interactions in non-perturbative QCD, we then went on to describe the various methods that are currently used to charaterize those PDFs in Chapter 2, specifically in the framework of global QCD analysis. The PDF uncertainties generated by these diferent approaches were compared to each other in Chapter 3, where all methods were found to produce the same results for a toy example of a PDF fit, unless they employed neural networks. We identified the partitioning and cross-validation steps inherent to the neural network approach as the primary drivers of this difference, which tended to produce inflated uncertainty bands compared to the other methodologies. A more extensive study using a realistic global QCD analysis in future may further nail down the extent to which such a neural network approach may inadvertantly warp PDF uncertainties.

In Chapter 4 we used the JAM global analysis framework to conduct the most comprehensive fit of PDFs by this group to date. Starting from the previous best JAM results, we added 1370 high-x polarized DIS data points including new JLab Hall C data, along with lattice QCD data that had never been included in JAM before. We found that after the inclusion of this new data we were able to constrain the sign of the Δg PDF, ruling out negative solutions after the quality of the fit worsened significantly for those replicas. The spin flavour asymmetry $\Delta\bar{u} - \Delta\bar{d}$ was seen to increase in magnitude, while the Δd PDF appeared smaller when compared to previous results and other fitting groups. We also observed little evidence for substantial higher twist effects, with d_2 consistent with zero. Introducing more lattice data would help to further constrain the polarized PDFs, as well as SIDIS at large p_T which will be produced at the Electron-Ion Collider.

Returning to the methodology of PDF global analyses, we devised a new MCMC algorithm in Chapter 5 that uses diffusion models to accelerate sampling of the Bayesian posterior, which we called `MCMC-Diffusion`. Applying this new technique to several complicated test functions, we found clear improvement in the performance of `MCMC-Diffusion` compared to the canonical Metropolis-Hastings MCMC algorithm. This same improvement could also be seen when applied to a toy example of a global QCD fit, suggesting that a future full analysis employing `MCMC-Diffusion` within a framework such as JAM could produce an increase in accuracy in characterising the PDFs as well as improved efficiency.

We then turned to BSM physics in Chapter 6, attempting for the first time to see how global QCD analysis could help search for dark matter candidates such as the dark photon or the Z' boson. Introducing modifications to the underlying theory of the JAM analysis to account for the potential existence of such BSM particles, we were able to place new constraints in large regions of parameter space. In the case of the dark photon model, we saw a dramatic improvement in the quality of the fit over the SM hypothesis, providing an inital hint of a dark photon signal. This could be explored further in future by dedicating other experimental searches to the preferred values for the dark photon parameters. There are also several other BSM models and variants that could be constrained by following the same process we have performed here.

With our knowledge of Bayesian sampling, we could then explore other BSM models such as the Λ-Cold Dark Matter model and the Minimal Supersymmetric Standard model in Chapter 7. Within the two categories of MCMC and nested sampling, we compared several algorithms across a variety of test functions along with these BSM likelihoods to determine which would show the best performance under what conditions. We found that sophisticated MCMC algorithms such as `MCMC-Diffusion` and `ptemcee` tended to utperform all others in terms of efficiency and accuracy, suggesting that they would be best suited for future applications to BSM searches.

Appendix A

Target mass corrections in collinear factorization

Following the derivation by Moffat *et al.* [290] of the unpolarized F_1 and F_2 structure functions at finite values of M^2/Q^2, in this appendix we outline the steps in the corresponding derivation of the spin-dependent structure functions g_1 and g_2 in the presence of hadron masses. As usual, we define the Nachtmann scaling variable as

$$x_{\rm N} = \frac{2x}{1+\rho}, \qquad \rho^2 = 1 + \frac{4x^2M^2}{Q^2}\,. \tag{A.1}$$

Neglecting lepton masses, the hadronic tensor for polarized DIS in the one-photon exchange approximation is given by

$$W_{\mu\nu} \;=\; i\epsilon_{\mu\nu\alpha\beta}\frac{q^\alpha}{P\cdot q}\left[S^\beta\, g_1(x,Q^2) + \left(S^\beta - \frac{S\cdot q}{P\cdot q}P^\beta\right) g_2(x,Q^2)\right], \tag{A.2}$$

where the structure functions g_1 and g_2 contain all M^2/Q^2 target mass corrections.

In the Breit frame the four-momenta of the target nucleon $P = (P^+, P^-, \boldsymbol{P}_\perp)$ and virtual photon $q = (q^+, q^-, \boldsymbol{q}_\perp)$ can be written in light-front coordinates as

$$P \;=\; \left(\frac{Q}{\sqrt{2}x_{\rm N}}, \frac{x_{\rm N}M^2}{\sqrt{2}Q}, \mathbf{0}_\perp\right), \qquad q = \left(-\frac{Q}{\sqrt{2}}, \frac{Q}{\sqrt{2}}, \mathbf{0}_\perp\right), \tag{A.3}$$

respectively. The covariant spin vector S is given by

$$S = M\left(\lambda\frac{P^+}{M}, -\lambda\frac{M}{2P^+}, \boldsymbol{S}_\perp\right), \tag{A.4}$$

and satisfies the relations $P\cdot S = 0$ and $S^2 = -M^2$. In the massless target limit, $M \to 0$, the "$-$" components of P and S vanish, and we can define "massless" four-vectors $\widetilde{P}$

and $\widetilde{S}$ by

$$\widetilde{P} = \left(\frac{Q}{\sqrt{2}x_{\rm N}}, 0, \mathbf{0}_\perp\right), \qquad \widetilde{S} = M\left(\lambda\frac{P^+}{M}, 0, \boldsymbol{S}_\perp\right), \tag{A.5}$$

keeping the transverse component of the spin vector.

With this notation one can write the hadronic tensor as

$$W_{\mu\nu} = i\epsilon_{\mu\nu\alpha\beta}\frac{q^\alpha}{\widetilde{P}\cdot q}\left[\widetilde{S}^\beta\,\widetilde{g}_1(x,Q^2) + \left(\widetilde{S}^\beta - \frac{\widetilde{S}\cdot q}{\widetilde{P}\cdot q}\widetilde{P}^\beta\right)\widetilde{g}_2(x,Q^2)\right], \tag{A.6}$$

where the structure functions $\widetilde{g}_{1,2}$ do not contain the M^2/Q^2 corrections, in contrast to those in Eq. **??**). The hadronic tensors in Eq. **??**) and Eq. **??**) must be equal, as they are simply the same object expressed in terms of two choices of basis tensors.

Using the relations

$$\widetilde{P}^\mu = \frac{1}{Q^2 + x_{\rm N}^2M^2}\left(Q^2P^\mu - x_{\rm N}M^2q^\mu\right), \tag{A.7a}$$

$$\widetilde{S}^\mu = S^\mu + \lambda\frac{x_{\rm N}M^2}{Q^2 + x_{\rm N}^2M^2}\left(x_{\rm N}P^\mu + q^\mu\right), \tag{A.7b}$$

one can then write

$$\begin{aligned} i\epsilon_{\mu\nu\alpha\beta}\frac{q^\alpha}{\widetilde{P}\cdot q}\widetilde{S}^\beta = i\epsilon_{\mu\nu\alpha\beta}\frac{q^\alpha}{P\cdot q}\Bigg[&\frac{(Q^2 - x_{\rm N}^2M^2)(Q^2 + 3x_{\rm N}^2M^2)}{(Q^2 + x_{\rm N}^2M^2)^2}S^\beta \\ &- \frac{x_{\rm N}^2M^2}{Q^2}\frac{(Q^2 - x_{\rm N}^2M^2)^2}{(Q^2 + x_{\rm N}^2M^2)^2}\left(S^\beta - \frac{S\cdot q}{P\cdot q}P^\beta\right)\Bigg], \end{aligned} \tag{A.8a}$$

and

$$\begin{aligned} i\epsilon_{\mu\nu\alpha\beta}\frac{q^\alpha}{\widetilde{P}\cdot q}\left(\widetilde{S}^\beta - \frac{\widetilde{S}\cdot q}{\widetilde{P}\cdot q}\widetilde{P}^\beta\right) = i\epsilon_{\mu\nu\alpha\beta}\frac{q^\alpha}{P\cdot q}\Bigg[&4x_{\rm N}^2M^2\frac{Q^2 - x_{\rm N}^2M^2}{(Q^2 + x_{\rm N}^2M^2)^2}S^\beta \\ &+ \frac{(Q^2 - x_{\rm N}^2M^2)^3}{Q^2(Q^2 + x_{\rm N}^2M^2)^2}\left(S^\beta - \frac{S\cdot q}{P\cdot q}P^\beta\right)\Bigg] \end{aligned} \tag{A.8b}$$

Substituting Eqs. (A.10a) and (A.10b) into Eq. **??**) and comparing the tensor structures to Eq. **??**) leads to the relations

$$g_1(x,Q^2) = \frac{(Q^2 - x_{\rm N}^2M^2)(Q^2 + 3x_{\rm N}^2M^2)}{(Q^2 + x_{\rm N}^2M^2)^2}\,\widetilde{g}_1(x,Q^2) + 4x_{\rm N}^2M^2\frac{(Q^2 - x_{\rm N}^2M^2)}{(Q^2 + x_{\rm N}^2M^2)^2}\,\widetilde{g}_2(x,Q^2), \tag{A.9a}$$

$$g_2(x,Q^2) = -\frac{x_{\rm N}^2}{Q^2}\frac{(Q^2 - x_{\rm N}^2M^2)^2}{(Q^2 + x_{\rm N}^2M^2)^2}\,\widetilde{g}_1(x,Q^2) + \frac{(Q^2 - x_{\rm N}^2M^2)^3}{Q^2(Q^2 + x_{\rm N}^2M^2)^2}\,\widetilde{g}_2(x,Q^2), \tag{A.9b}$$

which can be rewritten in terms of ρ as

$$g_1(x,Q^2) = \frac{2\rho-1}{\rho^2}\,\widetilde{g}_1(x,Q^2) + \frac{2(\rho-1)}{\rho^2}\,\widetilde{g}_2(x,Q^2), \tag{A.10a}$$

$$g_2(x,Q^2) = \frac{1-\rho}{(1+\rho)\rho^2}\,\widetilde{g}_1(x,Q^2) + \frac{2}{(1+\rho)\rho^2}\,\widetilde{g}_2(x,Q^2). \tag{A.10b}$$

Equations (A.10a) and (A.10b) then represent the general, model-independent relation between the massless limit structure functions and those evaluated at finite values of M^2/Q^2, without any kinematical or dynamical approximations.

Within the standard collinear factorization framework [288], in the leading twist approximation the structure functions can be written in factorized form in terms of PDFs and hard coefficient functions (4.8), but with the lower limit of integration in the convolution integral (4.9) equal to the Nachtmann scaling variable $x_{\rm N}$. Namely, without making any massless target approximations, in collinear factorization one has

$$\widetilde{g}_1(x,Q^2) \to g_1^{\rm LT}(x_{\rm N},Q^2), \tag{A.11a}$$

$$\widetilde{g}_2(x,Q^2) \to g_2^{\rm LT}(x_{\rm N},Q^2). \tag{A.11b}$$

Using the Wandzura-Wilczek relation is to write $g_2^{\rm LT}$ in terms of $g_1^{\rm LT}$ [130] then leads to

$$g_1^{\rm TMC}(x,Q^2) = \frac{2\rho-1}{\rho^2}\,g_1^{\rm LT}(x_{\rm N},Q^2) + \frac{2(\rho-1)}{\rho^2}\Big(-g_1^{\rm LT}(x_{\rm N},Q^2) + \int_{x_{\rm N}}^1 \frac{{\rm d}z}{z} g_1^{\rm LT}(z,Q^2)\Big), \tag{A.12a}$$

$$g_2^{\rm TMC}(x,Q^2) = \frac{1-\rho}{(1+\rho)\rho^2}\,g_1^{\rm LT}(x_{\rm N},Q^2) + \frac{2}{(1+\rho)\rho^2}\Big(-g_1^{\rm LT}(x_{\rm N},Q^2) + \int_{x_{\rm N}}^1 \frac{{\rm d}z}{z} g_1^{\rm LT}(z,Q^2)\Big). \tag{A.12b}$$

After simplifying, we obtain

$$g_1^{\rm TMC}(x,Q^2) = \frac{1}{\rho^2} g_1^{\rm LT}(x_{\rm N},Q^2) + \frac{2(\rho-1)}{\rho^2}\int_{x_{\rm N}}^1 \frac{{\rm d}z}{z} g_1^{\rm LT}(z,Q^2), \tag{A.13a}$$

$$g_2^{\rm TMC}(x,Q^2) = -\frac{1}{\rho^2} g_1^{\rm LT}(x_{\rm N},Q^2) + \frac{2}{(1+\rho)\rho^2}\int_{x_{\rm N}}^1 \frac{{\rm d}z}{z} g_1^{\rm LT}(z,Q^2), \tag{A.13b}$$

as in Eqs. (4.14).

Appendix B

Data to theory comparisons

In this appendix we present the full set of comparisons of the polarized data versus theory, and the detailed reduced χ^2 values for the spin-dependent observables used in this analysis. Tables B.1 and B.2 give the $\chi^2_{\rm red}$ values for longitudinally polarized and transversely polarized, respectively, inclusive DIS asymmetries A_1 and $A_\parallel$ for proton, deuteron, and ^{3}He targets. The number of data points ($N_{\rm dat}$) corresponds to a cut of $W^2 > 4$ GeV^2 on the DIS data. The $\chi^2_{\rm red}$ values for the longitudinally polarized SIDIS asymmetries A_1^h, for the production of hadrons h ($= \pi^\pm, K^\pm, h^\pm$) from p, d, and ^{3}He targets, are listed in Table B.3. Note that for the SLAC 155x data the measured transverse asymmetry $\widetilde{A}_\perp$ has a small contribution from longitudinal polarization since the target polarizations were not completely perpendicular to the beam line [195]. Finally, Table B.4 lists the $\chi^2_{\rm red}$ values for inclusive jet and weak boson production in longitudinally polarized pp collisions. Here the jet data are binned in the jet rapidity $y_{\rm jet}$ and transverse momentum p_T, while the W/Z production data are binned in p_T and the electron rapidity, η_e.

Comparisons between the data from the individual experiments and the fitted values are shown in Figs. B.1–B.9. The most basic polarization observable, the inclusive DIS asymmetry A_1, is shown in Fig. B.1 for the proton data from EMC [175], SMC [176, 177], and COMPASS [179, 180], along with the new lower-Q^2 Jefferson Lab data from the SANE experiment [373], which probed deeper into the high-x region. Good agreement is obtained for all the datasets, although the theory predictions for the SANE data tend to be slightly lower than the observed points, particularly for the higher-x data. Data on the parallel asymmetry $A_\parallel$ for the proton are shown in Fig. B.2 from the HERMES [188], SLAC [181, 184, 186], and Jefferson Lab [191, 192] experiments. The large number of points for the Jefferson Lab Hall B eg1b experiment illustrates why the χ^2 for this dataset is so high, particularly at lower Q^2 where many of these points have small error bars. Even though the theory prediction falls right in the middle of the data, the sheer number of points with low errors situated away from the mean of the data means that the quality of the fit is lower than it appears for this dataset.

The transverse A_2 DIS asymmetry for the proton is shown in Fig. B.3 from the HERMES [196] experiment and the more recent SANE data from Hall C at Jefferson Lab [373]. Unlike for the longitudinal A_1 asymmetry in Fig. B.1, our fit goes through the SANE data and shows no indication of underestimating the data with any consistency. The comparison with the transverse proton DIS asymmetries $A_\perp$ and $\widetilde{A}_\perp$ from SLAC [184, 194, 195] are shown in Fig. B.4. While the vast majority of points are fitted well, one datum in the SLAC E155 $A_\perp$ dataset at the very highest x and Q^2 values in the is anomalously higher than our fit, resulting in a noticeably larger uncertainty band.

The polarized DIS data for the deuteron A_1 asymmetry from COMPASS [178] and SMC [176, 177] are shown in Fig. B.5. The COMPASS data are fitted better than the SMC data for this observable, with some SMC points at the same x value are slightly separated from the theory prediction. For the polarized deuteron longitudinal asymmetry $A_\parallel$, Fig. B.6 shows the comparison with data from HERMES [188], SLAC [184, 185], and CLAS at Jefferson Lab [191, 193]. Once again the high number of points with very small error bars contributes to a high χ^2 for the Jefferson Lab eg1b experiment, particularly for the 5.7 GeV incident energy subset. The data for the transversely polarized deuteron DIS asymmetries $A_\perp$ and $\widetilde{A}_\perp$ are shown in Fig. B.7 from the SLAC E143 [184] and E155 [194, 195] experiments. The enlarged theoretical uncertainty bands at large x and Q^2 are visible for the E155 data as in the $A_\perp$ case, however, for the deuteron data the position of the measured values above and below zero is likely driving the fitted result.

For polarized ^{3}He targets, Fig. B.8 shows DIS data for the A_1 and $A_\parallel$ asymmetries from HERMES [187], SLAC [182, 183], and Jefferson Lab Hall A [189, 190]. For the HERMES data the ^{3}He asymmetry was converted, assuming a model ^{3}He wave function, to an asymmetry on the free neutron [187]. Despite the small number of data points from the ^{3}He experiments, the global fit gives a very good description of the data. Good description are also found for the ^{3}He nuclear A_2 and $A_\perp$ asymmetries in Fig. B.9 for SLAC [182, 183] and Jefferson Lab [189, 190] data.

Turning now to the polarized SIDIS asymmetries A_1^h, in Fig. B.10 we show the data versus theory comparison for the production of charged pions, $h = \pi^\pm$, for p and d data from HERMES [198] and COMPASS [199, 200]. The agreement with data is relatively good across all panels, with the exception of the deuteron HERMES data, where for both π^+ and π^- production the highest-x data points appear to be below both the theoretical curves and the general trend of the other data points at lower x. The polarized SIDIS kaon production asymmetries, $A_1^{K^\pm}$, are shown in Fig. B.11 for HERMES [198] and COMPASS [199, 200] data on proton and deuteron targets. The vast majority of these points are fitted fairly well, with the exception of the HERMES deuteron K^- asymmetry at the highest x values, which display greater fluctuations within large experimental uncertainties. For the production of unidentified

hadrons $h^{\pm}$, Fig. B.12 shows polarized SIDIS asymmetries $A_1^{h^{\pm}}$ from HERMES [198], COMPASS [199], and SMC [374]. Again the agreement between the data and the fitted values is quite good, with both reflecting the increasing numerical values of the asymmetries in the high-x region.

For polarized pp scattering data, Fig. B.13 shows the longitudinally polarized jet asymmetries $A_{LL}^{\rm jet}$ at $\sqrt{s} = 200$ and 510 GeV from the STAR Collaboration [201, 202, 203, 121, 122, 123, 124] at RHIC. These data are also fitted quite well, with only a single datum in the mid-range of the 2005 dataset falling outside of the theoretical uncertainty bands. The $\chi^2_{\rm red}$ values are generally below unity, with only the STAR data from Refs. [122, 202] giving a $\chi^2_{\rm red}$ exceeding 1.

The single spin asymmetries A_L^W and $A_L^{W/Z}$ for vector boson production in longitudinally polarized pp scattering at $\sqrt{s} = 510$ GeV is shown in Fig. B.14. The data from the STAR [204] and PHENIX [205, 206] Collaborations, integrated over specific p_T ranges, are well described by our fitted results, with the exception of the W^+ PHENIX data with $p_T > 30$ GeV cut [206], although the uncertainties there are relatively large.

Table B.1: Reduced χ^2 values for longitudinal DIS polarization asymmetries A_1 and $A_\parallel$ for p, d, and ^{3}He (or "neutron") targets. The number of data points ($N_{\rm dat}$) corresponds to a cut of $W^2 > 4$ GeV^2 on the DIS data.

Experiment	Target	Observable	$N_{\rm dat}$	$\chi^2_{\rm red}$
EMC [175]	p	A_1	10	0.37
SMC [176]	p	A_1	11	0.39
SMC [176]	d	A_1	11	1.59
SMC [177]	p	A_1	7	1.61
SMC [177]	d	A_1	7	0.63
COMPASS [179]	p	A_1	11	1.02
COMPASS [178]	d	A_1	11	0.45
COMPASS [180]	p	A_1	35	0.95
SLAC E80/E130 [181]	p	$A_\parallel$	19	0.61
SLAC E142 [182]	^{3}He	A_1	6	1.01
SLAC E143 [184]	p	$A_\parallel$	56	0.84
SLAC E143 [184]	d	$A_\parallel$	56	1.04
SLAC E154 [183]	^{3}He	$A_\parallel$	16	0.50
SLAC E155 [186]	p	$A_\parallel$	65	1.24
SLAC E155 [185]	d	$A_\parallel$	65	1.06
HERMES [187]	"n"	A_1	7	0.27
HERMES [188]	p	$A_\parallel$	25	0.58
HERMES [188]	d	$A_\parallel$	25	0.95
JLab E99-117 [189]	^{3}He	$A_\parallel$	3	0.18
JLab E06-014 [190]	^{3}He	$A_\parallel$	14	1.47
JLab eg1-dvcs [191]	p	$A_\parallel$	117	1.65
JLab eg1-dvcs [191]	d	$A_\parallel$	73	0.52
JLab eg1b [192]	p	$A_\parallel$	473	1.01
JLab eg1b [193]	d	$A_\parallel$	112	0.85
JLab SANE [373]	p	A_1	30	1.46
Total			1265	**1.01**

Table B.2: Reduced χ^2 values for transverse DIS polarization asymmetries A_2, $A_\perp$, and $\widetilde{A}_\perp$ for p, d, and ^{3}He targets. The number of data points ($N_{\rm dat}$) corresponds to a cut of $W^2 > 4$ GeV^2 on the DIS data.

Experiment	Target	Observable	$N_{\rm dat}$	$\chi^2_{\rm red}$
SLAC E142 [182]	^{3}He	A_2	6	0.68
SLAC E143 [184]	p	$A_\perp$	43	1.02
SLAC E143 [184]	d	$A_\perp$	43	0.92
SLAC E154 [183]	^{3}He	$A_\perp$	16	1.10
SLAC E155 [194]	p	$A_\perp$	58	0.93
SLAC E155 [194]	d	$A_\perp$	58	1.53
SLAC E155x [195]	p	$\widetilde{A}_\perp$	92	1.29
SLAC E155x [195]	d	$\widetilde{A}_\perp$	92	0.79
HERMES [196]	p	A_2	15	1.02
JLab E99-117 [189]	^{3}He	$A_\perp$	3	1.88
JLab E06-014 [197]	^{3}He	$A_\perp$	14	0.88
JLab SANE [373]	p	A_2	30	0.82
Total			470	**1.06**

Table B.3: Reduced χ^2 values for longitudinal SIDIS asymmetries A_1^h for production of hadrons h ($= \pi^{\pm}, K^{\pm}, h^{\pm}$) from p, d, and ^{3}He targets.

Experiment	Target	Hadron	$N_{\rm dat}$	$\chi^2_{\rm red}$
HERMES [198]	p	π^+	7	2.35
HERMES [198]	p	π^-	7	1.83
HERMES [198]	d	π^+	7	1.02
HERMES [198]	d	π^-	7	2.95
COMPASS [200]	p	π^+	10	4.74
COMPASS [200]	p	π^-	10	2.88
COMPASS [199]	d	π^+	8	0.34
COMPASS [199]	d	π^-	8	0.85
HERMES [198]	d	K^+	7	0.91
HERMES [198]	d	K^-	7	0.82
HERMES [198]	d	$K^{\pm}$	7	0.59
COMPASS [200]	p	K^+	10	0.55
COMPASS [200]	p	K^-	10	0.99
COMPASS [199]	d	K^+	8	0.88
COMPASS [199]	d	K^-	8	0.81
SMC [374]	p	h^+	12	2.41
SMC [374]	p	h^-	12	0.85
SMC [374]	d	h^+	12	0.70
SMC [374]	d	h^-	12	1.08
HERMES [198]	p	h^+	7	2.72
HERMES [198]	p	h^-	7	1.17
HERMES [198]	d	h^+	7	1.13
HERMES [198]	d	h^-	7	0.73
HERMES [375]	^{3}He	h^+	7	0.55
HERMES [375]	^{3}He	h^-	7	0.69
COMPASS [199]	d	h^+	10	0.61
COMPASS [199]	d	h^-	10	0.98
Total			231	**1.36**

Table B.4: Reduced χ^2 values for jet and weak boson production in polarized pp collisions. Inclusive jets data are binned according to y_{jet} and p_T, while W/Z production are binned according to p_T and the electron rapidity, η_e.

Reaction	Experiment	Observable	N_{dat}	χ^2_{red}
Inclusive jets	STAR [201]	$A_{\parallel}^{\text{jet}}$	4	0.23
	STAR [202]	$A_{\parallel}^{\text{jet}}$	7	1.53
	STAR [203]	$A_{\parallel}^{\text{jet}}$	9	0.36
	STAR [121]	$A_{\parallel}^{\text{jet}}$	18	0.54
	STAR [122]	$A_{\parallel}^{\text{jet}}$	12	1.58
	STAR [123]	$A_{\parallel}^{\text{jet}}$	18	0.84
	STAR [376]	$A_{\parallel}^{\text{jet}}$	13	0.75
	PHENIX [125]	$A_{\parallel}^{\text{jet}}$	2	0.38
W/Z production	STAR [204]	A_L^W	12	0.80
	PHENIX [205]	$A_L^{W/Z}$	2	0.05
	PHENIX [206]	$A_L^{W/Z}$	4	0.66
Total			104	**0.78**

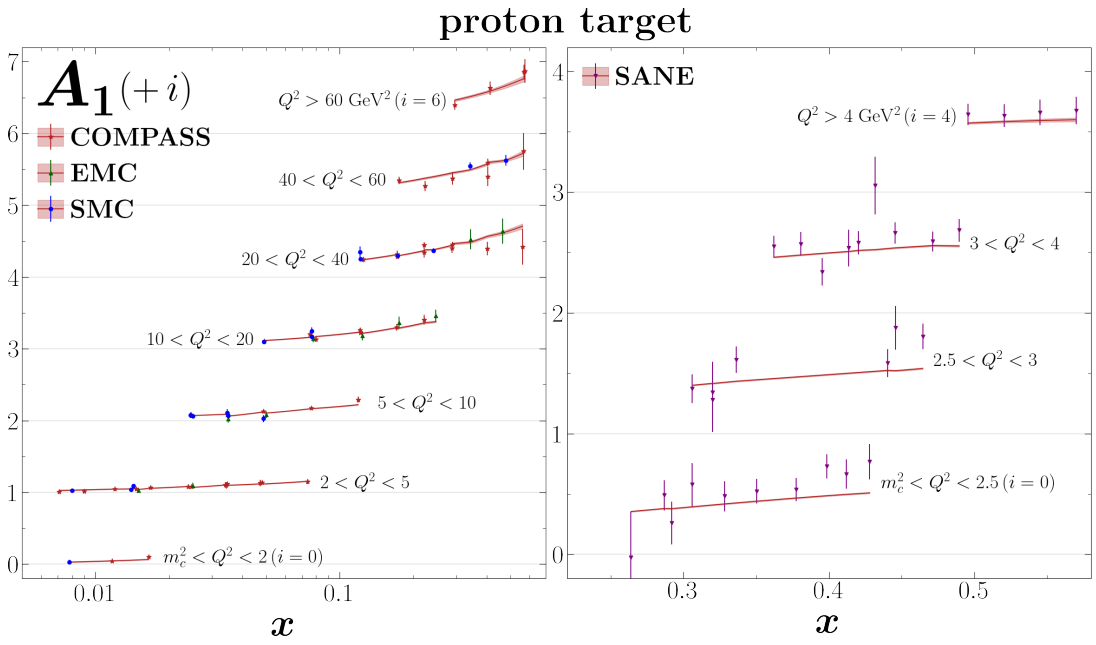


Figure B.1: Proton A_1 polarization asymmetry from COMPASS [179, 180], EMC [175], SMC [176, 177], and the SANE experiment at Jefferson Lab [373] versus Bjorken x, compared with the JAM fit (with 1σ uncertainty bands). Data in different Q^2 bins are offset by i for clarity.

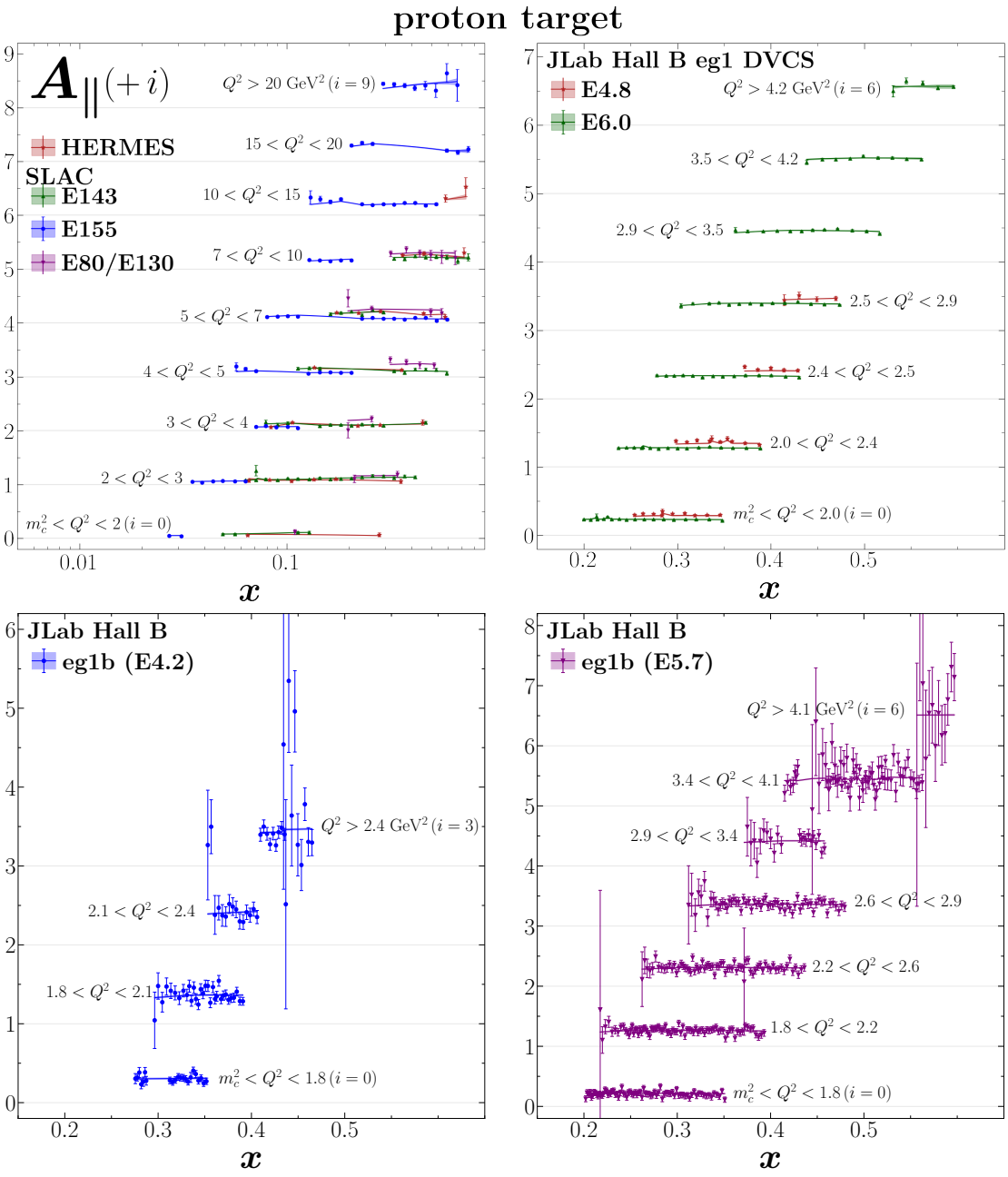


Figure B.2: Polarized DIS data for the observable $A_{\parallel}^p$ from HERMES [188], SLAC [181, 184, 186], and Jefferson Lab [191, 192] plotted as a function of x against the JAM result (colored 1σ bands). Data in different Q^2 bins are increased by i for clarity.

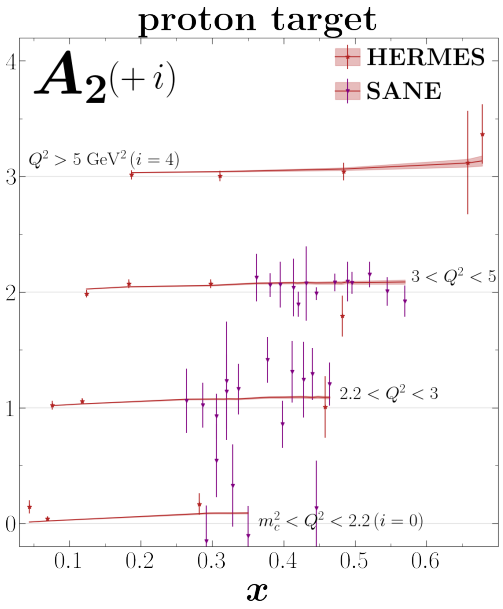


Figure B.3: Polarized DIS data for the proton A_2 asymmetry from HERMES [196] and SANE [373] plotted as a function of x against the JAM result (red 1σ bands). Data in different Q^2 bins are increased by i for clarity.

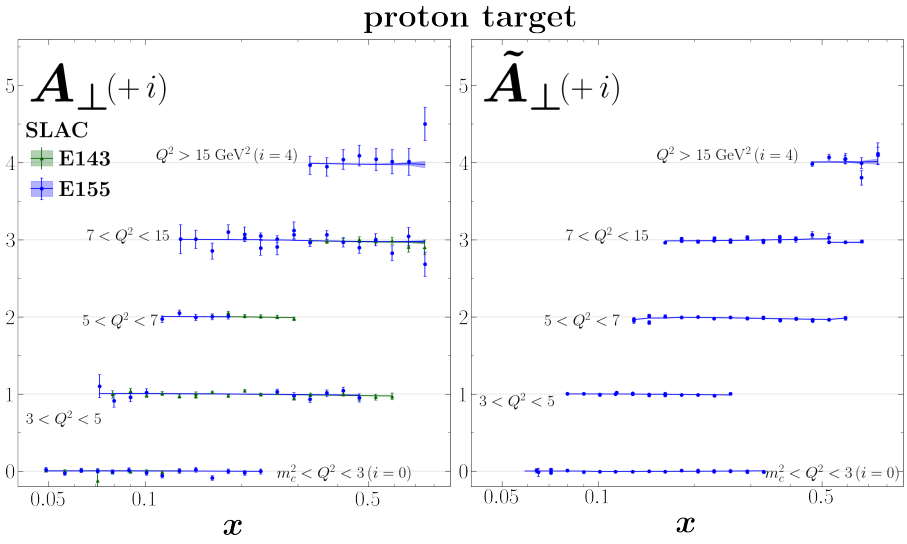


Figure B.4: Polarized DIS data for the observable $A_\perp$ (left panel) and $\widetilde{A}_\perp$ (right panel) from SLAC [184, 194, 195] plotted as a function of x against the JAM result (colored 1σ bands). Data in different Q^2 bins are increased by i for clarity.

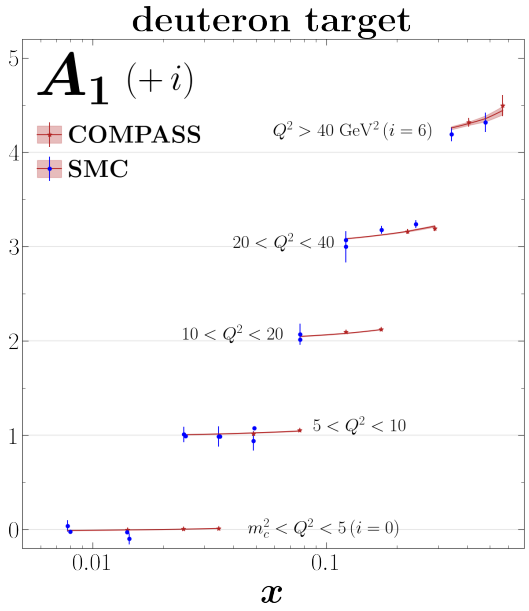


Figure B.5: Polarized DIS data for the observable A_1^D from COMPASS [178] and SMC [176, 177] plotted as a function of x against the JAM result (red 1σ bands). Data in different Q^2 bins are increased by i for clarity.

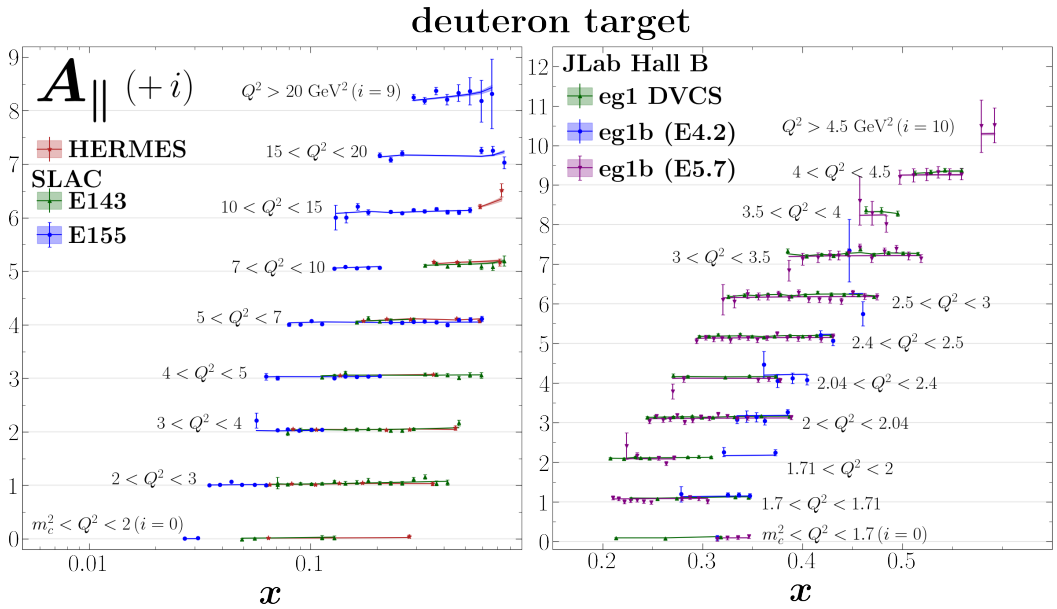


Figure B.6: Polarized DIS data for the observable $A_\parallel^D$ from from HERMES [188], SLAC [184, 185], and Jefferson Lab [191, 193] plotted as a function of x against the JAM result (colored 1σ bands). Data in different Q^2 bins are increased by i for clarity.

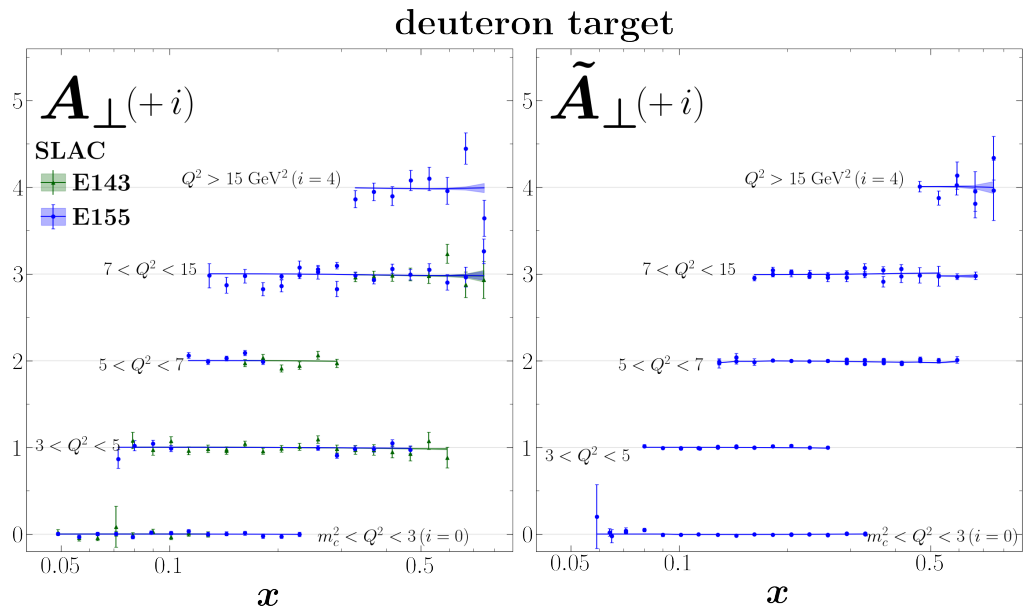


Figure B.7: Polarized DIS data for the observable $A_\perp^D$ (left panel) and $\widetilde{A}_\perp^D$ (right panel) from SLAC [184, 194, 195] plotted as a function of x against the JAM result (colored 1σ bands). Data in different Q^2 bins are increased by i for clarity.

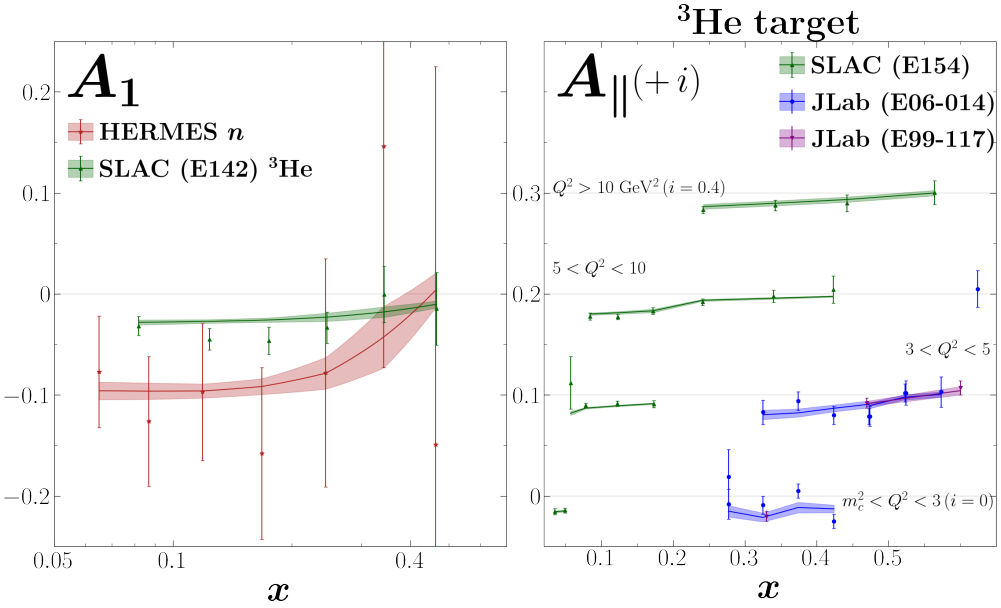


Figure B.8: Polarized DIS data for the observable $A_1^{n/^3\mathrm{He}}$ (left panel) and $A_\parallel^{^3\mathrm{He}}$ (right panel) from HERMES [187], SLAC [182, 183], and Jefferson Lab [189, 190] plotted as a function of x against the JAM result (colored 1σ bands). Data in different Q^2 bins are increased by i for clarity.

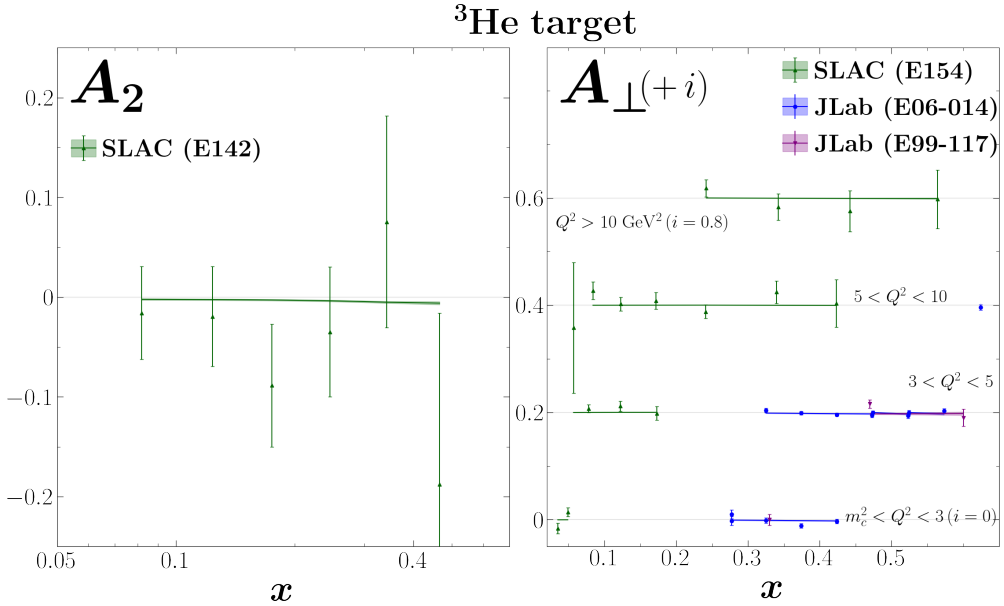


Figure B.9: Polarized DIS data for the observable $A_2^{^3\mathrm{He}}$ (left panel) and $A_\perp^{^3\mathrm{He}}$ (right panel) from SLAC [182, 183] and Jefferson Lab [189, 190] plotted as a function of x against the JAM result (colored 1σ bands). Data in different Q^2 bins are increased by i for clarity.

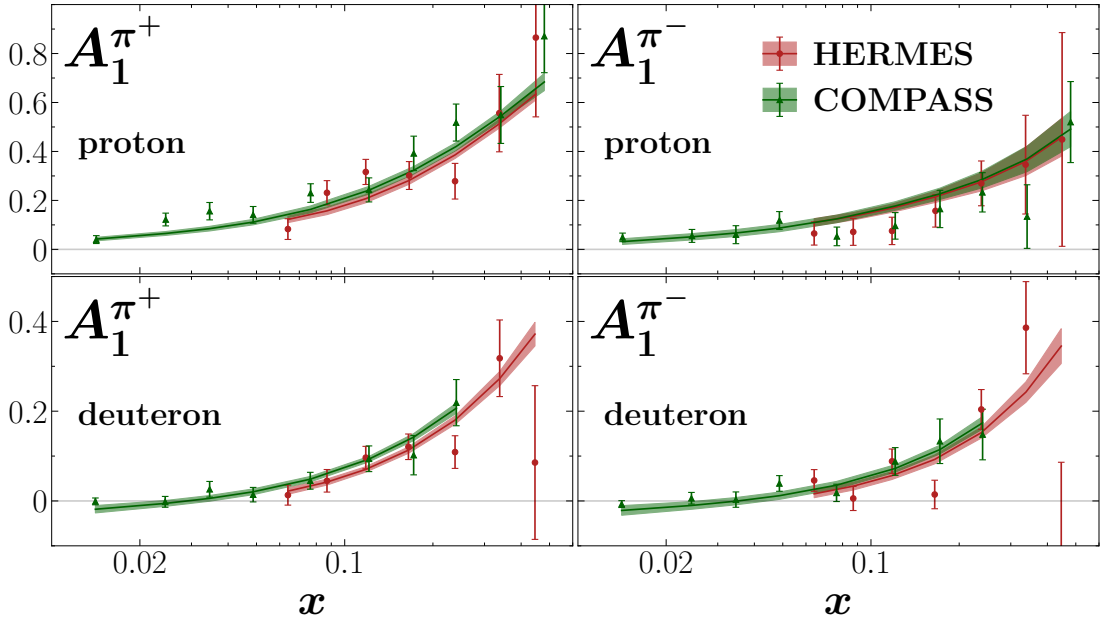


Figure B.10: Polarized SIDIS asymmetries $A_1^{\pi^\pm}$ from HERMES [198] (red circles) and COMPASS [199, 200] (green triangles) plotted as a function of x and compared to the JAM fit (1σ colored bands).

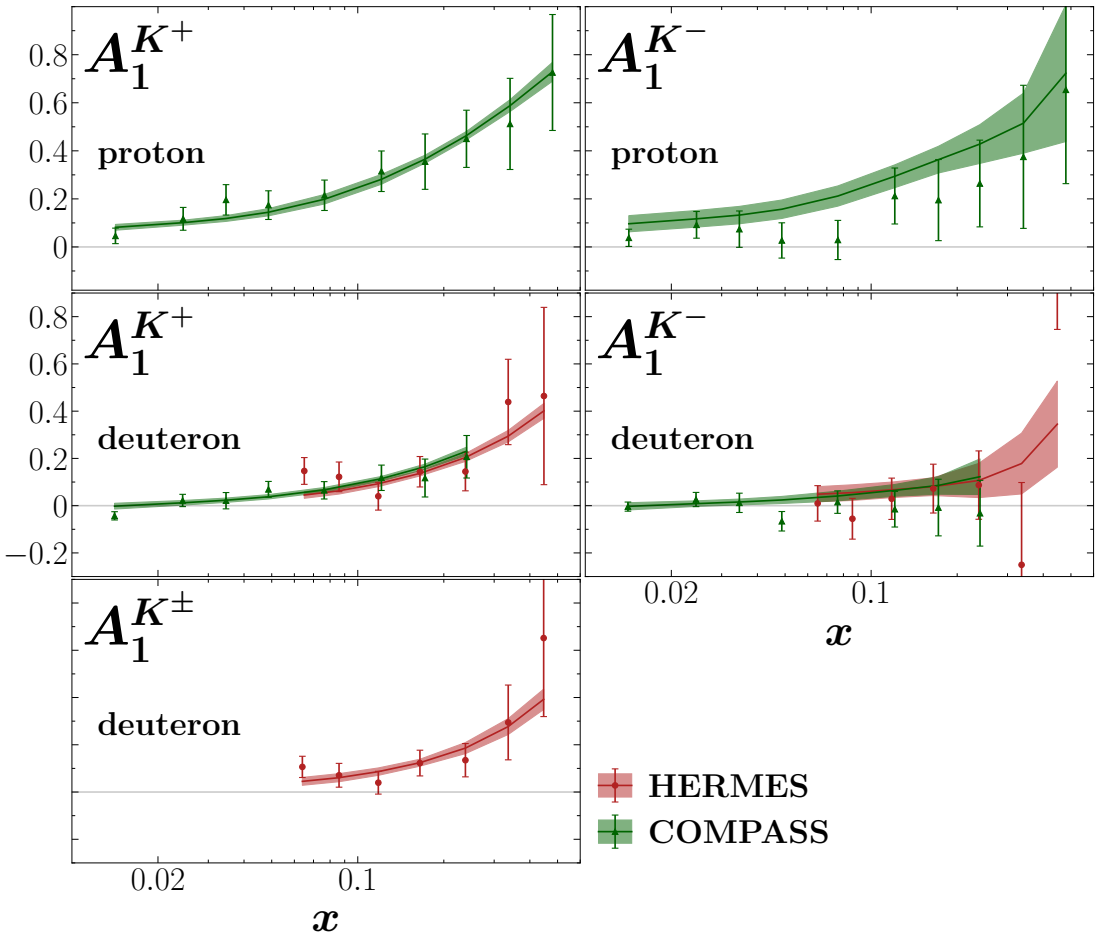


Figure B.11: Polarized SIDIS asymmetries $A_1^{K^\pm}$ from HERMES [198] (red circles) and COMPASS [199, 200] (green triangles) plotted as a function of x and compared to the JAM fit (1σ colored bands).

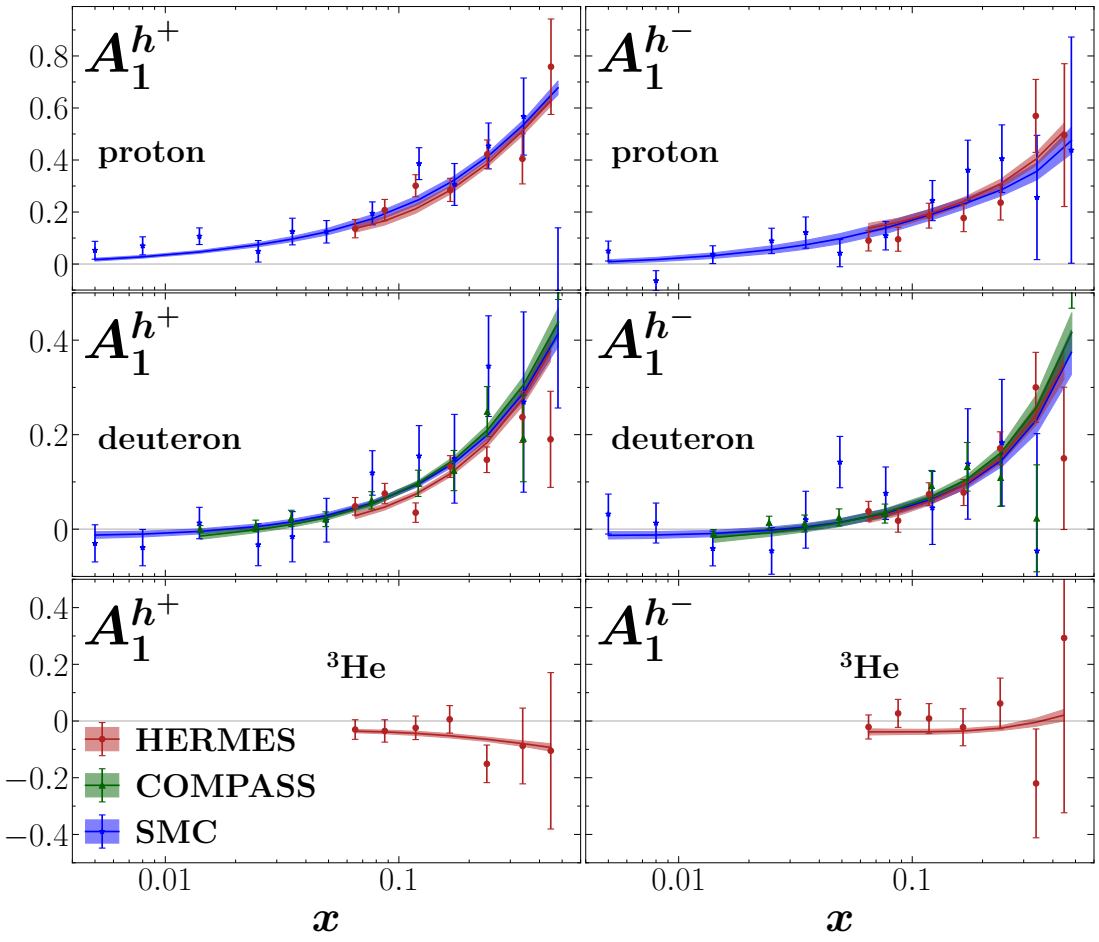


Figure B.12: Polarized proton (top row), deuteron (middle row), and ^{3}He (bottom row) SIDIS asymmetries $A_1^{h^\pm}$ from HERMES [198] (red circles), COMPASS [199] (green triangles), and SMC [374] (blue stars) as a function of x and compared to the JAM fit (1σ colored bands).

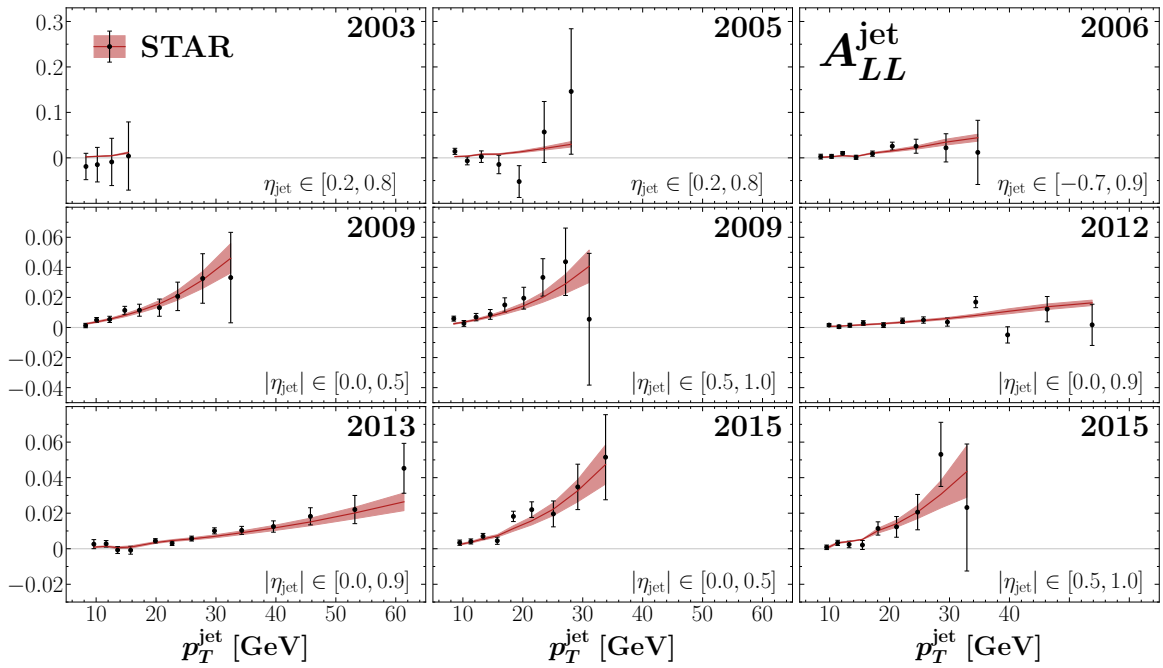


Figure B.13: Polarized double longitudinal jet asymmetries A_{LL}^{jet} from STAR [201, 202, 203, 121, 122, 123, 124] (black circles) plotted as a function of p_T^{jet} and compared to the JAM fit (red 1σ bands). Each subplot shows the year when the data was taken and the pseudorapidity bins η_{jet} or $|\eta_{\text{jet}}|$.

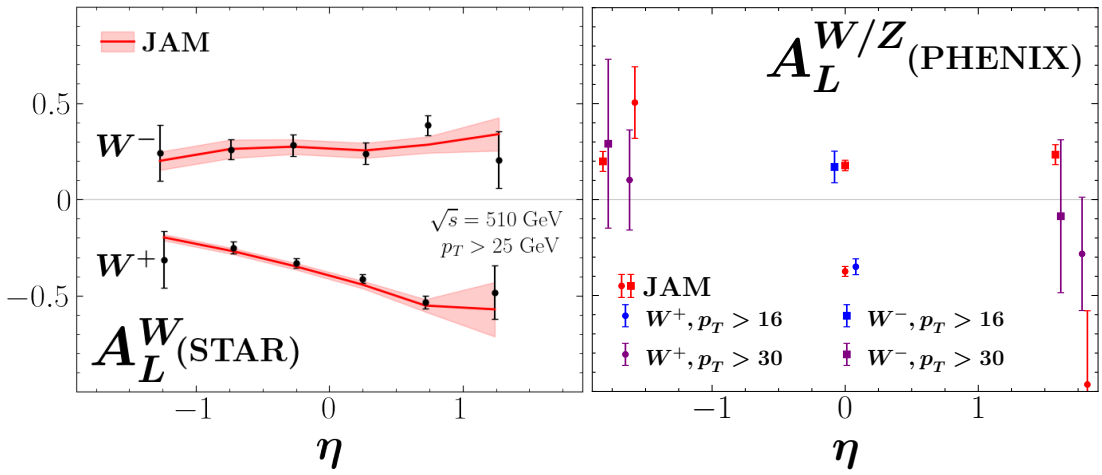


Figure B.14: Single spin asymmetries A_L^W and $A_L^{W/Z}$ versus lepton pseudorapidity η. *Left panel*: Asymmetries from STAR [204] (black circles) at $\sqrt{s} = 510$ GeV and integrated over $p_T > 25$ GeV, compared with the full JAM fit (red solid lines and 1σ uncertainty bands). *Right panel*: Asymmetries from PHENIX [205, 206] at $\sqrt{s} = 510$ GeV and integrated over $p_T > 16$ GeV (blue points) or $p_T > 30$ GeV (purple points), compared with the full JAM fit (red points). The W^+/Z asymmetries are shown with circles, while the W^-/Z asymmetries are shown with squares (we show points instead of bands due to the wide binning of the PHENIX data).

Appendix C

Accelerating Markov Chain Monte Carlo sampling with diffusion models

C.1 Hyperparameters

Here we outline each of the hyperparameters in `MCMC-diffusion`. If their values varied between test problems, the different settings are contained in Table C.1.

- The total number of retrains was set to 100 for each test problem.
- τ is the number of samples per retrain of the diffusion model. τ multiplied by (retrains) gives n, the total number of samples generated.
- p_{diff} is the probability of drawing a diffusion proposal sample rather than a Gaussian proposal sample.
- σ_{MH} is the width of the Gaussian proposal distribution.
- β is the amount of noise progressively added in the forward diffusion process, and is set to increase linearly from 0.1 to 0.3 for each test problem.
- `nsteps` is the number of noising steps in the forward/reverse diffusion process, and is set to 20 for each of the test problems.
- `bins` is the number of bins in the 1D histograms when calculating Eq. (5.4), and is set to 20 for each test problem.

Table C.1: MCMC-diffusion hyperparameter values assigned for each test problem.

Hyperparameter	2D Himmelblau	10D Gaussian	4D EggBox	4D PDF
τ	100	500	1000	100
p_{diff}	0.83	0.1	0.5	0.5
σ_{MH}	0.15	0.5	0.6	0.1

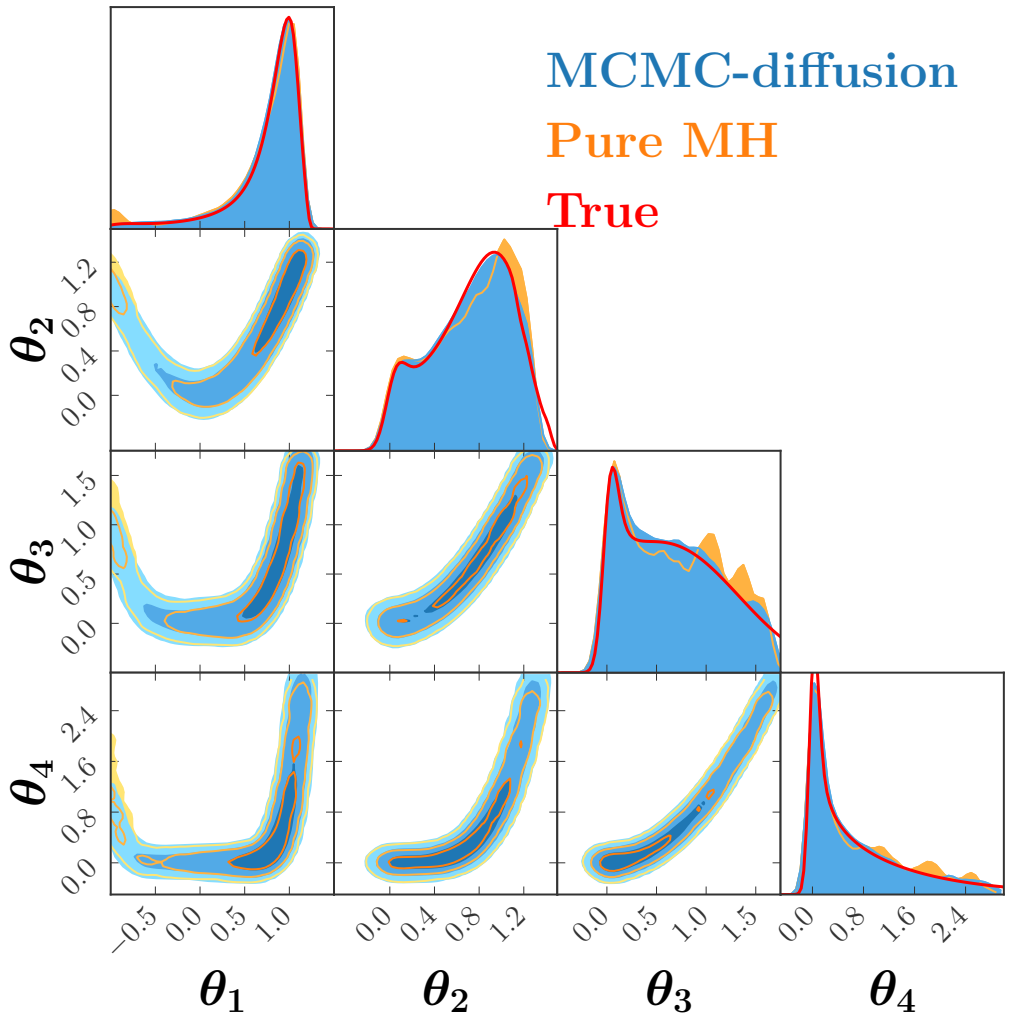


Figure C.1: Comparison of MCMC-diffusion and pure MH for the 4D Rosenbrock function (0.5M samples). The true marginalized posteriors are shown in red, coloured contours represent sample density.

C.2 4D Rosenbrock function

The Rosenbrock function is given by

$$f(\boldsymbol{\theta}) = \sum_{i=1}^{D-1} [100 \times (\theta_{i+1} - \theta_i)^2 + (1 - \theta_i)^2], \tag{C.1}$$

where D is the number of dimensions. This banana-shaped function consists of a single mode located within a long, flat bump. In 4D, it is particularly difficult to find this mode while also correctly mapping the curved shape of the surrounding area. Figure C.1 shows a comparison between pure MH and `MCMC-diffusion` for the 4D Rosenbrock function over 0.5M MCMC samples. Pure MH produces a false secondary mode, which is not present once the diffusion model is added in. Twice as many samples were required for pure MH to match `MCMC-diffusion` for this test function. One advantage the diffusion model has over pure MH is that it can make nonlocal jumps to key regions of interest and avoid getting stuck in the tails of the distribution, which explains why `MCMC-diffusion` is able to converge to the truth in fewer function evaluations.

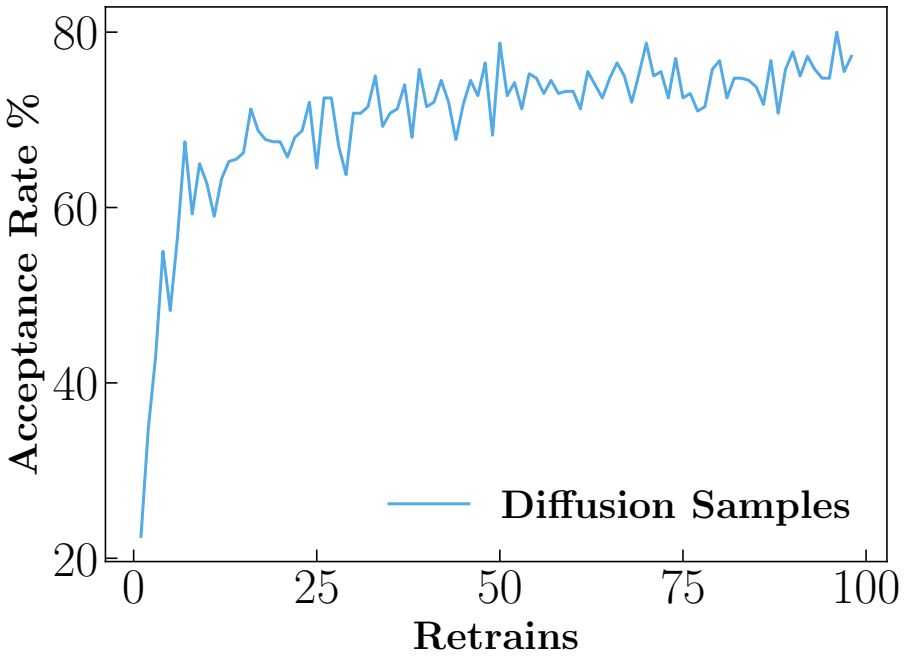


Figure C.2: Increase in diffusion sample acceptance rate with number of retrains for the 4D Rosenbrock function (0.5M samples).

We also show in Fig. C.2 a plot demonstrating asymptotic exactness for the Rosenbrock test function using `MCMC-diffusion`. As the MCMC chain progresses and the

diffusion model is repeatedly retrained, the acceptance rate increases well beyond 70%. This indicates that the proposal diffusion model is closely approximating the target posterior, as explained in Sec. 5.2.

Appendix D

A comparison of Bayesian sampling algorithms for particle astrophysics applications

D.1 Effective sample size

D.1.1 MCMC

The effective number of samples in an MCMC chain may be computed per parameter by taking into account the auto-correlation. We may define n_{eff} in this setting through the variance of a posterior mean,

$$\text{Var}(\bar{x}) = \frac{\text{Var}(x)}{n_{\text{eff}}} \tag{D.1}$$

This results in

$$n_{\text{eff}}^{\text{MCMC}} = \frac{n}{1 + 2\sum_{t=1}^{\infty} \rho(t)} \tag{D.2}$$

where $\rho(t)$ is the auto-correlation at lag t. We again use `arviz` [354] to compute this. This measure wouldn't be appropriate if one were particularly interested in samples from the tail of the distribution. To ensure that diagnostics and even n_{eff} itself may be computed reliably,it is reccommended in [357] that at least 400 effective samples for any scalar quantity of interest. Since the auto-correlation differs per parameter, as a conservative overall effective sample size we take the minimum n_{eff} among the model parameters.

D.1.2 Nested sampling

In NS, we estimate the effective sample size using a standard rule of thumb [377],

$$n_{\text{eff}}^{\text{Kish}} = \frac{1}{\sum_{i=1}^{N} w_i^2} \tag{D.3}$$

where $\sum w_{i=1}^{N} = 1$ and w_i are the NS weights. This is a well-known heuristic that unfortunately lacks formal justification [378] and should be used as a 'rough guide' [379]. We thus furthermore consider an information-theoretic alternative,

$$n_{\text{eff}}^{\text{Information}} = e^{-\sum_{i=1}^{N} w_i \log w_i} \tag{D.4}$$

In both cases, equally weighted samples, $w_i = 1/N$ result in $n_{\text{eff}} = N$, and we expect $n_{\text{eff}} \propto n_{\text{live}}$. Viewing them as expectations, by Jensen's inequality,

$$\langle w \rangle \geq e^{\langle \ln w \rangle} \tag{D.5}$$

such that the rule of thumb is more conservative, $n_{\text{eff}}^{\text{Kish}} \leq n_{\text{eff}}^{\text{Information}}$. To simulate the weights w_i we use `anesthetic` [380].

Unlike for MCMC, these are overall estimates, not per parameter. Thus, finally we build an estimator more directly comparable to the MCMC case per parameter by considering

$$n_{\text{eff}}^{\text{Bootstrap}} = \frac{\text{Var}(x)}{\text{Var}(\bar{x})} \tag{D.6}$$

where $\text{Var}(x)$ is the posterior variance of a parameter, and $\text{Var}(\bar{x})$ is the variance of our estimate of the posterior mean of the parameter. The latter takes into account two sources of variance: sampling of the compression factors, t, and sampling x from around a likelihood contour [381] using a bootstrap procedure in `nestcheck` [382, 383].

There is a computational subtlety about estimating $n_{\text{eff}}^{\text{Kish}}$ from NS results. If we write the weights as

$$w_i \propto L_i \Delta X_i = L_i X_i (1 - t_i) \tag{D.7}$$

such that

$$\sum_{i=1}^{N} w_i^2 \propto \sum_{i=1}^{N} L_i^2 X_i^2 (1 - t_i)^2 \tag{D.8}$$

and use estimators for $t_i \simeq \langle t \rangle$ in this equation, we typically overestimate $n_{\text{eff}}^{\text{Kish}}$ by about a factor of two as $\langle (1-t)^2 \rangle / (1 - \langle t \rangle)^2 \simeq 2$. A similar error occurs if using $\ln t_i \simeq \langle \ln t \rangle$.

Finally, chain sampler NS implementations produce a chain of phantom points when generating a draw from the constrained prior. These phantom points may contribute to the posterior samples, but due to their correlations their contribution to the effective sample size is challenging to quantify.

D.2 Posterior mean and standard deviation

D.2.1 Eggbox

Algorithm	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	θ_9	θ_{10}
True	15.7 ± 9.3	15.7 ± 9.3	15.7 ± 9.3	15.7 ± 9.3	15.7 ± 9.3	15.7 ± 9.3	15.7 ± 9.3	15.7 ± 9.3	15.7 ± 9.3	15.7 ± 9.3
MCMC: `emcee`	13.8 ± 7.8	15.1 ± 7.1	15.7 ± 8.6	15.3 ± 7.2	16.6 ± 8.5	15.1 ± 8.6	15.9 ± 7.7	16.5 ± 8.4	15.3 ± 7.8	16.1 ± 7.5
MCMC: `ptemcee`	15.9 ± 7.4	16.1 ± 7.8	15.8 ± 7.9	15.9 ± 8.2	15.5 ± 7.5	15.7 ± 7.6	15.3 ± 8.0	16.6 ± 8.1	16.2 ± 7.7	15.2 ± 8.2
MCMC: `zeus`	15.4 ± 9.2	15.5 ± 9.3	15.2 ± 9.4	15.7 ± 9.4	15.8 ± 9.2	15.8 ± 9.1	15.8 ± 9.2	16.5 ± 9.4	15.3 ± 9.2	16.1 ± 9.2
MCMC: `MCMC-diffusion`	16.5 ± 9.8	15.4 ± 9.3	15.1 ± 9.2	14.4 ± 8.7	15.2 ± 9.5	15.5 ± 9.1	14.6 ± 8.9	16.0 ± 8.5	15.7 ± 8.3	16.6 ± 9.4
NS: `dynesty`	15.6 ± 9.3	15.7 ± 9.4	15.5 ± 9.3	15.9 ± 9.2	15.9 ± 9.3	15.7 ± 9.4	15.5 ± 9.3	15.5 ± 9.4	15.7 ± 9.4	16.2 ± 9.4

Table D.1: Posterior mean for $10D$ Eggbox test function. The true mean for each parameter is 15.7 ± 9.3.

D.2.2 Rosenbrock

Algorithm	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	θ_9	θ_{10}
MCMC: `emcee`	0.66 ± 0.45	0.64 ± 0.32	0.52 ± 0.35	0.39 ± 0.35	0.28 ± 0.32	0.19 ± 0.27	0.12 ± 0.21	0.07 ± 0.15	0.03 ± 0.10	0.01 ± 0.08
MCMC: `ptemcee`	0.92 ± 0.33	0.94 ± 0.15	0.91 ± 0.19	0.86 ± 0.23	0.80 ± 0.28	0.72 ± 0.32	0.64 ± 0.36	0.55 ± 0.41	0.48 ± 0.49	0.46 ± 0.69
MCMC: `zeus`	0.65 ± 0.68	0.88 ± 0.24	0.84 ± 0.28	0.78 ± 0.32	0.72 ± 0.32	0.64 ± 0.38	0.56 ± 0.40	0.48 ± 0.42	0.41 ± 0.48	0.39 ± 0.63
MCMC: `MCMC-diffusion`	0.92 ± 0.31	0.93 ± 0.15	0.90 ± 0.20	0.86 ± 0.25	0.80 ± 0.28	0.73 ± 0.32	0.67 ± 0.34	0.56 ± 0.40	0.48 ± 0.49	0.50 ± 0.71
NS: `dynesty`	0.94 ± 0.22	0.93 ± 0.16	0.90 ± 0.20	0.86 ± 0.24	0.80 ± 0.29	0.72 ± 0.33	0.64 ± 0.37	0.55 ± 0.41	0.47 ± 0.48	0.45 ± 0.70
NS: `multinest`	0.91 ± 0.23	0.89 ± 0.20	0.83 ± 0.25	0.75 ± 0.30	0.66 ± 0.33	0.54 ± 0.34	0.42 ± 0.33	0.30 ± 0.30	0.18 ± 0.25	0.10 ± 0.20
NS: `nessai`	0.92 ± 0.19	0.89 ± 0.19	0.83 ± 0.24	0.76 ± 0.28	0.67 ± 0.31	0.54 ± 0.32	0.41 ± 0.31	0.27 ± 0.27	0.15 ± 0.21	0.07 ± 0.17
NS: `JaxNS`	0.93 ± 0.26	0.93 ± 0.14	0.89 ± 0.18	0.83 ± 0.22	0.73 ± 0.26	0.61 ± 0.29	0.46 ± 0.29	0.30 ± 0.26	0.16 ± 0.20	0.06 ± 0.15
NS: `nautilus`	0.93 ± 0.21	0.91 ± 0.18	0.87 ± 0.22	0.80 ± 0.27	0.72 ± 0.30	0.61 ± 0.33	0.49 ± 0.33	0.34 ± 0.31	0.21 ± 0.26	0.11 ± 0.20

Table D.2: Posterior mean for $10D$ Rosenbrock test function.

D.2.3 Rastrigin

Algorithm	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	θ_9	θ_{10}
True	0.00 ± 0.69	0.00 ± 0.69	0.00 ± 0.69	0.00 ± 0.69	0.00 ± 0.69	0.00 ± 0.69	0.00 ± 0.69	0.00 ± 0.69	0.00 ± 0.69	0.00 ± 0.69
MCMC: `emcee`	0.27 ± 0.93	-0.37 ± 0.92	-0.26 ± 0.85	-0.04 ± 0.99	0.37 ± 0.87	-0.09 ± 1.23	-0.53 ± 0.67	0.10 ± 0.65	0.22 ± 1.03	0.35 ± 1.10
MCMC: `ptemcee`	-0.03 ± 0.69	0.01 ± 0.74	0.03 ± 0.70	0.02 ± 0.69	0.00 ± 0.70	-0.02 ± 0.70	-0.05 ± 0.70	0.01 ± 0.69	0.03 ± 0.73	0.08 ± 0.67
MCMC: `zeus`	0.50 ± 2.39	-0.36 ± 2.65	0.03 ± 2.81	-0.36 ± 2.65	0.27 ± 2.67	-0.43 ± 3.00	-1.38 ± 2.70	-0.90 ± 2.97	-0.43 ± 2.60	0.57 ± 2.95
MC: `MCMC-diffusion`	-0.09 ± 0.82	0.24 ± 0.99	0.05 ± 0.72	0.14 ± 0.55	-0.21 ± 0.69	0.02 ± 0.71	0.17 ± 0.84	-0.38 ± 1.05	-0.06 ± 0.78	0.09 ± 0.72
NS: `JaxNS`	0.13 ± 0.71	0.09 ± 0.76	-0.27 ± 0.89	-0.07 ± 0.71	-0.05 ± 0.72	-0.16 ± 0.64	0.16 ± 0.74	-0.04 ± 0.56	-0.31 ± 0.71	-0.12 ± 0.76

Table D.3: Posterior mean for $10D$ Rastrigin test function. The true mean for each parameter is 0.

D.2.4 Spike-slab function

Algorithm	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	θ_9	θ_{10}
True	4.25 ± 1.84	4.25 ± 1.84	0.00 ± 0.57	0.00 ± 0.57	0.00 ± 0.57	0.00 ± 0.57	0.00 ± 0.57	0.00 ± 0.57	0.00 ± 0.57	0.00 ± 0.57
MCMC: `emcee`	2.60 ± 1.09	2.62 ± 1.08	0.01 ± 0.88	-0.00 ± 0.88	0.04 ± 0.88	0.01 ± 0.89	0.00 ± 0.87	-0.01 ± 0.88	-0.01 ± 0.87	0.00 ± 0.88
MCMC: `ptemcee`	2.52 ± 0.94	2.53 ± 0.93	0.01 ± 0.88	-0.01 ± 0.89	-0.03 ± 0.89	0.01 ± 0.88	0.00 ± 0.89	0.01 ± 0.89	-0.02 ± 0.89	0.02 ± 0.89
MCMC: `zeus`	3.20 ± 1.62	3.20 ± 1.62	0.01 ± 0.82	0.01 ± 0.82	0.0 ± 0.82	-0.01 ± 0.82	0.00 ± 0.82	0.00 ± 0.83	0.01 ± 0.82	-0.01 ± 0.82
CMC: `MCMC-diffusion`	4.38 ± 1.81	4.45 ± 1.92	-0.02 ± 0.72	-0.02 ± 0.72	0.03 ± 0.73	-0.01 ± 0.73	-0.00 ± 0.74	0.02 ± 0.73	0.00 ± 0.74	0.00 ± 0.74
NS: `Polychord`	4.80 ± 1.74	4.80 ± 1.76	0.00 ± 0.56	0.00 ± 0.56	0.00 ± 0.56	-0.01 ± 0.56	0.00 ± 0.57	0.01 ± 0.57	-0.02 ± 0.57	0.01 ± 0.57
NS: `JaxNS`	5.11 ± 1.60	5.11 ± 1.61	0.00 ± 0.50	0.00 ± 0.51	0.01 ± 0.50	-0.01 ± 0.51	0.00 ± 0.51	0.00 ± 0.50	0.00 ± 0.51	0.00 ± 0.51

Table D.4: Posterior mean and error for $10D$ spike-slab test function.

D.2.5 Cosmology-Planck

Algorithm	n_s	θ_s	ω_b	ω_c	τ	$\ln(A_s)$
MCMC: `emcee`	0.960 ± 0.001	1.04 ± 0.00	0.0222 ± 0.0001	0.119 ± 0.001	0.0616 ± 0.0021	3.08 ± 0.01
MCMC: `ptemcee`	0.960 ± 0.001	1.04 ± 0.00	0.0223 ± 0.0001	0.119 ± 0.001	0.0616 ± 0.0021	3.08 ± 0.01
MCMC: `zeus`	0.960 ± 0.001	1.04 ± 0.00	0.0223 ± 0.0001	0.119 ± 0.001	0.0617 ± 0.0021	3.08 ± 0.01
MCMC: `MCMC-diffusion`	0.960 ± 0.001	1.04 ± 0.00	0.0223 ± 0.0001	0.119 ± 0.001	0.0618 ± 0.0021	3.09 ± 0.01
NS: `Dynesty`	0.960 ± 0.002	1.04 ± 0.00	0.0223 ± 0.0001	0.119 ± 0.001	0.0617 ± 0.0021	3.08 ± 0.01
NS: `Multinest`	0.960 ± 0.002	1.04 ± 0.00	0.0223 ± 0.0001	0.119 ± 0.001	0.0617 ± 0.0020	3.08 ± 0.00
NS: `Polychord`	0.960 ± 0.001	1.04 ± 0.00	0.0223 ± 0.0001	0.119 ± 0.001	0.0615 ± 0.0021	3.08 ± 0.01
NS: `Nessai`	0.960 ± 0.002	1.04 ± 0.00	0.0223 ± 0.0001	0.119 ± 0.001	0.0615 ± 0.0020	3.08 ± 0.00
NS: `JaxNS`	0.960 ± 0.001	1.04 ± 0.00	0.0223 ± 0.0001	0.119 ± 0.001	0.0616 ± 0.0021	3.08 ± 0.00
NS: `Nautilus`	0.966 ± 0.000	1.04 ± 0.00	0.0201 ± 0.0001	0.105 ± 0.001	0.0529 ± 0.0001	3.39 ± 0.00
NS: `Ultranest`	0.960 ± 0.002	1.04 ± 0.00	0.0223 ± 0.0001	0.119 ± 0.001	0.0615 ± 0.0021	3.08 ± 0.00

Table D.5: Posterior mean and error for the CMB likelihood.

D.2.6 Particle physics - LHC

Algorithm	M_2	m_f^2	$m_{H_u}^2$	$m_{H_d}^2$	A_u	A_d	$\tan(\beta)$	m_t	α_s	ρ_0	σ
MCMC: `emcee`	1730 ± 3800	6.08 $\times 10^7 \pm$ 2.67 $\times 10^7$	-1.80 $\times 10^7 \pm$ 2.36 $\times 10^7$	-1.05 $\times 10^7 \pm$ 5.02 $\times 10^7$	1170 ± 6190	1500 ± 5240	31.5 ± 18.4	173 ± 1.13	0.119 ± 0.001	0.402 ± 0.143	42.3 ±
MCMC: `ptemcee`	-7160 ± 10	8.56 $\times 10^7 \pm$ 0.49 $\times 10^7$	-9.91 $\times 10^7 \pm$ 0.08 $\times 10^7$	9.66 $\times 10^7 \pm$ 0.34 $\times 10^7$	1510 ± 970	-9510 ± 500	69.9 ± 0.1	171 ± 0.01	0.120 ± 0.001	0.212 ± 0.012	63.1 ±
MCMC: `zeus`	4390 ± 3380	5.29 $\times 10^7 \pm$ 2.45 $\times 10^7$	-2.07 $\times 10^7 \pm$ 2.14 $\times 10^7$	-2.09 $\times 10^7 \pm$ 4.32 $\times 10^7$	2120 ± 7380	496 ± 5773	26.2 ± 18.6	173 ± 1.27	0.119 ± 0.001	0.424 ± 0.140	44.7 ±
MCMC: `MCMC-diffusion`	-7160 ± 10	8.55 $\times 10^7 \pm$ 0.50 $\times 10^7$	-9.91 $\times 10^7 \pm$ 0.08 $\times 10^7$	9.66 $\times 10^7 \pm$ 0.35 $\times 10^7$	1510 ± 990	-9510 ± 490	69.9 ± 0.1	171 ± 0.01	0.120 ± 0.001	0.212 ± 0.013	63.0 ±
NS: `Dynesty`	-7160 ± 10	8.57 $\times 10^7 \pm$ 0.49 $\times 10^7$	-9.90 $\times 10^7 \pm$ 0.07 $\times 10^7$	9.66 $\times 10^7 \pm$ 0.34 $\times 10^7$	1500 ± 970	-9520 ± 480	69.9 ± 0.1	171 ± 0.01	0.120 ± 0.001	0.213 ± 0.013	63.1 ±
NS: `JaxNS`	-7160 ± 10	9.38 $\times 10^7 \pm$ 0.54 $\times 10^7$	-6.45 $\times 10^6 \pm$ 5.45 $\times 10^6$	7.68 $\times 10^7 \pm$ 0.82 $\times 10^7$	-9860 ± 130	9670 ± 310	42.4 ± 3.0	171 ± 0.10	0.120 ± 0.001	0.201 ± 0.001	66.5 ±
NS: `Nautilus`	-7160 ± 10	8.55 $\times 10^7 \pm$ 0.50 $\times 10^7$	-9.91 $\times 10^7 \pm$ 0.07 $\times 10^7$	9.66 $\times 10^7 \pm$ 0.34 $\times 10^7$	1490 ± 1000	-9510 ± 490	69.9 ± 0.1	171 ± 0.01	0.120 ± 0.001	0.212 ± 0.012	63.1 ±
NS: `Ultranest`	-7160 ± 10	8.55 $\times 10^7 \pm$ 0.51 $\times 10^7$	-9.91 $\times 10^7 \pm$ 0.07 $\times 10^7$	9.66 $\times 10^7 \pm$ 0.35 $\times 10^7$	1510 ± 1000	-9510 ± 470	69.9 ± 0.1	171 ± 0.01	0.120 ± 0.001	0.213 ± 0.013	62.9 ±

Table D.6: Posterior mean and error for the MSSM7 likelihood.

Algorithm	n_s	θ_s	ω_b	ω_c	τ	$\ln(A_s)$
MCMC:emcee	0.0066	0.0031	0.0030	0.00081	0.0011	0.0011
MCMC:ptemcee	0.0046	0.0011	0.0010	0.00047	0.0016	0.0017
MCMC:zeus	0.0062	0.0036	0.0029	0.0020	0.0043	0.0035
MCMC:MCMC-diffusion	0.0063	0.0019	0.0015	0.00064	0.0015	0.0017
dynesty	0.0084	0.0014	0.0009	0.0007	0.0020	0.0022
jaxns	0.0136	0.0045	0.0013	0.0005	0.0022	0.0023
nautilus	0.0065	0.0029	0.0009	0.0005	0.0017	0.0027
nessai	0.0088	0.0013	0.0016	0.0008	0.0017	0.0024
pymultinest	0.0050	0.0011	0.0016	0.0006	0.0026	0.0025
pypolychord	0.0066	0.0021	0.0017	0.0006	0.0018	0.0018
ultranest	0.0053	0.0024	0.0017	0.0005	0.0026	0.0028

Table D.7: Marginal Wasserstein-1 distances for each parameter from reference posterior for CMB for cheapest converging NS runs. A '–' indicates none of the completed runs converged.

Algorithm	M_2	m_f^2	$m_{H_u}^2$	$m_{H_d}^2$	A_u	A_d	$\tan(\beta)$	m_t	α_s	ρ_0	σ_s	σ_l
MCMC:emcee	0.62	0.25	0.81	0.64	0.23	0.55	0.57	0.45	0.49	0.32	0.44	0.41
MCMC:ptemcee	0.0001	0.0013	0.0006	0.0019	0.0008	0.0003	0.0002	0.0002	0.0006	0.0023	0.0038	0.0031
MCMC:zeus	0.81	0.33	0.85	0.77	0.32	0.47	0.76	0.41	0.44	0.40	0.42	0.43
MCMC:MCMC-diffusion	0.0000	0.0024	0.0008	0.0026	0.0010	0.0007	0.0001	0.0001	0.0008	0.0023	0.0063	0.0029
dynesty	0.0000	0.0012	0.0005	0.0016	0.0011	0.0009	0.0002	0.0002	0.0004	0.0023	0.0027	0.0013
jaxns	0.0000	0.0096	0.0003	0.0027	0.0095	0.0030	0.0000	0.0001	0.0003	0.0004	0.0123	0.0336
nautilus	0.0001	0.0034	0.0005	0.0009	0.0025	0.0023	0.0004	0.0002	0.0010	0.0013	0.0099	0.0019
nessai	0.0001	0.0081	0.0007	0.0023	0.0050	0.0053	0.0005	0.0005	0.0014	0.0021	0.0178	0.0107
pymultinest	–	–	–	–	–	–	–	–	–	–	–	–
pypolychord	0.0001	0.0022	0.0011	0.0021	0.0027	0.0027	0.0004	0.0003	0.0015	0.0021	0.0087	0.0061
ultranest	0.0001	0.0019	0.0007	0.0019	0.0016	0.0017	0.0002	0.0001	0.0008	0.0024	0.0059	0.0033

Table D.8: Marginal Wasserstein-1 distances for each parameter from reference posterior for MSSM7 for cheapest converging NS runs. A '–' indicates none of the completed runs converged.

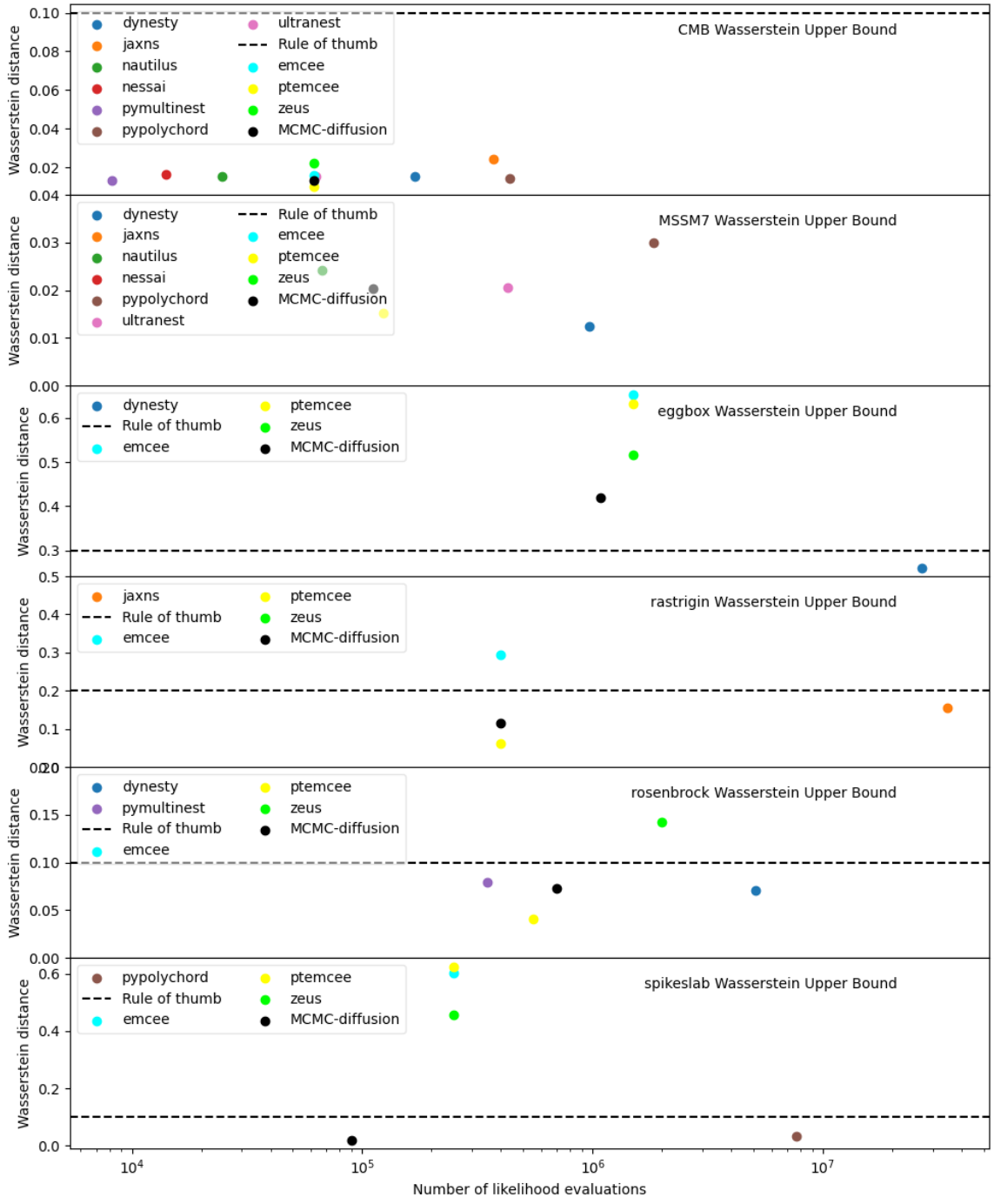


Figure D.1: Upper bound on Wasserstein distance from reference posterior for each problem. The black dashed line is the assumed rule of thumb for convergence for each problem in the unit-cube parameterised domain.

Algorithm	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	θ_9	θ_{10}
MCMC:emcee	0.092	0.087	0.048	0.087	0.043	0.048	0.065	0.046	0.057	0.082
MCMC:ptemcee	0.086	0.062	0.062	0.055	0.073	0.050	0.070	0.059	0.060	0.053
MCMC:zeus	0.066	0.028	0.11	0.040	0.042	0.044	0.078	0.032	0.030	0.050
MCMC:MCMC-diffusion	0.029	0.019	0.057	0.047	0.036	0.022	0.075	0.040	0.045	0.050
dynesty	0.0272	0.0175	0.0357	0.0354	0.0208	0.0189	0.0472	0.0134	0.0158	0.0292
jaxns	–	–	–	–	–	–	–	–	–	–
nautilus	–	–	–	–	–	–	–	–	–	–
nessai	–	–	–	–	–	–	–	–	–	–
pymultinest	–	–	–	–	–	–	–	–	–	–
pypolychord	–	–	–	–	–	–	–	–	–	–
ultranest	–	–	–	–	–	–	–	–	–	–

Table D.9: Marginal Wasserstein-1 distances for each parameter from reference posterior for Eggbox for cheapest converging NS runs. A '–' indicates none of the completed runs converged.

Algorithm	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	θ_9	θ_{10}
MCMC:emcee	0.037	0.027	0.027	0.026	0.027	0.026	0.036	0.026	0.034	0.030
MCMC:ptemcee	0.0090	0.0062	0.0071	0.011	0.0052	0.0048	0.0040	0.0024	0.0076	0.0040
MCMC:zeus	0.20	0.020	0.20	0.20	0.19	0.19	0.19	0.19	0.20	0.20
MCMC:MCMC-diffusion	0.013	0.0060	0.010	0.0072	0.015	0.011	0.0068	0.010	0.012	0.013
dynesty	–	–	–	–	–	–	–	–	–	–
jaxns	0.0193	0.0147	0.0391	0.0070	0.0130	0.0072	0.0220	0.0132	0.0151	0.0045
nautilus	–	–	–	–	–	–	–	–	–	–
nessai	–	–	–	–	–	–	–	–	–	–
pymultinest	–	–	–	–	–	–	–	–	–	–
pypolychord	–	–	–	–	–	–	–	–	–	–
ultranest	–	–	–	–	–	–	–	–	–	–

Table D.10: Marginal Wasserstein-1 distances for each parameter from reference posterior for rastrigin for cheapest converging NS runs. A '–' indicates none of the completed runs converged.

D.3 MCMC: Number of walkers

Algorithm	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	θ_9	θ_{10}
MCMC:emcee	0.028	0.031	0.041	0.049	0.055	0.058	0.058	0.055	0.052	0.059
MCMC:ptemcee	0.0025	0.0010	0.0014	0.0018	0.0025	0.0032	0.0041	0.0051	0.0070	0.012
MCMC:zeus	0.028	0.0066	0.0084	0.010	0.011	0.012	0.013	0.014	0.015	0.021
MCMC:MCMC-diffusion	0.0032	0.0012	0.0021	0.0036	0.0044	0.0050	0.0069	0.0098	0.016	0.021
dynesty	0.0011	0.0018	0.0026	0.0038	0.0054	0.0068	0.0082	0.0095	0.012	0.019
jaxns	–	–	–	–	–	–	–	–	–	–
nautilus	–	–	–	–	–	–	–	–	–	–
nessai	–	–	–	–	–	–	–	–	–	–
pymultinest	0.0021	0.0022	0.0031	0.0041	0.0056	0.0070	0.0086	0.011	0.014	0.023
pypolychord	–	–	–	–	–	–	–	–	–	–
ultranest	–	–	–	–	–	–	–	–	–	–

Table D.11: Marginal Wasserstein-1 distances for each parameter from reference posterior for rosenbrock for cheapest converging NS runs. A '–' indicates none of the completed runs converged.

Algorithm	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	θ_9	θ_{10}
MCMC:emcee	0.19	0.19	0.028	0.028	0.028	0.028	0.028	0.028	0.027	0.028
MCMC:ptemcee	0.20	0.20	0.028	0.029	0.029	0.029	0.029	0.029	0.029	0.029
MCMC:zeus	0.14	0.14	0.022	0.021	0.022	0.021	0.021	0.021	0.022	0.022
MCMC:MCMC-diffusion	0.0046	0.0041	0.0013	0.0013	0.0013	0.0016	0.0014	0.0016	0.0013	0.0014
dynesty	–	–	–	–	–	–	–	–	–	–
jaxns	–	–	–	–	–	–	–	–	–	–
nautilus	–	–	–	–	–	–	–	–	–	–
nessai	–	–	–	–	–	–	–	–	–	–
pymultinest	–	–	–	–	–	–	–	–	–	–
pypolychord	0.0072	0.0077	0.0021	0.0018	0.0020	0.0023	0.0023	0.0023	0.0021	0.0023
ultranest	–	–	–	–	–	–	–	–	–	–

Table D.12: Marginal Wasserstein-1 distances for each parameter from reference posterior for spikeslab for cheapest converging NS runs. A '–' indicates none of the completed runs converged.

Algorithm	Eggbox	Rosenbrock	Rastrigin	Spike-slab	Cosmology - Planck	Particle physics - LHC
MCMC: `zeus`	100	50	30	30	12	12
MCMC: `ptemcee`	100	100	30	30	12	12
MCMC: `emcee`	100	2000	30	30	12	12
MCMC: `MCMC-diffusion`	30	30	30	30	30	30

Table D.13: Number of MCMC walkers for each test problem.

Bibliography

[1] R. L. Workman et al. "Review of Particle Physics". In: *PTEP* 2022 (2022), p. 083C01. DOI: 10.1093/ptep/ptac097.

[2] Gerard 't Hooft. "Renormalization of Massless Yang-Mills Fields". In: *Nucl. Phys. B* 33 (1971), pp. 173–199. DOI: 10.1016/0550-3213(71)90395-6.

[3] David J. Gross and Frank Wilczek. "Ultraviolet Behavior of Nonabelian Gauge Theories". In: *Phys. Rev. Lett.* 30 (1973). Ed. by J. C. Taylor, pp. 1343–1346. DOI: 10.1103/PhysRevLett.30.1343.

[4] H. David Politzer. "Reliable Perturbative Results for Strong Interactions?" In: *Phys. Rev. Lett.* 30 (1973). Ed. by J. C. Taylor, pp. 1346–1349. DOI: 10.1103/PhysRevLett.30.1346.

[5] Peter W. Higgs. "Spontaneous Symmetry Breakdown without Massless Bosons". In: *Phys. Rev.* 145 (1966), pp. 1156–1163. DOI: 10.1103/PhysRev.145.1156.

[6] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. "Global Conservation Laws and Massless Particles". In: *Phys. Rev. Lett.* 13 (1964). Ed. by J. C. Taylor, pp. 585–587. DOI: 10.1103/PhysRevLett.13.585.

[7] F. Englert and R. Brout. "Broken Symmetry and the Mass of Gauge Vector Mesons". In: *Phys. Rev. Lett.* 13 (1964). Ed. by J. C. Taylor, pp. 321–323. DOI: 10.1103/PhysRevLett.13.321.

[8] Shehu S. AbdusSalam et al. "Simple and statistically sound strategies for analysing physical theories". In: (Dec. 2020). arXiv: 2012.09874 [hep-ph].

[9] P. Athron et al. "GAMBIT: The Global and Modular Beyond-the-Standard-Model Inference Tool". In: *Eur. Phys. J. C* 77 (May 2017), p. 784. arXiv: 1705.07908 [hep-ph].

[10] J. De Blas et al. "HEPfit: a Code for the Combination of Indirect and Direct Constraints on High Energy Physics Models". In: *Eur. Phys. J. C* 80 (2019), p. 456. DOI: 10.1140/epjc/s10052-020-7904-z. arXiv: 1910.14012 [hep-ph].

[11] O. Buchmueller et al. "Frequentist analysis of the parameter space of minimal supergravity". In: *Eur. Phys. J. C* 71, 1583 (Mar. 2011), p. 1583. DOI: 10.1140/epjc/s10052-011-1583-8. arXiv: 1011.6118 [hep-ph].

[12] Ivan Esteban et al. "NuFit 3.2". http://www.nu-fit.org (accessed 2018-05-08). 2018.

[13] Andrea Barontini et al. "Pineline: Industrialization of High-Energy Theory Predictions". In: (Feb. 2023). arXiv: 2302.12124 [hep-ph].

[14] H. Abdolmaleki et al. "xFitter: An Open Source QCD Analysis Framework. A resource and reference document for the Snowmass study". In: *DPF Snowmass Community Planning Exercise*. June 2022. arXiv: 2206.12465 [hep-ph].

[15] DianYu Liu, ChuanLe Sun, and Jun Gao. "Machine learning of log-likelihood functions in global analysis of parton distributions". In: *JHEP* 08 (2022), p. 088. DOI: 10.1007/JHEP08(2022)088. arXiv: 2201.06586 [hep-ph].

[16] S. Brown, G. Niculescu, and I. Niculescu. "Machine learning representation of the F2 structure function over all charted Q^2 and x range". In: *Phys. Rev. C* 104.6 (2021), p. 064321. DOI: 10.1103/PhysRevC.104.064321. arXiv: 2106.06390 [hep-ph].

[17] Maria Concepcion Gonzalez-Garcia, Michele Maltoni, and Thomas Schwetz. "NuFIT: Three-Flavour Global Analyses of Neutrino Oscillation Experiments". In: *Universe* 7.12 (2021), p. 459. DOI: 10.3390/universe7120459. arXiv: 2111.03086 [hep-ph].

[18] Ivan Esteban et al. "The fate of hints: updated global analysis of three-flavor neutrino oscillations". In: *JHEP* 09 (2020), p. 178. DOI: 10.1007/JHEP09(2020)178. arXiv: 2007.14792 [hep-ph].

[19] Kevin J. Kelly et al. "Neutrino mass ordering in light of recent data". In: *Phys. Rev. D* 103.1 (2021), p. 013004. DOI: 10.1103/PhysRevD.103.013004. arXiv: 2007.08526 [hep-ph].

[20] Francesco Capozzi et al. "Global constraints on absolute neutrino masses and their ordering". In: *Phys. Rev. D* 95.9 (2017). [Addendum: Phys.Rev.D 101, 116013 (2020)], p. 096014. DOI: 10.1103/PhysRevD.95.096014. arXiv: 2003.08511 [hep-ph].

[21] Juan Carlos Criado, Roman Kogler, and Michael Spannowsky. "Quantum fitting framework applied to effective field theories". In: *Phys. Rev. D* 107.1 (2023), p. 015023. DOI: 10.1103/PhysRevD.107.015023. arXiv: 2207.10088 [hep-ph].

[22] José Eliel Camargo-Molina, Rikard Enberg, and Johan Löfgren. "A new perspective on the electroweak phase transition in the Standard Model Effective Field Theory". In: *JHEP* 10 (2021), p. 127. DOI: 10.1007/JHEP10(2021)127. arXiv: 2103.14022 [hep-ph].

[23] John Ellis et al. "Top, Higgs, Diboson and Electroweak Fit to the Standard Model Effective Field Theory". In: *JHEP* 04 (2021), p. 279. DOI: 10.1007/JHEP04(2021)279. arXiv: 2012.02779 [hep-ph].

[24] Rafael Aoude et al. "The impact of flavour data on global fits of the MFV SMEFT". In: *JHEP* 12 (2020), p. 113. DOI: 10.1007/JHEP12(2020)113. arXiv: 2003.05432 [hep-ph].

[25] John Ellis et al. "The CMSSM survives Planck, the LHC, LUX-ZEPLIN, Fermi-LAT, H.E.S.S. and IceCube". In: *Eur. Phys. J. C* 83.3 (2023), p. 246. DOI: 10.1140/epjc/s10052-023-11405-1. arXiv: 2210.16337 [hep-ph].

[26] Jennet Dickinson et al. "A Grand Scan of the pMSSM Parameter Space for Snowmass 2021". In: (July 2022). arXiv: 2207.05103 [hep-ph].

[27] Henning Bahl et al. "Constraining the $\mathcal{CP}$ structure of Higgs-fermion couplings with a global LHC fit, the electron EDM and baryogenesis". In: *Eur. Phys. J. C* 82.7 (2022), p. 604. DOI: 10.1140/epjc/s10052-022-10528-1. arXiv: 2202.11753 [hep-ph].

[28] Andreas Crivellin, Dario Müller, and Luc Schnell. "Combined constraints on first generation leptoquarks". In: *Phys. Rev. D* 103.11 (2021). [Addendum: Phys.Rev.D 104, 055020 (2021)], p. 115023. DOI: 10.1103/PhysRevD.103.115023. arXiv: 2101.07811 [hep-ph].

[29] Michael J. Ramsey-Musolf and Juan Carlos Vasquez. "Left-right symmetry and electric dipole moments. A global analysis". In: *Phys. Lett. B* 815 (2021), p. 136136. DOI: 10.1016/j.physletb.2021.136136. arXiv: 2012.02799 [hep-ph].

[30] Margarete Mühlleitner et al. "ScannerS: parameter scans in extended scalar sectors". In: *Eur. Phys. J. C* 82.3 (2022), p. 198. DOI: 10.1140/epjc/s10052-022-10139-w. arXiv: 2007.02985 [hep-ph].

[31] Stephen P. Martin. "A Supersymmetry primer". In: *Adv. Ser. Direct. High Energy Phys.* 18 (1998). Ed. by Gordon L. Kane, pp. 1–98. DOI: 10.1142/9789812839657_0001. arXiv: hep-ph/9709356.

[32] John C. Collins, Davison E. Soper, and George F. Sterman. "Factorization of Hard Processes in QCD". In: *Adv. Ser. Direct. High Energy Phys.* 5 (1989), pp. 1–91. DOI: 10.1142/9789814503266_0001. arXiv: hep-ph/0409313.

[33] Yuri L. Dokshitzer. "Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum Chromodynamics." In: *Sov. Phys. JETP* 46 (1977), pp. 641–653.

[34] V. N. Gribov and L. N. Lipatov. "Deep inelastic e p scattering in perturbation theory". In: *Sov. J. Nucl. Phys.* 15 (1972), pp. 438–450.

[35] Guido Altarelli and G. Parisi. "Asymptotic Freedom in Parton Language". In: *Nucl. Phys. B* 126 (1977), pp. 298–318. DOI: 10.1016/0550-3213(77)90384-4.

[36] J. T. Londergan, J. C. Peng, and A. W. Thomas. "Charge Symmetry at the Partonic Level". In: *Rev. Mod. Phys.* 82 (2010), pp. 2009–2052. DOI: 10.1103/RevModPhys.82.2009. arXiv: 0907.2352 [hep-ph].

[37] E. G. Floratos, C. Kounnas, and R. Lacaze. "Higher Order QCD Effects in Inclusive Annihilation and Deep Inelastic Scattering". In: *Nucl. Phys. B* 192 (1981), pp. 417–462. DOI: 10.1016/0550-3213(81)90434-X.

[38] Karol Kovařik, Pavel M. Nadolsky, and Davison E. Soper. "Hadronic structure in high-energy collisions". In: *Rev. Mod. Phys.* 92.4 (2020), p. 045003. DOI: 10.1103/RevModPhys.92.045003. arXiv: 1905.06957 [hep-ph].

[39] J. Pumplin et al. "Uncertainties of predictions from parton distribution functions. 2. The Hessian method". In: *Phys. Rev. D* 65 (2001), p. 014013. DOI: 10.1103/PhysRevD.65.014013. arXiv: hep-ph/0101032 [hep-ph].

[40] D. Stump et al. "Uncertainties of predictions from parton distribution functions. 1. The Lagrange multiplier method". In: *Phys. Rev. D* 65 (2001), p. 014012. DOI: 10.1103/PhysRevD.65.014012. arXiv: hep-ph/0101051 [hep-ph].

[41] Tie-Jiun Hou et al. "New CTEQ global analysis of quantum chromodynamics with high-precision data from the LHC". In: *Phys. Rev. D* 103.1 (2021), p. 014013. DOI: 10.1103/PhysRevD.103.014013. arXiv: 1912.10053 [hep-ph].

[42] S. Bailey et al. "Parton distributions from LHC, HERA, Tevatron and fixed target data: MSHT20 PDFs". In: *Eur. Phys. J. C* 81.4 (2021), p. 341. DOI: 10.1140/epjc/s10052-021-09057-0. arXiv: 2012.04684 [hep-ph].

[43] Jun Gao, Lucian Harland-Lang, and Juan Rojo. "The Structure of the Proton in the LHC Precision Era". In: *Phys. Rep.* 742 (2018), p. 1. DOI: 10.1016/j.physrep.2018.03.002. arXiv: 1709.04922 [hep-ph].

[44] Daniel de Florian et al. "Global Analysis of Helicity Parton Densities and Their Uncertainties". In: *Phys. Rev. Lett.* 101 (2008), p. 072001. DOI: 10.1103/PhysRevLett.101.072001. arXiv: 0804.0422 [hep-ph].

[45] Daniel de Florian et al. "Extraction of Spin-Dependent Parton Densities and Their Uncertainties". In: *Phys. Rev. D* 80 (2009), p. 034030. DOI: 10.1103/PhysRevD.80.034030. arXiv: 0904.3821 [hep-ph].

[46] Daniel de Florian et al. "Evidence for polarization of gluons in the proton". In: *Phys. Rev. Lett.* 113.1 (2014), p. 012001. DOI: 10.1103/PhysRevLett.113.012001. arXiv: 1404.4293 [hep-ph].

[47] Stefano Forte et al. "Neural network parametrization of deep inelastic structure functions". In: *JHEP* 05 (2002), p. 062. DOI: 10.1088/1126-6708/2002/05/062. arXiv: hep-ph/0204232 [hep-ph].

[48] Luigi Del Debbio et al. "Unbiased determination of the proton structure function F(2)**p with faithful uncertainty estimation". In: *JHEP* 03 (2005), p. 080. DOI: 10.1088/1126-6708/2005/03/080. arXiv: hep-ph/0501067 [hep-ph].

[49] Luigi Del Debbio et al. "Neural network determination of parton distributions: The Nonsinglet case". In: *JHEP* 03 (2007), p. 039. DOI: 10.1088/1126-6708/2007/03/039. arXiv: hep-ph/0701127 [hep-ph].

[50] Richard D. Ball et al. "A Determination of parton distributions with faithful uncertainty estimation". In: *Nucl. Phys.* B809 (2009). [Erratum: Nucl. Phys. **B816**, 293 (2009)], pp. 1–63. DOI: 10.1016/j.nuclphysb.2008.09.037,10.1016/j.nuclphysb.2009.02.027. arXiv: 0808.1231 [hep-ph].

[51] Richard D. Ball et al. "The path to proton structure at 1% accuracy". In: *Eur. Phys. J. C* 82.5 (2022), p. 428. DOI: 10.1140/epjc/s10052-022-10328-7. arXiv: 2109.02653 [hep-ph].

[52] Emanuele R. Nocera et al. "A first unbiased global determination of polarized PDFs and their uncertainties". In: *Nucl. Phys.* B887 (2014), pp. 276–308. DOI: 10.1016/j.nuclphysb.2014.08.008. arXiv: 1406.5539 [hep-ph].

[53] Nobuo Sato et al. "Iterative Monte Carlo analysis of spin-dependent parton distributions". In: *Phys. Rev. D* 93.7 (2016), p. 074005. DOI: 10.1103/PhysRevD.93.074005. arXiv: 1601.07782 [hep-ph].

[54] C. Cocuzza et al. "Isovector EMC Effect from Global QCD Analysis with MARATHON Data". In: *Phys. Rev. Lett.* 127.24 (2021), p. 242001. DOI: 10.1103/PhysRevLett.127.242001. arXiv: 2104.06946 [hep-ph].

[55] C. Cocuzza et al. "Bayesian Monte Carlo extraction of the sea asymmetry with SeaQuest and STAR data". In: *Phys. Rev. D* 104.7 (2021), p. 074031. DOI: 10.1103/PhysRevD.104.074031. arXiv: 2109.00677 [hep-ph].

[56] N. Y. Cao et al. "Towards the three-dimensional parton structure of the pion: Integrating transverse momentum data into global QCD analysis". In: *Phys. Rev. D* 103.11 (2021), p. 114014. DOI: 10.1103/PhysRevD.103.114014. arXiv: 2103.02159 [hep-ph].

[57] P. C. Barry et al. "Global QCD Analysis of Pion Parton Distributions with Threshold Resummation". In: *Phys. Rev. Lett.* 127.23 (2021), p. 232001. DOI: 10.1103/PhysRevLett.127.232001. arXiv: 2108.05822 [hep-ph].

[58] John Skilling. "Nested sampling for general Bayesian computation". In: *Bayesian Anal.* 1.4 (Dec. 2006), pp. 833–859. DOI: 10.1214/06-BA127. URL: https://doi.org/10.1214/06-BA127.

[59] F. Feroz and M. P. Hobson. "Multimodal nested sampling: an efficient and robust alternative to Markov Chain Monte Carlo methods for astronomical data analyses". In: *Monthly Notices of the Royal Astronomical Society* 384.2 (2008), pp. 449–463. DOI: 10.1111/j.1365-2966.2007.12353.x.

[60] F. Feroz, M. P. Hobson, and M. Bridges. "MultiNest: an efficient and robust Bayesian inference tool for cosmology and particle physics". In: *Monthly Notices of the Royal Astronomical Society* 398.4 (2009), pp. 1601–1614. DOI: 10.1111/j.1365-2966.2009.14548.x.

[61] W. J. Handley, M. P. Hobson, and A. N. Lasenby. "polychord: nested sampling for cosmology". In: *Monthly Notices of the Royal Astronomical Society: Letters* 450.1 (2015), pp. L61–L65. DOI: 10.1093/mnrasl/slv047. URL: http://dx.doi.org/10.1093/mnrasl/slv047.

[62] Edward Higson et al. *Sampling Errors in Nested Sampling Parameter Estimation*. Advance publication. 2017. DOI: 10.1214/17-BA1075. URL: https://doi.org/10.1214/17-BA1075.

[63] Joao Alves et al. "Planck 2015 results". In: *Astronomy & Astrophysics* 594 (2016), E1. DOI: 10.1051/0004-6361/201629543. URL: https://doi.org/10.1051/0004-6361/201629543.

[64] Huey-Wen Lin et al. "First Monte Carlo Global Analysis of Nucleon Transversity with Lattice QCD Constraints". In: *Phys. Rev. Lett.* 120.15 (2018), p. 152502. DOI: 10.1103/PhysRevLett.120.152502. arXiv: 1710.09858 [hep-ph].

[65] P. C. Barry et al. "First Monte Carlo Global QCD Analysis of Pion Parton Distributions". In: *Phys. Rev. Lett.* 121.15 (2018), p. 152001. DOI: 10.1103/PhysRevLett.121.152001. arXiv: 1804.01965 [hep-ph].

[66] S. Duane et al. "Hybrid Monte Carlo". In: *Phys. Lett. B* 195 (1987), pp. 216–222. DOI: 10.1016/0370-2693(87)91197-X.

[67] Steve Brooks et al. *Handbook of Markov Chain Monte Carlo*. CRC press, 2011.

[68] N. Metropolis et al. "Equation of state calculations by fast computing machines". In: *J. Chem. Phys.* 21 (1953), pp. 1087–1092. DOI: 10.1063/1.1699114.

[69] W. K. Hastings. "Monte Carlo Sampling Methods Using Markov Chains and Their Applications". In: *Biometrika* 57 (1970), pp. 97–109. DOI: 10.1093/biomet/57.1.97.

[70] S. Alekhin et al. "HERAFitter". In: *Eur. Phys. J. C* 75.7 (2015), p. 304. DOI: 10.1140/epjc/s10052-015-3480-z. arXiv: 1410.4412 [hep-ph].

[71] Yémalin Gabin Gbedo and Mariane Mangin-Brinet. "Markov chain Monte Carlo techniques applied to parton distribution functions determination: Proof of concept". In: *Phys. Rev. D* 96.1 (2017), p. 014015. DOI: 10.1103/PhysRevD.96.014015. arXiv: 1701.07678 [hep-ph].

[72] Shayan Iranipour and Maria Ubiali. "A new generation of simultaneous fits to LHC data using deep learning". In: *JHEP* 05 (2022), p. 032. DOI: 10.1007/JHEP05(2022)032. arXiv: 2201.07240 [hep-ph].

[73] Luigi Del Debbio, Tommaso Giani, and Michael Wilson. "Bayesian approach to inverse problems: an application to NNPDF closure testing". In: *Eur. Phys. J. C* 82.4 (2022), p. 330. DOI: 10.1140/epjc/s10052-022-10297-x. arXiv: 2111.05787 [hep-ph].

[74] S. Camarda et al. "QCD analysis of W- and Z-boson production at Tevatron". In: *Eur. Phys. J. C* 75.9 (2015), p. 458. DOI: 10.1140/epjc/s10052-015-3655-7. arXiv: 1503.05221 [hep-ph].

[75] J. de Blas et al. "Global analysis of electroweak data in the Standard Model". In: *Phys. Rev. D* 106.3 (2022), p. 033003. DOI: 10.1103/PhysRevD.106.033003. arXiv: 2112.07274 [hep-ph].

[76] Andrzej J. Buras et al. "Global analysis of leptophilic Z' bosons". In: *JHEP* 06 (2021), p. 068. DOI: 10.1007/JHEP06(2021)068. arXiv: 2104.07680 [hep-ph].

[77] Ethan Carragher et al. "Convergent Bayesian Global Fits of 4D Composite Higgs Models". In: *JHEP* 05 (2021), p. 237. DOI: 10.1007/JHEP05(2021)237. arXiv: 2101.00428 [hep-ph].

[78] Anisha et al. "Connecting electroweak-scale observables to BSM physics through EFT and Bayesian statistics". In: *Phys. Rev. D* 103.7 (2021), p. 076007. DOI: 10.1103/PhysRevD.103.076007. arXiv: 2010.04088 [hep-ph].

[79] Andreas Crivellin et al. "Global Electroweak Fit and Vector-Like Leptons in Light of the Cabibbo Angle Anomaly". In: *JHEP* 12 (2020), p. 166. DOI: 10.1007/JHEP12(2020)166. arXiv: 2008.01113 [hep-ph].

[80] Kamila Kowalska, Dinesh Kumar, and Enrico Maria Sessolo. "Implications for new physics in $b \to s\mu\mu$ transitions after recent measurements by Belle and LHCb". In: *The European Physical Journal C* 79.10 (Oct. 2019). DOI: 10.1140/epjc/s10052-019-7330-2. URL: https://doi.org/10.1140%2Fepjc%2Fs10052-019-7330-2.

[81] P. F. de Salas et al. "2020 global reassessment of the neutrino oscillation picture". In: *JHEP* 02 (2021), p. 071. DOI: 10.1007/JHEP02(2021)071. arXiv: 2006.11237 [hep-ph].

[82] Stefano Dell'Oro, Simone Marcocci, and Francesco Vissani. "Empirical Inference on the Majorana Mass of the Ordinary Neutrinos". In: *Phys. Rev. D* 100.7 (2019), p. 073003. DOI: 10.1103/PhysRevD.100.073003. arXiv: 1909.05381 [hep-ph].

[83] J. Bergström et al. "Bayesian global analysis of neutrino oscillation data". In: *JHEP* 9, 200 (Sept. 2015), p. 200. DOI: 10.1007/JHEP09(2015)200. arXiv: 1507.04366 [hep-ph].

[84] D. Bose et al. "Bayesian approach for counting experiment statistics applied to a neutrino point source analysis". In: *Astropart. Phys.* 50-52 (2013), pp. 57–64. DOI: 10.1016/j.astropartphys.2013.09.009. arXiv: 1212.2008 [astro-ph.HE].

[85] Andrés Romero-Wolf and Máximo Ave. "Bayesian Inference Constraints on Astrophysical Production of Ultra-high Energy Cosmic Rays and Cosmogenic Neutrino Flux Predictions". In: *JCAP* 07 (2018), p. 025. DOI: 10.1088/1475-7516/2018/07/025. arXiv: 1712.07290 [astro-ph.HE].

[86] Francesca Capel, Daniel J. Mortlock, and Chad Finley. "Bayesian constraints on the astrophysical neutrino source population from IceCube data". In: *Phys. Rev. D* 101.12 (2020). [Erratum: Phys.Rev.D 105, 129904 (2022)], p. 123017. DOI: 10.1103/PhysRevD.101.123017. arXiv: 2005.02395 [astro-ph.HE].

[87] Andrew Fowlie. "DAMPE squib? Significance of the 1.4 TeV DAMPE excess". In: *Phys. Lett. B* 780 (2018), pp. 181–184. DOI: 10.1016/j.physletb.2018.03.006. arXiv: 1712.05089 [hep-ph].

[88] P. Athron et al. "Status of the scalar singlet dark matter model". In: *Eur. Phys. J. C* 77.8 (May 2017), p. 568. arXiv: 1705.07931 [hep-ph].

[89] Peter Athron et al. "Impact of vacuum stability, perturbativity and XENON1T on global fits of $\mathbb{Z}_2$ and $\mathbb{Z}_3$ scalar singlet dark matter". In: *Eur. Phys. J. C* 78.10 (2018), p. 830. DOI: 10.1140/epjc/s10052-018-6314-y. arXiv: 1806.11281 [hep-ph].

[90] Peter Athron et al. "Global analyses of Higgs portal singlet dark matter models using GAMBIT". In: *Eur. Phys. J. C* 79.1 (2019), p. 38. DOI: 10.1140/epjc/s10052-018-6513-6. arXiv: 1808.10465 [hep-ph].

[91] S. Liem et al. "Effective field theory of dark matter: a global analysis". In: *JHEP* 9, 77 (Sept. 2016), p. 77. DOI: 10.1007/JHEP09(2016)077. arXiv: 1603.05994 [hep-ph].

[92] Gregory Ashton and Colm Talbot. "B ilby-MCMC: an MCMC sampler for gravitational-wave inference". In: *Mon. Not. Roy. Astron. Soc.* 507.2 (2021), pp. 2037–2051. DOI: 10.1093/mnras/stab2236. arXiv: 2106.08730 [gr-qc].

[93] Gregory Ashton and Colm Talbot. *Bilby MCMC Guide*. URL: https://lscsoft.docs.ligo.org/bilby/bilby-mcmc-guide.html. (accessed: 10.10.2022).

[94] Gregory Ashton et al. "BILBY: A user-friendly Bayesian inference library for gravitational-wave astronomy". In: *Astrophys. J. Suppl.* 241.2 (2019), p. 27. DOI: 10.3847/1538-4365/ab06fc. arXiv: 1811.02042 [astro-ph.IM].

[95] Peter Athron et al. "GAMBIT: The Global and Modular Beyond-the-Standard-Model Inference Tool". In: *Eur. Phys. J. C* 77.11 (2017). [Addendum: Eur.Phys.J.C 78, 98 (2018)], p. 784. DOI: 10.1140/epjc/s10052-017-5321-8. arXiv: 1705.07908 [hep-ph].

[96] O. Buchmueller et al. "Prediction for the Lightest Higgs Boson Mass in the CMSSM using Indirect Experimental Constraints". In: *Phys. Lett. B* 657 (2007), pp. 87–94. DOI: 10.1016/j.physletb.2007.09.058. arXiv: 0707.3447 [hep-ph].

[97] Giulio D'Agostini. "Probability and measurement uncertainty in physics: A Bayesian primer". In: (). arXiv: hep-ph/9512295 [hep-ph].

[98] G D'Agostini. "Bayesian inference in processing experimental data: principles and basic applications". In: *Rep. Prog. Phys.* 66.9 (2003), p. 1383. URL: http://stacks.iop.org/0034-4885/66/i=9/a=201.

[99] G. Cowan. *Statistical data analysis*. Oxford University Press, USA, 1998.

[100] Mark de Berg et al. "Computational Geometry". In: *Computational Geometry: Algorithms and Applications*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2000, pp. 1–17. ISBN: 978-3-662-04245-8. DOI: 10.1007/978-3-662-04245-8_1. URL: https://doi.org/10.1007/978-3-662-04245-8_1.

[101] Radford M. Neal. "MCMC Using Hamiltonian Dynamics". In: *Handbook of Markov Chain Monte Carlo* 54 (2010), pp. 113–162.

[102] A. D. Martin et al. "Uncertainties of predictions from parton distributions. 1: Experimental errors". In: *Eur. Phys. J. C* 28 (2003), pp. 455–473. DOI: 10.1140/epjc/s2003-01196-2. arXiv: hep-ph/0211080 [hep-ph].

[103] A. D. Martin et al. "Parton distributions for the LHC". In: *Eur. Phys. J. C* 63 (2009), pp. 189–285. DOI: 10.1140/epjc/s10052-009-1072-5. arXiv: 0901.0002 [hep-ph].

[104] L. A. Harland-Lang et al. "Parton distributions in the LHC era: MMHT 2014 PDFs". In: *Eur. Phys. J. C* 75.5 (2015), p. 204. DOI: 10.1140/epjc/s10052-015-3397-6. arXiv: 1412.3989 [hep-ph].

[105] Daniel de Florian et al. "Parton-to-Pion Fragmentation Reloaded". In: *Phys. Rev. D* 91.1 (2015), p. 014035. DOI: 10.1103/PhysRevD.91.014035. arXiv: 1410.6027 [hep-ph].

[106] D. de Florian et al. "Parton-to-Kaon Fragmentation Revisited". In: *Phys. Rev. D* 95.9 (2017), p. 094019. DOI: 10.1103/PhysRevD.95.094019. arXiv: 1702.06353 [hep-ph].

[107] A. Accardi et al. "Deuterium scattering experiments in CTEQ global QCD analyses: a comparative investigation". In: *Eur. Phys. J. C* 81.7 (2021), p. 603. DOI: 10.1140/epjc/s10052-021-09318-y. arXiv: 2102.01107 [hep-ph].

[108] Nobuo Sato et al. "First Monte Carlo analysis of fragmentation functions from single-inclusive e^+e^- annihilation". In: *Phys. Rev. D* 94.11 (2016), p. 114004. DOI: 10.1103/PhysRevD.94.114004. arXiv: 1609.00899 [hep-ph].

[109] J. J. Ethier, N. Sato, and W. Melnitchouk. "First simultaneous extraction of spin-dependent parton distributions and fragmentation functions from a global QCD analysis". In: *Phys. Rev. Lett.* 119.13 (2017), p. 132001. DOI: 10.1103/PhysRevLett.119.132001. arXiv: 1705.05889 [hep-ph].

[110] Richard D. Ball et al. "The PDF4LHC21 combination of global PDF fits for the LHC Run III". In: (). arXiv: 2203.05506 [hep-ph].

[111] Jon Butterworth et al. "PDF4LHC recommendations for LHC Run II". In: *J. Phys. G* 43 (2016), p. 023001. DOI: 10.1088/0954-3899/43/2/023001. arXiv: 1510.03865 [hep-ph].

[112] A. Accardi et al. "A Critical Appraisal and Evaluation of Modern PDFs". In: *Eur. Phys. J. C* 76.8 (2016), p. 471. DOI: 10.1140/epjc/s10052-016-4285-4. arXiv: 1603.08906 [hep-ph].

[113] Richard D. Ball et al. "Parton distributions for the LHC Run II". In: *JHEP* 04 (2015), p. 040. DOI: 10.1007/JHEP04(2015)040. arXiv: 1410.8849 [hep-ph].

[114] N. T. Hunt-Smith et al. "Determination of uncertainties in parton densities". In: *Phys. Rev. D* 106.3 (2022), p. 036003. DOI: 10.1103/PhysRevD.106.036003. arXiv: 2206.10782 [hep-ph].

[115] Matthew D. Hoffman and Andrew Gelman. arXiv: 1111.4246 [stat.CO].

[116] F. Feroz, M. P. Hobson, and M. Bridges. "MultiNest: an efficient and robust Bayesian inference tool for cosmology and particle physics". In: *Mon. Not. Roy. Astron. Soc.* 398 (2009), pp. 1601–1614. DOI: 10.1111/j.1365-2966.2009.14548.x. arXiv: 0809.3437 [astro-ph].

[117] Richard D. Ball et al. "The path to proton structure at 1% accuracy". In: *Eur. Phys. J. C* 82.5 (2022), p. 428. DOI: 10.1140/epjc/s10052-022-10328-7. arXiv: 2109.02653 [hep-ph].

[118] N. Sato et al. "Strange quark suppression from a simultaneous Monte Carlo analysis of parton distributions and fragmentation functions". In: *Phys. Rev. D* 101.7 (2020), p. 074020. DOI: 10.1103/PhysRevD.101.074020.

[119] Eric Moffat et al. "Simultaneous Monte Carlo analysis of parton densities and fragmentation functions". In: *Phys. Rev. D* 104.1 (2021), p. 016015. DOI: 10.1103/PhysRevD.104.016015.

[120] C. Cocuzza et al. "Polarized antimatter in the proton from a global QCD analysis". In: *Phys. Rev. D* 106.3 (2022), p. L031502. DOI: 10.1103/PhysRevD.106.L031502.

[121] L. Adamczyk et al. "Precision Measurement of the Longitudinal Double-spin Asymmetry for Inclusive Jet Production in Polarized Proton Collisions at $\sqrt{s} =$ 200 GeV". In: *Phys. Rev. Lett.* 115.9 (2015), p. 092002. DOI: 10.1103/PhysRevLett.115.092002.

[122] J. Adam et al. "Longitudinal double-spin asymmetry for inclusive jet and dijet production in pp collisions at $\sqrt{s} = 510$ GeV". In: *Phys. Rev. D* 100.5 (2019), p. 052005. DOI: 10.1103/PhysRevD.100.052005.

[123] M. S. Abdallah et al. "Longitudinal double-spin asymmetry for inclusive jet and dijet production in polarized proton collisions at $\sqrt{s} = 200$ GeV". In: *Phys. Rev. D* 103.9 (2021), p. L091103. DOI: 10.1103/PhysRevD.103.L091103.

[124] M. S. Abdallah et al. "Longitudinal double-spin asymmetry for inclusive jet and dijet production in polarized proton collisions at $\sqrt{s} = 510$ GeV". In: *Phys. Rev. D* 105.9 (2022), p. 092011. DOI: 10.1103/PhysRevD.105.092011.

[125] A. Adare et al. "Event Structure and Double Helicity Asymmetry in Jet Production from Polarized $p + p$ Collisions at $\sqrt{s} = 200$˜GeV". In: *Phys. Rev. D* 84 (2011), p. 012006. DOI: 10.1103/PhysRevD.84.012006.

[126] Y. Zhou, N. Sato, and W. Melnitchouk. "How well do we know the gluon polarization in the proton?" In: *Phys. Rev. D* 105.7 (2022), p. 074022. DOI: 10.1103/PhysRevD.105.074022.

[127] J. Karpie et al. "Gluon helicity from global analysis of experimental data and lattice QCD Ioffe time distributions". In: *Phys. Rev. D* 109 (2024), p. 036031. DOI: 10.1103/PhysRevD.109.036031.

[128] N. T. Hunt-Smith et al. in preparation. 2024.

[129] N. T. Hunt-Smith et al. "On the resolution of the sign of gluon polarization in the proton". In: (). arXiv: 2403.08117 [hep-ph].

[130] S. Wandzura and Frank Wilczek. "Sum Rules for Spin Dependent Electroproduction: Test of Relativistic Constituent Quarks". In: *Phys. Lett. B* 72 (1977), pp. 195–198. DOI: 10.1016/0370-2693(77)90700-6.

[131] P. Jimenez-Delgado, A. Accardi, and W. Melnitchouk. "Impact of hadronic and nuclear corrections on global analysis of spin-dependent parton distributions". In: *Phys. Rev. D* 89.3 (2014), p. 034025. DOI: 10.1103/PhysRevD.89.034025.

[132] Ingo Schienbein et al. "Target mass corrections". In: *J. Phys. G.* 35.5 (Mar. 2008), p. 053101. DOI: 10.1088/0954-3899/35/5/053101. URL: https://dx.doi.org/10.1088/0954-3899/35/5/053101.

[133] Otto Nachtmann. "Positivity constraints for anomalous dimensions". In: *Nucl. Phys.* B63 (1973), pp. 237–247. DOI: 10.1016/0550-3213(73)90144-2.

[134] O. W. Greenberg and D. Bhaumik. "Parton model with variable intermediate-state parton mass". In: *Phys. Rev. D* 4 (1971), pp. 2048–2055. DOI: 10.1103/PhysRevD.4.2048.

[135] Satoshi Matsuda and Tsuneo Uematsu. "Target Mass Effects in Polarized Electroproduction". In: *Nucl. Phys.* B168 (1980), pp. 181–188. DOI: 10.1016/0550-3213(80)90283-7.

[136] Andrea Piccione and Giovanni Ridolfi. "Target mass effects in polarized deep inelastic scattering". In: *Nucl. Phys.* B513 (1998), pp. 301–316. DOI: 10.1016/S0550-3213(97)00716-5.

[137] Johannes Blumlein and Avtandil Tkabladze. "Target mass corrections for polarized structure functions and new sum rules". In: *Nucl. Phys.* B553 (1999), pp. 427–464. DOI: 10.1016/S0550-3213(99)00289-8.

[138] A. Accardi and W. Melnitchouk. "Target mass corrections for spin-dependent structure functions in collinear factorization". In: *Phys. Lett. B* 670 (2008), pp. 114–118. DOI: 10.1016/j.physletb.2008.10.036.

[139] Howard Georgi and H. David Politzer. "Freedom at moderate energies: Masses in color dynamics". In: *Phys. Rev. D* 14 (Oct. 1976), pp. 1829–1848. DOI: 10.1103/PhysRevD.14.1829. URL: https://link.aps.org/doi/10.1103/PhysRevD.14.1829.

[140] Alvaro De Rujula, Howard Georgi, and H. David Politzer. "Demythification of Electroproduction, Local Duality and Precocious Scaling". In: *Annals Phys.* 103 (1977), p. 315. DOI: 10.1016/S0003-4916(97)90003-8.

[141] A. De Rujula, Howard Georgi, and H. David Politzer. "Trouble with xi Scaling?" In: *Phys. Rev. D* 15 (1977), p. 2495. DOI: 10.1103/PhysRevD.15.2495.

[142] Khalil Bitar, Porter W. Johnson, and W.-K. Tung. "QCD Asymptotics and Kinematic Thresholds in Deep Inelastic Scattering". In: *Phys. Lett. B* 83 (1979), pp. 114–118. DOI: 10.1016/0370-2693(79)90902-X.

[143] F. M. Steffens and W. Melnitchouk. "Target mass corrections revisited". In: *Phys. Rev. C* 73 (2006), p. 055202. DOI: 10.1103/PhysRevC.73.055202.

[144] F. M. Steffens et al. "Parton distributions in the presence of target mass corrections". In: *Phys. Rev. C* 86 (2012), p. 065208. DOI: 10.1103/PhysRevC.86.065208.

[145] J. F. Owens, A. Accardi, and W. Melnitchouk. "Global parton distributions with nuclear and finite-Q^2 corrections". In: *Phys. Rev. D* 87.9 (2013), p. 094012. DOI: 10.1103/PhysRevD.87.094012.

[146] A. Accardi et al. "Constraints on large-x parton distributions from new weak boson production and deep-inelastic scattering data". In: *Phys. Rev. D* 93.11 (2016), p. 114017. DOI: 10.1103/PhysRevD.93.114017.

[147] W. Melnitchouk, R. Ent, and C. Keppel. "Quark-hadron duality in electron scattering". In: *Phys. Rept.* 406 (2005), pp. 127–301. DOI: 10.1016/j.physrep.2004.10.004. arXiv: hep-ph/0501217.

[148] R. L. Jaffe and Xiang-Dong Ji. "Studies of the Transverse Spin Dependent Structure Function $g(2)$ (X, Q^2)". In: *Phys. Rev. D* 43 (1991), pp. 724–732. DOI: 10.1103/PhysRevD.43.724.

[149] M. Anselmino, A. Efremov, and E. Leader. "The Theory and phenomenology of polarized deep inelastic scattering". In: *Phys. Rep.* 261 (1995). [Erratum: Phys. Rep. **281**, 399 (1997)], pp. 1–124. DOI: 10.1016/0370-1573(95)00011-5.

[150] Hugh Burkhardt and W. N. Cottingham. "Sum rules for forward virtual Compton scattering". In: *Annals Phys.* 56 (1970), pp. 453–463. DOI: 10.1016/0003-4916(70)90025-4.

[151] Xiang-Dong Ji and Peter Unrau. "Q**2 dependence of the proton's G1 structure function sum rule". In: *Phys. Lett. B* 333 (1994), p. 228. DOI: 10.1016/0370-2693(94)91035-9.

[152] E. Stein et al. "QCD sum rule calculation of twist - three contributions to polarized nucleon structure functions". In: *Phys. Lett. B* 343 (1995), p. 369. DOI: 10.1016/0370-2693(94)01443-G.

[153] E. Stein et al. "QCD sum rule calculation of twist four corrections to Bjorken and Ellis-Jaffe sum rules". In: *Phys. Lett. B* 353 (1995), pp. 107–113. DOI: 10.1016/0370-2693(95)00544-U.

[154] Matthias Burkardt. "Transverse force on quarks in deep-inelastic scattering". In: *Phys. Rev. D* 88 (2013), p. 114502. DOI: 10.1103/PhysRevD.88.114502.

[155] Alessandro Bacchetta et al. "Semi-inclusive deep inelastic scattering at small transverse momentum". In: *JHEP* 02 (2007), p. 093. DOI: 10.1088/1126-6708/2007/02/093. arXiv: hep-ph/0611265.

[156] Marco Stratmann and Werner Vogelsang. "Towards a global analysis of polarized parton distributions". In: *Phys. Rev. D* 64 (2001), p. 114007. DOI: 10.1103/PhysRevD.64.114007.

[157] A. Accardi, T. Hobbs, and W. Melnitchouk. "Hadron mass corrections in semi-inclusive deep inelastic scattering". In: *JHEP* 11 (2009), p. 084. DOI: 10.1088/1126-6708/2009/11/084.

[158] J. V. Guerrero et al. "Hadron mass corrections in semi-inclusive deep-inelastic scattering". In: *JHEP* 09 (2015), p. 169. DOI: 10.1007/JHEP09(2015)169.

[159] Juan V. Guerrero and Alberto Accardi. "Gauge invariance and kaon production in deep inelastic scattering at low scales". In: *Phys. Rev. D* 97.11 (2018), p. 114012. DOI: 10.1103/PhysRevD.97.114012.

[160] T. Anderson et al. in preparation. 2024.

[161] D. Adamiak et al. in preparation. 2024.

[162] Felix Ringer and Werner Vogelsang. "Single-Spin Asymmetries in W Boson Production at Next-to-Leading Order". In: *Phys. Rev. D* 91 (2015), p. 094033. DOI: 10.1103/PhysRevD.91.094033.

[163] B. Jager, M. Stratmann, and W. Vogelsang. "Single inclusive jet production in polarized pp collisions at $O(alpha_s^3)$". In: *Phys. Rev. D* 70 (2004), p. 034010. DOI: 10.1103/PhysRevD.70.034010.

[164] A. V. Radyushkin. "Quasi-parton distribution functions, momentum distributions, and pseudo-parton distribution functions". In: *Phys. Rev. D* 96.3 (2017), p. 034025. DOI: 10.1103/PhysRevD.96.034025.

[165] Ian Balitsky, Wayne Morris, and Anatoly Radyushkin. "Polarized gluon pseudodistributions at short distances". In: *JHEP* 02 (2022), p. 193. DOI: 10.1007/JHEP02(2022)193.

[166] Szabolcs Borsányi et al. "High-precision scale setting in lattice QCD". In: *JHEP* 09 (2012), p. 010. DOI: 10.1007/JHEP09(2012)010.

[167] Colin Egerer et al. "Toward the determination of the gluon helicity distribution in the nucleon from lattice quantum chromodynamics". In: *Phys. Rev. D* 106.9 (2022), p. 094511. DOI: 10.1103/PhysRevD.106.094511.

[168] J. Bringewatt et al. "Confronting lattice parton distributions with global QCD analysis". In: *Phys. Rev. D* 103.1 (2021), p. 016003. DOI: 10.1103/PhysRevD.103.016003.

[169] P. C. Barry et al. "Complementarity of experimental and lattice QCD data on pion parton distributions". In: *Phys. Rev. D* 105.11 (2022), p. 114051. DOI: 10.1103/PhysRevD.105.114051.

[170] Leonard Gamberg et al. "Updated QCD global analysis of single transverse-spin asymmetries: Extracting H˜, and the role of the Soffer bound and lattice QCD". In: *Phys. Rev. D* 106.3 (2022), p. 034014. DOI: 10.1103/PhysRevD.106.034014.

[171] P. C. Barry et al. "Tomography of pions and protons via transverse momentum dependent distributions". In: *Phys. Rev. D* 108.9 (2023), p. L091504. DOI: 10.1103/PhysRevD.108.L091504.

[172] C. Cocuzza et al. "Transversity Distributions and Tensor Charges of the Nucleon: Extraction from Dihadron Production and Their Universal Nature". In: *Phys. Rev. Lett.* 132.9 (2024), p. 091901. DOI: 10.1103/PhysRevLett.132.091901.

[173] C. Cocuzza et al. "First simultaneous global QCD analysis of dihadron fragmentation functions and transversity parton distribution functions". In: *Phys. Rev. D* 109.3 (2024), p. 034024. DOI: 10.1103/PhysRevD.109.034024.

[174] R. L. Workman et al. "Review of Particle Physics". In: *PTEP* 2022 (2022), p. 083C01. DOI: 10.1093/ptep/ptac097.

[175] J. Ashman et al. "An Investigation of the Spin Structure of the Proton in Deep Inelastic Scattering of Polarized Muons on Polarized Protons". In: *Nucl. Phys.* B328 (1989). Ed. by V. W. Hughes and C. Cavata, p. 1. DOI: 10.1016/0550-3213(89)90089-8.

[176] B. Adeva et al. "Spin asymmetries A(1) and structure functions g1 of the proton and the deuteron from polarized high-energy muon scattering". In: *Phys. Rev. D* 58 (1998), p. 112001. DOI: 10.1103/PhysRevD.58.112001.

[177] B. Adeva et al. "Spin asymmetries A(1) of the proton and the deuteron in the low x and low Q**2 region from polarized high-energy muon scattering". In: *Phys. Rev. D* 60 (1999). [Erratum: Phys. Rev. D **62**, 079902 (2000)], p. 072004. DOI: 10.1103/PhysRevD.60.072004.

[178] V. Yu. Alexakhin et al. "The Deuteron Spin-dependent Structure Function g1(d) and its First Moment". In: *Phys. Lett. B* 647 (2007), pp. 8–17. DOI: 10.1016/j.physletb.2006.12.076.

[179] M.G. Alekseev et al. "The spin-dependent structure function of the proton g1p and a test of the Bjorken sum rule". In: *Phys. Lett. B* 690.5 (2010), pp. 466–472. ISSN: 0370-2693. DOI: https://doi.org/10.1016/j.physletb.2010.05.069. URL: https://www.sciencedirect.com/science/article/pii/S0370269310006726.

[180] C. Adolph et al. "Longitudinal double spin asymmetries in single hadron quasi-real photoproduction at high pT". In: *Phys. Lett. B* 753 (2016), pp. 573–579. ISSN: 0370-2693. DOI: https://doi.org/10.1016/j.physletb.2015.12.035. URL: https://www.sciencedirect.com/science/article/pii/S0370269315009831.

[181] G. Baum et al. "New Measurement of Deep-Inelastic $e-p$ Asymmetries". In: *Phys. Rev. Lett.* 51 (13 Sept. 1983), pp. 1135–1138. DOI: 10.1103/PhysRevLett.51.1135. URL: https://link.aps.org/doi/10.1103/PhysRevLett.51.1135.

[182] P. L. Anthony et al. "Deep inelastic scattering of polarized electrons by polarized ^{3}He and the study of the neutron spin structure". In: *Phys. Rev. D* 54 (11 Dec. 1996), pp. 6620–6650. DOI: 10.1103/PhysRevD.54.6620. URL: https://link.aps.org/doi/10.1103/PhysRevD.54.6620.

[183] K. Abe et al. "Precision Determination of the Neutron Spin Structure Function g_1^n". In: *Phys. Rev. Lett.* 79 (1 July 1997), pp. 26–30. DOI: 10.1103/PhysRevLett.79.26. URL: https://link.aps.org/doi/10.1103/PhysRevLett.79.26.

[184] K. Abe et al. "Measurements of the proton and deuteron spin structure functions g_1 and g_2". In: *Phys. Rev. D* 58 (11 Oct. 1998), p. 112003. DOI: 10.1103/PhysRevD.58.112003. URL: https://link.aps.org/doi/10.1103/PhysRevD.58.112003.

[185] P.L. Anthony et al. "Measurement of the deuteron spin structure function g1d(x) for 1 (GeV/c)2¡Q2¡40 (GeV/c)2". In: *Phys. Lett. B* 463.2 (1999), pp. 339–345. ISSN: 0370-2693. DOI: https://doi.org/10.1016/S0370-2693(99)00940-5. URL: https://www.sciencedirect.com/science/article/pii/S0370269399009405.

[186] P.L Anthony et al. "Measurements of the Q2-dependence of the proton and neutron spin structure functions g1p and g1n". In: *Phys. Lett. B* 493.1 (2000), pp. 19–28. ISSN: 0370-2693. DOI: https://doi.org/10.1016/S0370-2693(00)01014-5. URL: https://www.sciencedirect.com/science/article/pii/S0370269300010145.

[187] K. Ackerstaff et al. "Measurement of the neutron spin structure function gn1 with a polarized 3He internal target". In: *Phys. Lett. B* 404.3 (1997), pp. 383–389. ISSN: 0370-2693. DOI: https://doi.org/10.1016/S0370-2693(97)00611-4. URL: https://www.sciencedirect.com/science/article/pii/S0370269397006114.

[188] A. Airapetian et al. "Precise determination of the spin structure function g_1 of the proton, deuteron, and neutron". In: *Phys. Rev. D* 75 (1 Jan. 2007), p. 012007. DOI: 10.1103/PhysRevD.75.012007. URL: https://link.aps.org/doi/10.1103/PhysRevD.75.012007.

[189] X. Zheng et al. "Precision measurement of the neutron spin asymmetries and spin-dependent structure functions in the valence quark region". In: *Phys. Rev. C* 70 (2004), p. 065207. DOI: 10.1103/PhysRevC.70.065207.

[190] D. S. Parno et al. "Precision Measurements of A_1^n in the Deep Inelastic Regime". In: *Phys. Lett. B* 744 (2015), pp. 309–314. DOI: 10.1016/j.physletb.2015.03.067.

[191] Y. Prok et al. "Precision measurements of g_1 of the proton and of the deuteron with 6 GeV electrons". In: *Phys. Rev. C* 90 (2 Aug. 2014), p. 025212. DOI: 10.1103/PhysRevC.90.025212. URL: https://link.aps.org/doi/10.1103/PhysRevC.90.025212.

[192] R. G. Fersch et al. "Determination of the proton spin structure functions for $0.05 < Q^2 < 5\text{GeV}^2$ using CLAS". In: *Phys. Rev. C* 96 (6 Dec. 2017), p. 065208. DOI: 10.1103/PhysRevC.96.065208. URL: https://link.aps.org/doi/10.1103/PhysRevC.96.065208.

[193] N. Guler et al. "Precise determination of the deuteron spin structure at low to moderate Q^2 with CLAS and extraction of the neutron contribution". In: *Phys. Rev. C* 92 (5 Nov. 2015), p. 055201. DOI: 10.1103/PhysRevC.92.055201. URL: https://link.aps.org/doi/10.1103/PhysRevC.92.055201.

[194] P. L. Anthony et al. "Measurement of the proton and deuteron spin structure functions g(2) and asymmetry A(2)". In: *Phys. Lett. B* 458 (1999), pp. 529–535. DOI: 10.1016/S0370-2693(99)00590-0.

[195] P. L. Anthony et al. "Precision measurement of the proton and deuteron spin structure functions g(2) and asymmetries A(2)". In: *Phys. Lett. B* 553 (2003), pp. 18–24. DOI: 10.1016/S0370-2693(02)03015-0.

[196] A. Airapetian et al. "Measurement of the virtual-photon asymmetry A_2 and the spin-structure function g_2 of the proton". In: *Eur. Phys. J. C* 72 (2012), p. 1921. DOI: 10.1140/epjc/s10052-012-1921-5.

[197] M. Posik et al. "A Precision Measurement of the Neutron Twist-3 Matrix Element d_2^n: Probing Color Forces". In: *Phys. Rev. Lett.* 113.2 (2014), p. 022002. DOI: 10.1103/PhysRevLett.113.022002.

[198] A. Airapetian et al. "Quark helicity distributions in the nucleon for up, down, and strange quarks from semi-inclusive deep-inelastic scattering". In: *Phys. Rev. D* 71 (2005), p. 012003. DOI: 10.1103/PhysRevD.71.012003.

[199] M. Alekseev et al. "Flavour Separation of Helicity Distributions from Deep Inelastic Muon-Deuteron Scattering". In: *Phys. Lett. B* 680 (2009), pp. 217–224. DOI: 10.1016/j.physletb.2009.08.065.

[200] Roland Windmolders. "Quark helicity distributions from longitudinal spin asymmetries in muon-proton and muon-deuteron scattering". In: *J. Phys. Conf. Ser.* 295 (2011). Ed. by Hans Stroher and Frank Rathmann, p. 012069. DOI: 10.1088/1742-6596/295/1/012069.

[201] B. I. Abelev et al. "Longitudinal double-spin asymmetry and cross section for inclusive jet production in polarized proton collisions at s**(1/2) = 200-GeV". In: *Phys. Rev. Lett.* 97 (2006), p. 252001. DOI: 10.1103/PhysRevLett.97.252001.

[202] B. I. Abelev et al. "Longitudinal double-spin asymmetry for inclusive jet production in p+p collisions at s**(1/2) = 200-GeV". In: *Phys. Rev. Lett.* 100 (2008), p. 232003. DOI: 10.1103/PhysRevLett.100.232003.

[203] L. Adamczyk et al. "Longitudinal and transverse spin asymmetries for inclusive jet production at mid-rapidity in polarized $p+p$ collisions at $\sqrt{s} = 200$ GeV". In: *Phys. Rev. D* 86 (2012), p. 032006. DOI: 10.1103/PhysRevD.86.032006.

[204] Jaroslav Adam et al. "Measurement of the longitudinal spin asymmetries for weak boson production in proton-proton collisions at $\sqrt{s} = 510$ GeV". In: *Phys. Rev. D* 99 (2019), p. 051102. DOI: 10.1103/PhysRevD.99.051102.

[205] A. Adare et al. "Measurement of parity-violating spin asymmetries in $W^{\pm}$ production at midrapidity in longitudinally polarized $p+p$ collisions". In: *Phys. Rev. D* 93 (2016), p. 051103. DOI: 10.1103/PhysRevD.93.051103.

[206] A. Adare et al. "Cross section and longitudinal single-spin asymmetry A_L for forward $W^{\pm} \to \mu^{\pm}\nu$ production in polarized $p+p$ collisions at $\sqrt{s} = 510$ GeV". In: *Phys. Rev. D* 98 (2018), p. 032007. DOI: 10.1103/PhysRevD.98.032007.

[207] Daniel de Florian, Stefano Forte, and Werner Vogelsang. "Higgs production at RHIC and the positivity of the gluon helicity distribution". In: *Phys. Rev. D* 109.7 (2024), p. 074007. DOI: 10.1103/PhysRevD.109.074007. arXiv: 2401.10814 [hep-ph].

[208] Yi-Bo Yang et al. "Glue Spin and Helicity in the Proton from Lattice QCD". In: *Phys. Rev. Lett.* 118.10 (2017), p. 102001. DOI: 10.1103/PhysRevLett.118.102001.

[209] R. M. Whitehill et al. "Accessing gluon polarization with high-PT hadrons in SIDIS". In: *Phys. Rev. D* 107.3 (2023), p. 034033. DOI: 10.1103/PhysRevD.107.034033.

[210] Y. Zhou et al. "Revisiting quark and gluon polarization in the proton at the EIC". In: *Phys. Rev. D* 104.3 (2021), p. 034028. DOI: 10.1103/PhysRevD.104.034028.

[211] Csaba Balázs et al. "A comparison of optimisation algorithms for high-dimensional particle and astrophysics applications". In: *JHEP* 05 (2021), p. 108. DOI: 10.1007/JHEP05(2021)108. arXiv: 2101.04525 [hep-ph].

[212] Anders Kvellestad, Pat Scott, and Martin White. "GAMBIT and its Application in the Search for Physics Beyond the Standard Model". In: (Dec. 2019). DOI: 10.1016/j.ppnp.2020.103769. arXiv: 1912.04079 [hep-ph].

[213] Jascha Sohl-Dickstein et al. "Deep Unsupervised Learning using Nonequilibrium Thermodynamics". In: *CoRR* abs/1503.03585 (2015). arXiv: 1503.03585. URL: http://arxiv.org/abs/1503.03585.

[214] Yang Song and Stefano Ermon. *Generative Modeling by Estimating Gradients of the Data Distribution*. 2019. DOI: 10.48550/ARXIV.1907.05600. URL: https://arxiv.org/abs/1907.05600.

[215] Jonathan Ho, Ajay Jain, and Pieter Abbeel. *Denoising Diffusion Probabilistic Models*. 2020. DOI: 10.48550/ARXIV.2006.11239. URL: https://arxiv.org/abs/2006.11239.

[216] Yang Song and Stefano Ermon. *Improved Techniques for Training Score-Based Generative Models*. 2020. DOI: 10.48550/ARXIV.2006.09011. URL: https://arxiv.org/abs/2006.09011.

[217] Alex Nichol and Prafulla Dhariwal. "Improved Denoising Diffusion Probabilistic Models". In: *CoRR* abs/2102.09672 (2021). arXiv: 2102.09672. URL: https://arxiv.org/abs/2102.09672.

[218] Prafulla Dhariwal and Alex Nichol. *Diffusion Models Beat GANs on Image Synthesis*. 2021. DOI: 10.48550/ARXIV.2105.05233. URL: https://arxiv.org/abs/2105.05233.

[219] Alex Nichol et al. *GLIDE: Towards Photorealistic Image Generation and Editing with Text-Guided Diffusion Models*. 2021. DOI: 10.48550/ARXIV.2112.10741. URL: https://arxiv.org/abs/2112.10741.

[220] Aditya Ramesh et al. *Hierarchical Text-Conditional Image Generation with CLIP Latents*. 2022. DOI: 10.48550/ARXIV.2204.06125. URL: https://arxiv.org/abs/2204.06125.

[221] Andrea Montanari and Yuchen Wu. *Posterior Sampling from the Spiked Models via Diffusion Processes*. 2023. arXiv: 2304.11449 [math.ST].

[222] M. S. Albergo, G. Kanwar, and P. E. Shanahan. "Flow-based generative models for Markov chain Monte Carlo in lattice field theory". In: *Phys. Rev. D* 100.3 (2019), p. 034515. DOI: 10.1103/PhysRevD.100.034515. arXiv: 1904.12072 [hep-lat].

[223] Denis Boyda et al. "Sampling using $SU(N)$ gauge equivariant flows". In: *Phys. Rev. D* 103.7 (2021), p. 074504. DOI: 10.1103/PhysRevD.103.074504. arXiv: 2008.05456 [hep-lat].

[224] Michael S. Albergo et al. "Introduction to Normalizing Flows for Lattice Field Theory". In: (Jan. 2021). arXiv: 2101.08176 [hep-lat].

[225] Marylou Gabrié, Grant M. Rotskoff, and Eric Vanden-Eijnden. *Efficient Bayesian Sampling Using Normalizing Flows to Assist Markov Chain Monte Carlo Methods*. 2021. DOI: 10.48550/ARXIV.2107.08001. URL: https://arxiv.org/abs/2107.08001.

[226] Marylou Gabrié, Grant M. Rotskoff, and Eric Vanden-Eijnden. "Adaptive Monte Carlo augmented with normalizing flows". In: *Proc. Nat. Acad. Sci.* 119.10 (2022), e2109420119. DOI: 10.1073/pnas.2109420119. arXiv: 2105.12603 [physics.data-an].

[227] Qinsheng Zhang and Yongxin Chen. *Diffusion Normalizing Flow*. 2021. arXiv: 2110.07579 [cs.LG].

[228] Csaba Balázs et al. "Cosmological constraints on decaying axion-like particles: a global analysis". In: *JCAP* 12 (2022), p. 027. DOI: 10.1088/1475-7516/2022/12/027. arXiv: 2205.13549 [astro-ph.CO].

[229] Viktor Ananyev et al. "Collider constraints on electroweakinos in the presence of a light gravitino". In: *Eur. Phys. J. C* 83.6 (2023), p. 493. DOI: 10.1140/epjc/s10052-023-11574-z. arXiv: 2303.09082 [hep-ph].

[230] Christopher Chang et al. "Global fits of simplified models for dark matter with GAMBIT II. Vector dark matter with an s-channel vector mediator". In: (Mar. 2023). arXiv: 2303.08351 [hep-ph].

[231] Rob Cornish et al. *Relaxing Bijectivity Constraints with Continuously Indexed Normalising Flows*. 2021. arXiv: 1909.13833 [stat.ML].

[232] N. T. Hunt-Smith et al. "Accelerating Markov Chain Monte Carlo sampling with diffusion models". In: *Comput. Phys. Commun.* 296 (2024), p. 109059. DOI: 10.1016/j.cpc.2023.109059. arXiv: 2309.01454 [hep-ph].

[233] William Feller. "On the Theory of Stochastic Processes, with Particular Reference to Applications". In: 1949.

[234] Luke Tierney. "Markov Chains for Exploring Posterior Distributions". In: *The Annals of Statistics* 22.4 (1994), pp. 1701–1728. DOI: 10.1214/aos/1176325750. URL: https://doi.org/10.1214/aos/1176325750.

[235] D.J.C. MacKay. *Information Theory, Inference and Learning Algorithms*. Cambridge University Press, 2003. ISBN: 9780521642989. URL: https://books.google.com.au/books?id=AKuMj4PN%5C_EMC.

[236] Sergey Samsonov et al. "Local-Global MCMC kernels: the best of both worlds". In: *Advances in Neural Information Processing Systems*. Ed. by S. Koyejo et al. Vol. 35. Curran Associates, Inc., 2022, pp. 5178–5193. URL: https://proceedings.neurips.cc/paper%5C_files/paper/2022/file/21c86d5b10cdc28664ccdPaper-Conference.pdf.

[237] Daniel Foreman-Mackey et al. "emcee: The MCMC Hammer". In: *Publications of the Astronomical Society of the Pacific* 125.925 (2013), pp. 306–312. DOI: 10.1086/670067. URL: https://doi.org/10.1086%2F670067.

[238] C Andrieu and J Thoms. "A tutorial on adaptive MCMC". English. In: *Statistics and Computing* 18, issue 4 (Dec. 2008). Publisher: Springer Netherlands, pp. 343–373. ISSN: 0960-3174. DOI: 10.1007/s11222-008-9110-y.

[239] Pierre Fayet. "Effects of the spin 1 partner of the goldstino (gravitino) on neutral current phenomenology". In: *Phys. Lett. B* 95 (1980), p. 285. DOI: 10.1016/0370-2693(80)90488-8.

[240] Pierre Fayet. "On the search for a new spin 1 boson". In: *Nucl. Phys. B* 187 (1981), p. 184. DOI: 10.1016/0550-3213(81)90122-X.

[241] L. B. Okun. "Limits of electrodynamics: Paraphotons?" In: *Sov. Phys. JETP* 56 (1982), p. 502.

[242] Peter Galison and Aneesh Manohar. "TWO Z's OR NOT TWO Z's?" In: *Phys. Lett. B* 136 (1984), pp. 279–283. DOI: 10.1016/0370-2693(84)91161-4.

[243] Bob Holdom. "Two U(1)'s and epsilon charge shifts". In: *Phys. Lett. B* 166 (1986), p. 196. DOI: 10.1016/0370-2693(86)91377-8.

[244] Bob Holdom. "Searching for ϵ Charges and a New U(1)". In: *Phys. Lett. B* 178 (1986), pp. 65–70. DOI: 10.1016/0370-2693(86)90470-3.

[245] Pierre Fayet. "Extra U(1)'s and New Forces". In: *Nucl. Phys.* B347 (1990), pp. 743–768. DOI: 10.1016/0550-3213(90)90381-M.

[246] A. Leike. "The Phenomenology of extra neutral gauge bosons". In: *Phys. Rep.* 317 (1999), pp. 143–250. DOI: 10.1016/S0370-1573(98)00133-1. arXiv: hep-ph/9805494.

[247] Thomas Appelquist, Bogdan A. Dobrescu, and Adam R. Hopper. "Nonexotic Neutral Gauge Bosons". In: *Phys. Rev. D* 68 (2003), p. 035012. DOI: 10.1103/PhysRevD.68.035012. arXiv: hep-ph/0212073.

[248] Paul Langacker. "The Physics of Heavy Z' Gauge Bosons". In: *Rev. Mod. Phys.* 81 (2009), pp. 1199–1228. DOI: 10.1103/RevModPhys.81.1199. arXiv: 0801.1345 [hep-ph].

[249] Maxim Pospelov. "Secluded U(1) below the weak scale". In: *Phys. Rev. D* 80 (2009), p. 095002. DOI: 10.1103/PhysRevD.80.095002. arXiv: 0811.1030 [hep-ph].

[250] Hooman Davoudiasl, Hye-Sung Lee, and William J. Marciano. "Muon anomaly and dark parity violation". In: *Phys. Rev. Lett.* 109 (2012), p. 031802. DOI: 10.1103/PhysRevLett.109.031802. arXiv: 1205.2709 [hep-ph].

[251] Andrea Caputo et al. "Dark photon limits: A handbook". In: *Phys. Rev. D* 104.9 (2021), p. 095029. DOI: 10.1103/PhysRevD.104.095029. arXiv: 2105.04565 [hep-ph].

[252] Kai-Yu Zhang and Wan-Zhe Feng. "Explaining the W boson mass anomaly and dark matter with a U(1) dark sector*". In: *Chin. Phys. C* 47.2 (2023), p. 023107. DOI: 10.1088/1674-1137/aca585. arXiv: 2204.08067 [hep-ph].

[253] Yu-Pan Zeng et al. "Z boson mixing and the mass of the W boson". In: *Phys. Rev. D* 107.5 (2023), p. 056004. DOI: 10.1103/PhysRevD.107.056004. arXiv: 2204.09487 [hep-ph].

[254] Yu Cheng et al. "Dark photon kinetic mixing effects for the CDF W-mass measurement". In: *Phys. Rev. D* 106.5 (2022), p. 055011. DOI: 10.1103/PhysRevD.106.055011. arXiv: 2204.10156 [hep-ph].

[255] A. W. Thomas and X. G. Wang. "Constraints on the dark photon from parity violation and the W mass". In: *Phys. Rev. D* 106 (2022), p. 056017. DOI: 10.1103/PhysRevD.106.056017. arXiv: 2205.01911 [hep-ph].

[256] Hooman Davoudiasl, Hye-Sung Lee, and William J. Marciano. "'Dark' Z implications for parity violation, rare meson decays, and Higgs physics". In: *Phys. Rev. D* 85 (2012), p. 115019. DOI: 10.1103/PhysRevD.85.115019. arXiv: 1203.2947 [hep-ph].

[257] Alakabha Datta et al. "Dark photon and dark Z mediated B meson decays". In: *JHEP* 03 (2023), p. 108. DOI: 10.1007/JHEP03(2023)108. arXiv: 2210.15662 [hep-ph].

[258] X. G. Wang and A. W. Thomas. "Dark photon effect on the rare kaon decay $K_L \to \pi^0 \nu\bar{\nu}$". In: (). arXiv: 2301.08367 [hep-ph].

[259] S. Bilmis et al. "Constraints on Dark Photon from Neutrino-Electron Scattering Experiments". In: *Phys. Rev. D* 92.3 (2015), p. 033009. DOI: 10.1103/PhysRevD.92.033009. arXiv: 1502.07763 [hep-ph].

[260] Yu. M. Andreev et al. "Search for a New B-L Z' Gauge Boson with the NA64 Experiment at CERN". In: *Phys. Rev. Lett.* 129.16 (2022), p. 161801. DOI: 10.1103/PhysRevLett.129.161801. arXiv: 2207.09979 [hep-ex].

[261] J. P. Lees et al. "Search for a Dark Photon in e^+e^- Collisions at BaBar". In: *Phys. Rev. Lett.* 113.20 (2014), p. 201801. DOI: 10.1103/PhysRevLett.113.201801. arXiv: 1406.2980 [hep-ex].

[262] J. P. Lees et al. "Search for invisible decays of a dark photon produced in e^+e^- collisions at BaBar". In: *Phys. Rev. Lett.* 119 (2017), p. 131804. DOI: 10.1103/PhysRevLett.119.131804. arXiv: 1702.03327 [hep-ex].

[263] Roel Aaij et al. "Search for $A' \to \mu^+\mu^-$ decays". In: *Phys. Rev. Lett.* 124 (2020), p. 041801. DOI: 10.1103/PhysRevLett.124.041801. arXiv: 1910.06926 [hep-ex].

[264] Albert M Sirunyan et al. "Search for a narrow resonance lighter than 200 GeV decaying to a pair of muons in proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *Phys. Rev. Lett.* 124 (2020), p. 131802. DOI: 10.1103/PhysRevLett.124.131802. arXiv: 1912.04776 [hep-ex].

[265] Albert M Sirunyan et al. "Search for resonant and nonresonant new phenomena in high-mass dilepton final states at $\sqrt{s} = 13$ TeV". In: *JHEP* 07 (2021), p. 208. DOI: 10.1007/JHEP07(2021)208. arXiv: 2103.02708 [hep-ex].

[266] Matt Graham, Christopher Hearty, and Mike Williams. "Searches for dark photons at accelerators". In: *Ann. Rev. Nucl. Part. Sci.* 71 (2021), p. 37. DOI: 10.1146/annurev-nucl-110320-051823. arXiv: 2104.10280 [hep-ph].

[267] Anson Hook, Eder Izaguirre, and Jay G. Wacker. "Model independent bounds on kinetic mixing". In: *Adv. High Energy Phys.* 2011 (2011), p. 859762. DOI: 10.1155/2011/859762. arXiv: 1006.0973 [hep-ph].

[268] David Curtin et al. "Illuminating dark photons with high-energy colliders". In: *JHEP* 02 (2015), p. 157. DOI: 10.1007/JHEP02(2015)157. arXiv: 1412.0018 [hep-ph].

[269] B. M. Loizos et al. "Constraints on the dark sector from electroweak precision observables". In: (June 2023). arXiv: 2306.13408 [hep-ph].

[270] Graham D. Kribs, David McKeen, and Nirmal Raj. "Breaking up the Proton: An Affair with Dark Forces". In: *Phys. Rev. Lett.* 126.1 (2021), p. 011801. DOI: 10.1103/PhysRevLett.126.011801. arXiv: 2007.15655 [hep-ph].

[271] A. W. Thomas, X. G. Wang, and A. G. Williams. "Constraints on the dark photon from deep inelastic scattering". In: *Phys. Rev. D* 105.3 (2022), p. L031901. DOI: 10.1103/PhysRevD.105.L031901. arXiv: 2111.05664 [hep-ph].

[272] Bin Yan. "Probing the dark photon via polarized DIS scattering at HERA and EIC". In: *Phys. Lett. B* 833 (2022), p. 137384. DOI: 10.1016/j.physletb.2022.137384. arXiv: 2203.01510 [hep-ph].

[273] A. W. Thomas, X. G. Wang, and A. G. Williams. "Sensitivity of parity-violating electron scattering to a dark photon". In: *Phys. Rev. Lett.* 129 (2022), p. 011807. DOI: 10.1103/PhysRevLett.129.011807. arXiv: 2201.06760 [hep-ph].

[274] Matthew McCullough, James Moore, and Maria Ubiali. "The dark side of the proton". In: *JHEP* 08 (2022), p. 019. DOI: 10.1007/JHEP08(2022)019. arXiv: 2203.12628 [hep-ph].

[275] R. Abdul Khalek et al. "Science Requirements and Detector Concepts for the Electron-Ion Collider: EIC Yellow Report". In: *Nucl. Phys.* A1026 (2022), p. 122447. DOI: 10.1016/j.nuclphysa.2022.122447.

[276] S. Abrahamyan et al. "Search for a New Gauge Boson in Electron-Nucleus Fixed-Target Scattering by the APEX Experiment". In: *Phys. Rev. Lett.* 107 (2011), p. 191804. DOI: 10.1103/PhysRevLett.107.191804. arXiv: 1108.2750 [hep-ex].

[277] M. Battaglieri et al. "The Heavy Photon Search Test Detector". In: *Nucl. Instrum. Meth. A* 777 (2015), pp. 91–101. DOI: 10.1016/j.nima.2014.12.017. arXiv: 1406.6115 [physics.ins-det].

[278] T. Beranek, H. Merkel, and M. Vanderhaeghen. "Theoretical framework to analyze searches for hidden light gauge bosons in electron scattering fixed target experiments". In: *Phys. Rev. D* 88 (2013), p. 015032. DOI: 10.1103/PhysRevD.88.015032. arXiv: 1303.2540 [hep-ph].

[279] N. T. Hunt-Smith et al. "Global QCD analysis and dark photons". In: *JHEP* 09 (2023), p. 096. DOI: 10.1007/JHEP09(2023)096. arXiv: 2302.11126 [hep-ph].

[280] X. G. Wang et al. "Constraints on the $U(1)_{B-L}$ model from global QCD analysis". In: *Phys. Rev. D* 111 (1 Jan. 2025), p. 015019. DOI: 10.1103/PhysRevD.111.015019. URL: https://link.aps.org/doi/10.1103/PhysRevD.111.015019.

[281] Stefano Carrazza et al. "Can New Physics hide inside the proton?" In: *Phys. Rev. Lett.* 123 (2019), p. 132001. DOI: 10.1103/PhysRevLett.123.132001. arXiv: 1905.05215 [hep-ph].

[282] Zahari Kassabov et al. "The top quark legacy of the LHC Run II for PDF and SMEFT analyses". In: *JHEP* 05 (2023), p. 205. DOI: 10.1007/JHEP05(2023)205. arXiv: 2303.06159 [hep-ph].

[283] Jun Gao et al. "Simultaneous CTEQ-TEA extraction of PDFs & SMEFT parameters from jet and $t\bar{t}$ data". In: *JHEP* 05 (2023), p. 003. DOI: 10.1007/JHEP05(2023)003. arXiv: 2211.01094 [hep-ph].

[284] Shayan Iranipour and Maria Ubiali. "A new generation of simultaneous fits to LHC data using deep learning". In: *JHEP* 05 (2022), p. 032. DOI: 10.1007/JHEP05(2022)032. arXiv: 2201.07240 [hep-ph].

[285] Admir Greljo et al. "Parton distributions in the SMEFT from high-energy Drell-Yan tails". In: *JHEP* 07 (2021), p. 122. DOI: 10.1007/JHEP07(2021)122. arXiv: 2104.02723 [hep-ph].

[286] Rabah Abdul Khalek et al. "Parton distributions with theory uncertainties: General formalism and first phenomenological studies". In: *Eur. Phys. J. C* 79 (2019), p. 931. DOI: 10.1140/epjc/s10052-019-7401-4. arXiv: 1906.10698 [hep-ph].

[287] John Collins, Ted C. Rogers, and Nobuo Sato. "Positivity and renormalization of parton densities". In: *Phys. Rev. D* 105 (2022), p. 076010. DOI: 10.1103/PhysRevD.105.076010. eprint: 2111.01170 (hep-ph).

[288] John Collins. *Foundations of Perturbative QCD*. Vol. 32. Cambridge Monographs on Particle Physics, Nuclear Physics and Cosmology. Cambridge University Press, July 2023. ISBN: 978-1-009-40184-5, 978-1-009-40183-8, 978-1-009-40182-1. DOI: 10.1017/9781009401845.

[289] M. A. G. Aivazis, Frederick I. Olness, and Wu-Ki Tung. "Leptoproduction of heavy quarks. 1. General formalism and kinematics of charged current and neutral current production processes". In: *Phys. Rev. D* 50 (1994), p. 3085. DOI: 10.1103/PhysRevD.50.3085. arXiv: hep-ph/9312318.

[290] E. Moffat et al. "What does kinematical target mass sensitivity in DIS reveal about hadron structure?" In: *Phys. Rev. D* 99.9 (2019), p. 096008. DOI: 10.1103/PhysRevD.99.096008.

[291] W. Melnitchouk, Andreas W. Schreiber, and Anthony William Thomas. "Deep inelastic scattering from off-shell nucleons". In: *Phys. Rev. D* 49 (1994), p. 1183. DOI: 10.1103/PhysRevD.49.1183. arXiv: nucl-th/9311008.

[292] W. Melnitchouk, Andreas W. Schreiber, and Anthony William Thomas. "Relativistic deuteron structure function". In: *Phys. Lett. B* 335 (1994), p. 11. DOI: 10.1016/0370-2693(94)91550-4. arXiv: nucl-th/9407007.

[293] Sergey A. Kulagin, G. Piller, and W. Weise. "Shadowing, binding and off-shell effects in nuclear deep inelastic scattering". In: *Phys. Rev. C* 50 (1994), p. 1154. DOI: 10.1103/PhysRevC.50.1154. arXiv: nucl-th/9402015.

[294] L. W. Whitlow et al. "Precise measurements of the proton and deuteron structure functions from a global analysis of the SLAC deep-inelastic electron scattering cross sections". In: *Phys. Lett. B* 282 (1992), p. 475. DOI: 10.1016/0370-2693(92)90672-Q.

[295] A. C. Benvenuti et al. "A high statistics measurement of the proton structure functions F_2 and R from deep-inelastic muon scattering at high Q^2". In: *Phys. Lett. B* 223 (1989), p. 485. DOI: 10.1016/0370-2693(89)91637-7.

[296] M. Arneodo et al. "Measurement of the proton and deuteron structure functions, F_2^p and F_2^d, and of the ratio σ_L/σ_T". In: *Nucl. Phys.* B483 (1997), p. 3. DOI: 10.1016/S0550-3213(96)00538-X. arXiv: hep-ph/9610231.

[297] M. Arneodo et al. "Accurate measurement of F_2^d/F_2^p and $R^d - R^p$". In: *Nucl. Phys.* B487 (1997), p. 3. DOI: 10.1016/S0550-3213(96)00673-6. arXiv: hep-ex/9611022.

[298] H. Abramowicz et al. "Combination of measurements of inclusive deep inelastic $e^{\pm}p$ scattering cross sections and QCD analysis of HERA data". In: *Eur. Phys. J. C* 75 (2015), p. 580. DOI: 10.1140/epjc/s10052-015-3710-4. arXiv: 1506.06042 [hep-ex].

[299] R. S. Towell et al. "Improved measurement of the anti-d / anti-u asymmetry in the nucleon sea". In: *Phys. Rev. D* 64 (2001), p. 052002. DOI: 10.1103/PhysRevD.64.052002. arXiv: hep-ex/0103030.

[300] J. Dove et al. "The asymmetry of antimatter in the proton". In: *Nature* 590 (2021). [Erratum: Nature **604** (2022) E26], p. 561. DOI: 10.1038/s41586-022-04707-z. arXiv: 2103.04024 [hep-ph].

[301] Timo Antero Aaltonen et al. "Measurement of $d\sigma/dy$ of Drell-Yan e^+e^- pairs in the Z Mass Region from $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV". In: *Phys. Lett. B* 692 (2010), p. 232. DOI: 10.1016/j.physletb.2010.06.043. arXiv: 0908.3914 [hep-ex].

[302] V. M. Abazov et al. "Measurement of the shape of the boson rapidity distribution for $p\bar{p} \to Z/\gamma^* \to e^+e^- + X$ events produced at $\sqrt{s}$ of 1.96-TeV". In: *Phys. Rev. D* 76 (2007), p. 012003. DOI: 10.1103/PhysRevD.76.012003. arXiv: hep-ex/0702025.

[303] T. Aaltonen et al. "Direct measurement of the W production charge asymmetry in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV". In: *Phys. Rev. Lett.* 102 (2009), p. 181801. DOI: 10.1103/PhysRevLett.102.181801. arXiv: 0901.2169 [hep-ex].

[304] Victor Mukhamedovich Abazov et al. "Measurement of the W boson production charge asymmetry in $p\bar{p} \to W+X \to e\nu+X$ events at $\sqrt{s} = 1.96$ TeV". In: *Phys. Rev. Lett.* 112 (2014). [Erratum: Phys. Rev. Lett. **114** (2015) 049901], p. 151803. DOI: 10.1103/PhysRevLett.112.151803. arXiv: 1312.2895 [hep-ex].

[305] A. Abulencia et al. "Measurement of the inclusive jet cross section using the k_T algorithmin $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV with the CDF II detector". In: *Phys. Rev. D* 75 (2007). [Erratum: Phys. Rev. D **75** (2007) 119901], p. 092006. DOI: 10.1103/PhysRevD.75.092006. arXiv: hep-ex/0701051.

[306] Victor Mukhamedovich Abazov et al. "Measurement of the inclusive jet cross section in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV". In: *Phys. Rev. D* 85 (2012), p. 052006. DOI: 10.1103/PhysRevD.85.052006. arXiv: 1110.3771 [hep-ex].

[307] B. Abi et al. "Measurement of the positive muon anomalous magnetic moment to 0.46 ppm". In: *Phys. Rev. Lett.* 126 (2021), p. 141801. DOI: 10.1103/PhysRevLett.126.141801. arXiv: 2104.03281 [hep-ex].

[308] G. W. Bennett et al. "Final report of the muon E821 anomalous magnetic moment measurement at BNL". In: *Phys. Rev. D* 73 (2006), p. 072003. DOI: 10.1103/PhysRevD.73.072003. arXiv: hep-ex/0602035.

[309] Sz. Borsanyi et al. "Leading hadronic contribution to the muon magnetic moment from lattice QCD". In: *Nature* 593 (2021), p. 51. DOI: 10.1038/s41586-021-03418-1. arXiv: 2002.12347 [hep-lat].

[310] T. Aoyama et al. "The anomalous magnetic moment of the muon in the Standard Model". In: *Phys. Rept.* 887 (2020), p. 1. DOI: 10.1016/j.physrep.2020.07.006. arXiv: 2006.04822 [hep-ph].

[311] Peter Athron et al. "Combined collider constraints on neutralinos and charginos". In: *Eur. Phys. J. C* 79 (2019), p. 395. DOI: 10.1140/epjc/s10052-019-6837-x. arXiv: 1809.02097 [hep-ph].

[312] L. A. Harland-Lang et al. "The impact of the final HERA combined data on PDFs obtained from a global fit". In: *Eur. Phys. J. C* 76 (2016), p. 186. DOI: 10.1140/epjc/s10052-016-4020-1. arXiv: 1601.03413 [hep-ph].

[313] Georges Aad et al. "Measurement of the W-boson mass and width with the ATLAS detector using proton-proton collisions at $\sqrt{s} = 7$ TeV". In: (Mar. 2024). arXiv: 2403.15085 [hep-ex].

[314] *Measurement of the W boson mass in proton-proton collisions at sqrts = 13 TeV*. Tech. rep. Geneva: CERN, 2024. URL: https://cds.cern.ch/record/2910372.

[315] Léo Morel et al. "Determination of the fine-structure constant with an accuracy of 81 parts per trillion". In: *Nature* 588 (2020), p. 61. DOI: 10.1038/s41586-020-2964-7.

[316] X. Fan et al. "Measurement of the electron magnetic moment". In: *Phys. Rev. Lett.* 130 (2023), p. 071801. DOI: 10.1103/PhysRevLett.130.071801. arXiv: 2209.13084 [physics.atom-ph].

[317] S. Navas et al. "Review of particle physics". In: *Phys. Rev. D* 110.3 (2024), p. 030001. DOI: 10.1103/PhysRevD.110.030001.

[318] Asli M. Abdullahi et al. "Semi-Visible Dark Photon Phenomenology at the GeV Scale". In: *Phys. Rev. D* 108.1 (2023), p. 015032. DOI: 10.1103/PhysRevD.108.015032. arXiv: 2302.05410 [hep-ph].

[319] M. Williams et al. "New Constraints (and Motivations) for Abelian Gauge Bosons in the MeV-TeV Mass Range". In: *JHEP* 08 (2011), p. 106. DOI: 10.1007/JHEP08(2011)106. arXiv: 1103.4556 [hep-ph].

[320] Julian Heeck. "Unbroken B – L symmetry". In: *Phys. Lett. B* 739 (2014), pp. 256–262. DOI: 10.1016/j.physletb.2014.10.067. arXiv: 1408.6845 [hep-ph].

[321] G. P. Zeller et al. "A Precise Determination of Electroweak Parameters in Neutrino Nucleon Scattering". In: *Phys. Rev. Lett.* 88 (2002). [Erratum: Phys. Rev. Lett. **90**, 239902 (2003)], p. 091802. DOI: 10.1103/PhysRevLett.88.091802. arXiv: hep-ex/0110059.

[322] F. J. Escrihuela et al. "Global constraints on muon-neutrino non-standard interactions". In: *Phys. Rev. D* 83 (2011), p. 093002. DOI: 10.1103/PhysRevD.83.093002. arXiv: 1103.1366 [hep-ph].

[323] W. Bentz et al. "Reassessment of the NuTeV determination of the weak mixing angle". In: *Phys. Lett. B* 693 (2010), pp. 462–466. DOI: 10.1016/j.physletb.2010.09.001. arXiv: 0908.3198 [nucl-th].

[324] Vernon Barger, Paul Langacker, and Hye-Sung Lee. "Primordial nucleosynthesis constraints on Z' properties". In: *Phys. Rev. D* 67 (2003), p. 075009. DOI: 10.1103/PhysRevD.67.075009. arXiv: hep-ph/0302066.

[325] Joshua Albert et al. "A comparison of Bayesian sampling algorithms for high-dimensional particle physics and cosmology applications". In: (Sept. 2024). arXiv: 2409.18464 [hep-ph].

[326] Andrew Fowlie, Sebastian Hoof, and Will Handley. "Nested Sampling for Frequentist Computation: Fast Estimation of Small p-Values". In: *Phys. Rev. Lett.* 128.2 (2022), p. 021801. DOI: 10.1103/PhysRevLett.128.021801. arXiv: 2105.13923 [physics.data-an].

[327] Charles C. Margossian and Andrew Gelman. "For how many iterations should we run Markov chain Monte Carlo?" In: *arXiv e-prints*, arXiv:2311.02726 (Nov. 2023), arXiv:2311.02726. DOI: 10.48550/arXiv.2311.02726. arXiv: 2311.02726 [stat.CO].

[328] Edward Higson et al. "Dynamic nested sampling: an improved algorithm for parameter estimation and evidence calculation". In: *Statistics and Computing* 29.5 (Sept. 2019), pp. 891–913. DOI: 10.1007/s11222-018-9844-0. arXiv: 1704.03459 [stat.CO].

[329] Alan Heavens et al. "Marginal Likelihoods from Monte Carlo Markov Chains". In: (Apr. 2017). arXiv: 1704.03472 [stat.CO].

[330] Rahul Srinivasan et al. "*floZ*: Improved Bayesian evidence estimation from posterior samples with normalizing flows". In: (Apr. 2024). arXiv: 2404.12294 [stat.ML].

[331] Alicja Polanska et al. "Learned harmonic mean estimation of the Bayesian evidence with normalizing flows". In: (May 2024). arXiv: 2405.05969 [astro-ph.IM].

[332] Daniel Foreman-Mackey et al. "emcee: The MCMC Hammer". In: *pasp* 125.925 (Mar. 2013), p. 306. DOI: 10.1086/670067. arXiv: 1202.3665 [astro-ph.IM].

[333] Daniel Foreman-Mackey et al. "emcee v3: A Python ensemble sampling toolkit for affine-invariant MCMC". In: *The Journal of Open Source Software* 4.43, 1864 (Nov. 2019), p. 1864. DOI: 10.21105/joss.01864. arXiv: 1911.07688 [astro-ph.IM].

[334] Jonathan Goodman and Jonathan Weare. "Ensemble samplers with affine invariance". In: *Communications in Applied Mathematics and Computational Science* 5.1 (Jan. 2010), pp. 65–80. DOI: 10.2140/camcos.2010.5.65.

[335] W. D. Vousden, W. M. Farr, and I. Mandel. "Dynamic temperature selection for parallel tempering in Markov chain Monte Carlo simulations". In: *mnras* 455.2 (Jan. 2016), pp. 1919–1937. DOI: 10.1093/mnras/stv2422. arXiv: 1501.05823 [astro-ph.IM].

[336] Robert H. Swendsen and Jian-Sheng Wang. "Replica Monte Carlo Simulation of Spin-Glasses". In: *Phys. Rev. Lett.* 57 (21 Nov. 1986), pp. 2607–2609. DOI: 10.1103/PhysRevLett.57.2607. URL: https://link.aps.org/doi/10.1103/PhysRevLett.57.2607.

[337] Charles J Geyer. *Estimating normalizing constants and reweighting mixtures*. Tech. rep. School of Statistics, University of Minnesota, 1994. URL: https://hdl.handle.net/11299/58433.

[338] David J. Earl and Michael W. Deem. "Parallel tempering: Theory, applications, and new perspectives". In: *Physical Chemistry Chemical Physics (Incorporating Faraday Transactions)* 7.23 (Jan. 2005), p. 3910. DOI: 10.1039/B509983H. arXiv: physics/0508111 [physics.comp-ph].

[339] Minas Karamanis, Florian Beutler, and John A. Peacock. "zeus: a python implementation of ensemble slice sampling for efficient Bayesian parameter inference". In: *Mon. Not. Roy. Astron. Soc.* 508.3 (2021), pp. 3589–3603. DOI: 10.1093/mnras/stab2867. arXiv: 2105.03468 [astro-ph.IM].

[340] Radford M. Neal. "Slice sampling". In: *The Annals of Statistics* 31.3 (June 2003). ISSN: 0090-5364. DOI: 10.1214/aos/1056562461. URL: http://dx.doi.org/10.1214/aos/1056562461.

[341] Minas Karamanis and Florian Beutler. "Ensemble slice sampling". In: *Stat. Comput.* 31.5 (2021), p. 61. DOI: 10.1007/s11222-021-10038-2. arXiv: 2002.06212 [stat.ML].

[342] F. Feroz, M. P. Hobson, and M. Bridges. "MULTINEST: an efficient and robust Bayesian inference tool for cosmology and particle physics". In: *mnras* 398 (Oct. 2009), pp. 1601–1614. DOI: 10.1111/j.1365-2966.2009.14548.x. arXiv: 0809.3437.

[343] F. Feroz et al. "Importance Nested Sampling and the MultiNest Algorithm". In: (June 2013). arXiv: 1306.2144 [astro-ph.IM].

[344] Joshua S. Speagle. "dynesty: a dynamic nested sampling package for estimating Bayesian posteriors and evidences". In: *Mon. Not. Roy. Astron. Soc.* 493.3 (2020), pp. 3132–3158. DOI: 10.1093/mnras/staa278. arXiv: 1904.02180 [astro-ph.IM].

[345] Johannes Buchner. "A statistical test for Nested Sampling algorithms". In: *Statistics and Computing* 26.1-2 (Jan. 2016), pp. 383–392. DOI: 10.1007/s11222-014-9512-y. arXiv: 1407.5459 [stat.CO].

[346] Johannes Buchner. "Collaborative Nested Sampling: Big Data versus Complex Physical Models". In: *pasp* 131.1004 (Oct. 2019), p. 108005. DOI: 10.1088/1538-3873/aae7fc. arXiv: 1707.04476 [stat.CO].

[347] Michael J. Williams, John Veitch, and Chris Messenger. "Importance nested sampling with normalising flows". In: *Mach. Learn. Sci. Tech.* 4.3 (2023), p. 035011. DOI: 10.1088/2632-2153/acd5aa. arXiv: 2302.08526 [astro-ph.IM].

[348] Joshua G. Albert. "JAXNS: a high-performance nested sampling package based on JAX". In: *arXiv e-prints*, arXiv:2012.15286 (Dec. 2020), arXiv:2012.15286. DOI: 10.48550/arXiv.2012.15286. arXiv: 2012.15286 [astro-ph.IM].

[349] Johannes U. Lange. *NAUTILUS: boosting Bayesian importance nested sampling with deep learning*. 2023. arXiv: 2306.16923 [astro-ph.IM].

[350] Greg Ashton et al. "Nested sampling for physical scientists". In: *Nature* 2 (2022). DOI: 10.1038/s43586-022-00121-x. arXiv: 2205.15570 [stat.CO].

[351] Johannes Buchner. "Comparison of Step Samplers for Nested Sampling". In: *Physical Sciences Forum*. Vol. 5. Physical Sciences Forum. Dec. 2022, 46, p. 46. DOI: 10.3390/psf2022005046. arXiv: 2211.09426 [stat.CO].

[352] Mary Kathryn Cowles and Bradley P. Carlin. "Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review". In: *Journal of the American Statistical Association* 91.434 (1996), pp. 883–904. DOI: 10.1080/01621459.1996.10476956.

[353] Charles J Geyer. "Introduction to Markov Chain Monte Carlo". In: *Handbook of Markov Chain Monte Carlo*. Ed. by Steve Brooks et al. Chapman and Hall/CRC, May 2011. DOI: 10.1201/b10905. URL: https://doi.org/10.1201/b10905.

[354] Ravin Kumar et al. "ArviZ a unified library for exploratory analysis of Bayesian models in Python". In: *Journal of Open Source Software* 4.33 (2019), p. 1143. DOI: 10.21105/joss.01143.

[355] Andrew Gelman and Donald B. Rubin. "Inference from Iterative Simulation Using Multiple Sequences". In: *Statistical Science* 7.4 (1992), pp. 457–472. ISSN: 08834237.

[356] Stephen P. Brooks and Andrew Gelman. "General Methods for Monitoring Convergence of Iterative Simulations". In: *Journal of Computational and Graphical Statistics* 7.4 (1998), pp. 434–455. DOI: 10.1080/10618600.1998.10474787.

[357] Aki Vehtari et al. "Rank-normalization, folding, and localization: An improved $\widehat{R}$ for assessing convergence of MCMC". In: *Bayesian Analysis* 16.2, arXiv:1903.08008 (Mar. 2019). DOI: 10.1214/20-ba1221. arXiv: 1903.08008 [stat.CO].

[358] Charles C. Margossian et al. "Nested $\widehat{R}$: Assessing the convergence of Markov chain Monte Carlo when running many short chains". In: (Oct. 2021). arXiv: 2110.13017 [stat.ME].

[359] Andrew Fowlie, Will Handley, and Liangliang Su. "Nested sampling cross-checks using order statistics". In: *Mon. Not. Roy. Astron. Soc.* 497.4 (2020), pp. 5256–5263. DOI: 10.1093/mnras/staa2345. arXiv: 2006.03371 [stat.CO].

[360] H. H. Rosenbrock. "An Automatic Method for Finding the Greatest or Least Value of a Function". In: *The Computer Journal* 3.3 (Mar. 1960), pp. 175–184. ISSN: 1460-2067. DOI: 10.1093/comjnl/3.3.175. URL: http://dx.doi.org/10.1093/comjnl/3.3.175.

[361] Leonard Andreevič Rastrigin. "Systems of extremal control". In: *Nauka* (1974).

[362] Pablo Lemos et al. "Improving Gradient-guided Nested Sampling for Posterior Inference". In: *arXiv e-prints*, arXiv:2312.03911 (Dec. 2023), arXiv:2312.03911. DOI: 10.48550/arXiv.2312.03911. arXiv: 2312.03911 [cs.LG].

[363] Julien Lesgourgues. *The Cosmic Linear Anisotropy Solving System (CLASS) I: Overview*. 2011. arXiv: 1104.2932 [astro-ph.IM].

[364] Alex Cole et al. "Fast and credible likelihood-free cosmology with truncated marginal neural ratio estimation". In: *Journal of Cosmology and Astroparticle Physics* 2022.09 (Sept. 2022), p. 004. ISSN: 1475-7516. DOI: 10.1088/1475-7516/2022/09/004. URL: http://dx.doi.org/10.1088/1475-7516/2022/09/004.

[365] Alessio Spurio Mancini et al. "CosmoPower: emulating cosmological power spectra for accelerated Bayesian inference from next-generation surveys". In: *Mon. Not. Roy. Astron. Soc.* 511.2 (2022), pp. 1771–1788. DOI: 10.1093/mnras/stac064. arXiv: 2106.03846 [astro-ph.CO].

[366] Günter Klambauer et al. *Self-Normalizing Neural Networks*. 2017. arXiv: 1706.02515 [cs.LG].

[367] Diederik P. Kingma and Jimmy Ba. *Adam: A Method for Stochastic Optimization*. 2017. arXiv: 1412.6980 [cs.LG].

[368] Davide Piras et al. "The future of cosmological likelihood-based inference: accelerated high-dimensional parameter estimation and model comparison". In: (May 2024). arXiv: 2405.12965 [astro-ph.CO].

[369] A. G. Adame et al. "DESI 2024 VI: Cosmological Constraints from the Measurements of Baryon Acoustic Oscillations". In: (Apr. 2024). arXiv: 2404.03002 [astro-ph.CO].

[370] P. Athron et al. "A global fit of the MSSM with GAMBIT". In: *Eur. Phys. J. C* 77, 879 (Dec. 2017), p. 879. DOI: 10.1140/epjc/s10052-017-5196-8. arXiv: 1705.07917 [hep-ph].

[371] A. Buckley, A. Shilton, and M. J. White. "Fast supersymmetry phenomenology at the Large Hadron Collider using machine learning techniques". In: *Comp. Phys. Comm.* 183 (Apr. 2012), pp. 960–970. DOI: 10.1016/j.cpc.2011.12.026. arXiv: 1106.4613 [hep-ph].

[372] G. Brooijmans et al. "Les Houches 2019 Physics at TeV Colliders: New Physics Working Group Report". In: *11th Les Houches Workshop on Physics at TeV Colliders: PhysTeV Les Houches*. Feb. 2020. arXiv: 2002.12220 [hep-ph].

[373] W. Armstrong et al. "Revealing Color Forces with Transverse Polarized Electron Scattering". In: *Phys. Rev. Lett.* 122.2 (2019), p. 022002. DOI: 10.1103/PhysRevLett.122.022002.

[374] B. Adeva et al. "Polarized quark distributions in the nucleon from semiinclusive spin asymmetries". In: *Phys. Lett. B* 420 (1998), pp. 180–190. DOI: 10.1016/S0370-2693(97)01546-3.

[375] K. Ackerstaff et al. "Flavor decomposition of the polarized quark distributions in the nucleon from inclusive and semiinclusive deep inelastic scattering". In: *Phys. Lett. B* 464 (1999), pp. 123–134. DOI: 10.1016/S0370-2693(99)00964-8.

[376] M. S. Abdallah et al. "Longitudinal double-spin asymmetry for inclusive jet and dijet production in polarized proton collisions at $\sqrt{s} = 510$ GeV". In: *Phys. Rev. D* 105.9 (2022), p. 092011. DOI: 10.1103/PhysRevD.105.092011.

[377] L. Kish. "Survey Sampling". In: *Biometrische Zeitschrift* 10.1 (1968), pp. 88–89. DOI: 10.1002/bimj.19680100122.

[378] Víctor Elvira, Luca Martino, and Christian P. Robert. "Rethinking the Effective Sample Size". In: *International Statistical Review* 90.3 (2022), pp. 525–550. DOI: 10.1111/insr.12500.

[379] Andrew Gelman et al. *Bayesian data analysis*. Third. Chapman & Hall/CRC Texts in Statistical Science Series. CRC, 2013. ISBN: 9781439840955 1439840954.

[380] Will Handley. "anesthetic: nested sampling visualisation". In: *The Journal of Open Source Software* 4.37 (June 2019), p. 1414. DOI: 10.21105/joss.01414. URL: http://dx.doi.org/10.21105/joss.01414.

[381] Edward Higson et al. "Sampling Errors in Nested Sampling Parameter Estimation". In: *Bayesian Analysis* 13.3 (Mar. 2018), pp. 873–896. DOI: 10.1214/17-BA1075. arXiv: 1703.09701 [stat.ME].

[382] Edward Higson. "nestcheck: error analysis, diagnostic tests and plots for nested sampling calculations". In: *Journal of Open Source Software* 3.29 (2018), p. 916. DOI: 10.21105/joss.00916. URL: http://joss.theoj.org/papers/10.21105/joss.00916.

[383] Edward Higson et al. "nestcheck: diagnostic tests for nested sampling calculations". In: *Monthly Notices of the Royal Astronomical Society* 483.2 (2019), pp. 2044–2056. DOI: 10.1093/mnras/sty3090. URL: http://doi.org/10.1093/mnras/sty3090.