Fitting parton distributions with LHC data

Chris Deans

NNPDF Collaboration
University of Edinburgh

11th September 2012

THE UNIVERSITY *of* EDINBURGH

Parton distribution functions (PDFs)

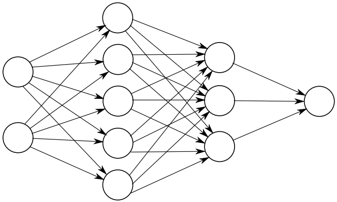
- Parton distribution functions (PDFs) characterize the structure of the proton
- PDF $f_i(x, Q^2)$ describes the probability of finding a parton of type i (i.e. quark flavor or gluon) with fraction x of the total momentum of the proton
- Factorization theorems bundle soft interactions in proton into PDFs, giving them scale dependence and evolution
- PDFs are a necessary for calculation of theory predictions e.g. for LHC

$$\sigma\left(Q^2\right) = \int dx_1 dx_2 \sum_{i,j} f_i\left(x_1, Q^2\right) f_j\left(x_2, Q^2\right) \hat{\sigma}_{i,j}\left(x_1, x_2, Q^2\right)$$

NNPDF approach - key features

- Use data from wide variety of hadronic processes (e.g DIS, W/Z cross sections from colliders, fixed target Drell-Yan experiments)
- Model PDFs using neural networks which provide an unbiased parameterization
- Vary parameters comparing theoretical predictions made with PDFs to data until best fit is found
- Uncertainties on PDFs are determined by the generation of Monte Carlo replica datasets using experimental uncertainties and fitting to each replica separately

Neural network PDFs



- Neural networks are mathematical models inspired by the structure of the brain
- Comprised of connected nodes; Value at each node is a function of the values of those around it
- For NNPDF fits, seven PDF combinations are modeled each with a separate neural network
- This provides a flexible, unbiased parameterization

Minimization

- Best fit is found by minimizing χ^2 comparing experimental data with predictions from PDFs:

$$\chi^2(f_i) = \sum_{i,j} \left(s_i^{dat} - s_i^{th}(f_i) \right) cov_{ij}^{-1} \left(s_j^{dat} - s_j^{th}(f_i) \right)$$

- In NNPDF approach a genetic algorithm is used to perform the minimization
- However, large flexibility of neural networks introduces potential for overfitting
- Need to use cross validation: Train networks on half of the data, use other half to detect overfitting

Monte Carlo replicas

- Best fit values alone are not sufficient for modern colliders, need corresponding PDF uncertainties
- NNPDF strategy:
  - Generate large number of replica datasets according to probability distributions suggested by experimental uncertainties
  - Fit PDFs to each replica dataset separately
  - Resulting set of replica PDFs allows uncertainties to be calculated by taking standard deviations or confidence intervals

NNPDF2.3 dataset

New NNPDF fit: NNPDF2.3 [arXiv:1207.1303]

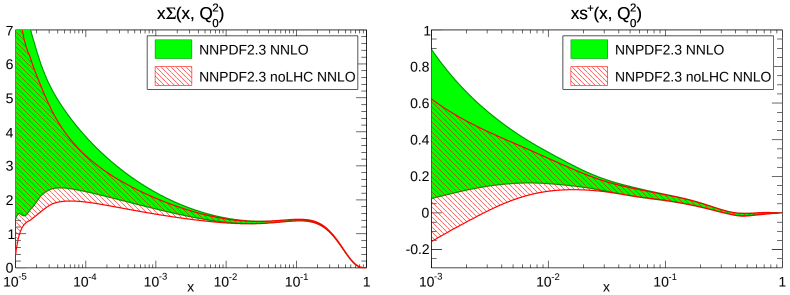
This is the first PDF fit to include LHC data, including all data where full covariance matrix is available:

- ATLAS Inclusive Jets, $36\mathrm{pb}^{-1}$ [arXiv:1112.6297]
- ATLAS W/Z lepton rapidity distributions, $36\mathrm{pb}^{-1}$ [arXiv:1109.5141]
- CMS W lepton asymmetry, $840\mathrm{pb}^{-1}$ [arXiv:1206.2598]
- LHCb W rapidity distributions, $36\mathrm{pb}^{-1}$ [arXiv:1204.1620]

plus data from DIS, Tevatron, fixed target Drell-Yan...

Impact of LHC data

Even with only (relatively) small amount of data some non-negligible impact can be observed



With more statistics constraining power will increase

Future LHC data

In the future we will see:

- Better statistics
- New observables e.g.:
  - top pair production - constrains gluon at large x
  - $W+$ charm - constrains strange distributions
- 2014+: Higher energies, constrains new regions in Q^2 and x

Summary

- NNPDF methodology provides a successful way to determine PDFs and uncertainties
- NNPDF2.3 is the first PDF fit to include LHC data
- Available LHC data already has small impact on PDFs, and as the LHC program continues and more data is released this impact will increase

NNPDF website: `nnpdf.hepforge.org`