xFitter – a few new ideas

xFitter Workshop

March 10th, 2022

Simone Amoroso (DESY)

Introduction

- Increasing interest in the community towards global interpretation of LHC data
  - Global SMEFT interpretations, fits of EW or other SM parameters from multiple datasets/ processes, large combinations
  - Circular interplay with the PDFs, which routinely include LHC data in fits
- Simultaneous determinations of PDFs with other parameters will become more and more common

- xFitter in an ideal position to become the standard tool for these kind of analyses by the LHC Collaboration
- While very flexible, xFitter has so far concentrated mostly (and obviously so) on PDF related developments
- Needs a strong push towards implementing the key features needed to adapt it to the new problems

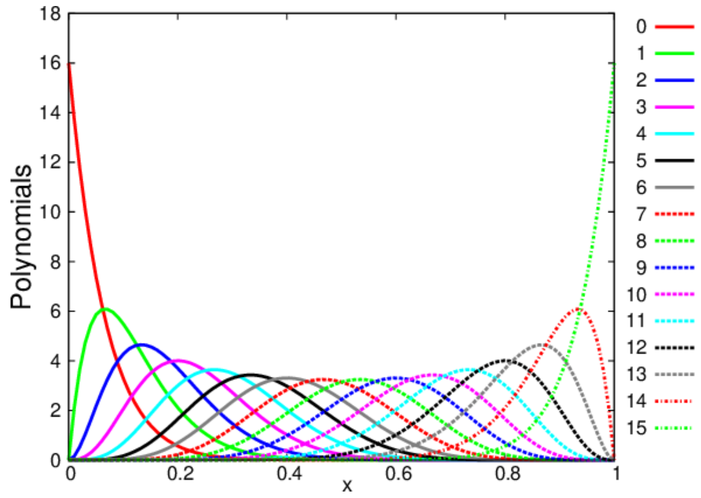
Minimisers

- Since a while Sasha wrote an interface to the CERES minimiser
  - For typical fits faster convergence (time/iterations) than MINUIT
  - But not yet fully integrated:
    - One doesn't get the full result outputs as for MINUIT
    - Minimiser options (many) should also be propagated to parameters.yaml
  - CERES supports multithreading, but should be enabled/testes
  - And can exploit numeric and automatic differentiation.
    Can we give it a try (at least for some of the parametrisation)?

- Fixing the parameters interface, opens the way to interfacing other minimisers,
  and different approaches to minimisation (evolutionary algorithms, particle swarm, …)
  - LPOPT, ALGENCAN, scipy.optimize, NSGA

New simple parametrisation

- With more and more complex analyses the xFitter HERAPDF PDF parametrisation begins to show its limits
  - Polynomial parameters act over a wide range of x
  - Often large correlations arise between parameters, making minimisation difficult



- Bernstein Polynomials, as in CTEQ
  - Available from this library
- Chebyschev Polynomials, as in MSHT
  - Taken from BOOST library
- Private implementation in xFitter, pending a bit of validation and some test case usage
- Also, Sasha's idea of a Q^2 independent parametrisation (more later today)

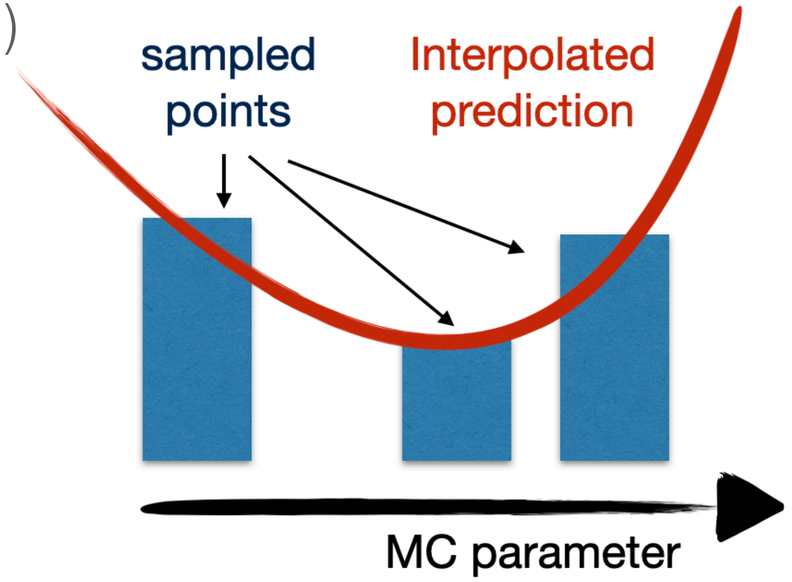
And more complex ones

- Neural-network based parametrisation used in PDF fits since > 10 years (NNPDF)
  - Allow greater flexibility than any given functional form
  - Comes at the price of a much larger number of parameters

- Recently, work by Valerio to implement analytic derivatives for NN
  - Implemented in C++ allows for fast backpropagation/training of the NN parameters, done using CERES
  - Should be technically trivial to interface to xFitter
- Can also exploit ML approaches to approximate the χ^2/likelihood dependence on PDFs (arXiv:2201.06586)
  - Obtains a speedup of several order of magnitudes over a naive sampling of the χ^2 distribution by fitting pseudo-data
  - Allows costless χ^2 scans which would be relevant for PDF+EW/SMEFT fits

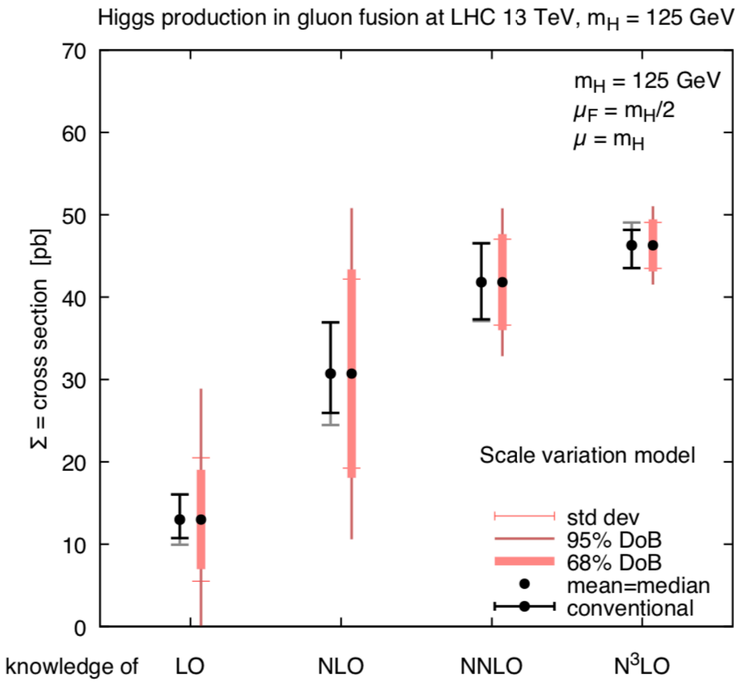
Minimising external functions

- Dependence on non-PDF parameters (SMEFT, $m_{t/W}$, …) typically cannot be included in grids, and needs to be parametrised from external predictions
  - I.e. CMS top fits parametrised top mass dependence using splines
- Typical problem in MC tuning, and we could reuse existing solutions
  - Analytic interpolation of the dependence of each bin on variations of external parameters derived from a discretised sampling of the parameter space
  - Professor code allows to take Rivet input and performs a fast polynomial interpolation
  - Other solutions exists (rational approximants, SVN/NN, …)
- An interface of xFitter to these interpolation functions would open the way to many interesting studies
  - Interpolation derived externally and passed to xFitter through a txt (pickle?) file



Theoretical uncertainties

- Inclusion of theoretical uncertainties in PDF fits is nowadays a hot topic
  - Very relevant for PDF+SM parameters or PDF+SMEFT combined fits
- For the usually dominant *missing higher orders* no “correct” solution exists, as scale variations do not have any statistical interpretation
- Yet, several approaches have came out with recipes on how these uncertainties can be included in fits
  - Including the missing higher orders as nuisances, but ensuring the fit constraints do not affect the parameter(s) of interest - *custodial nuisance parameters*
  - Building a *probabilistic model* making some assumptions on the behaviour of the pQCD series (Cacciari-Hudeau, Bonvini, Duhr et al, …)



EPJC 80 (2020) 989

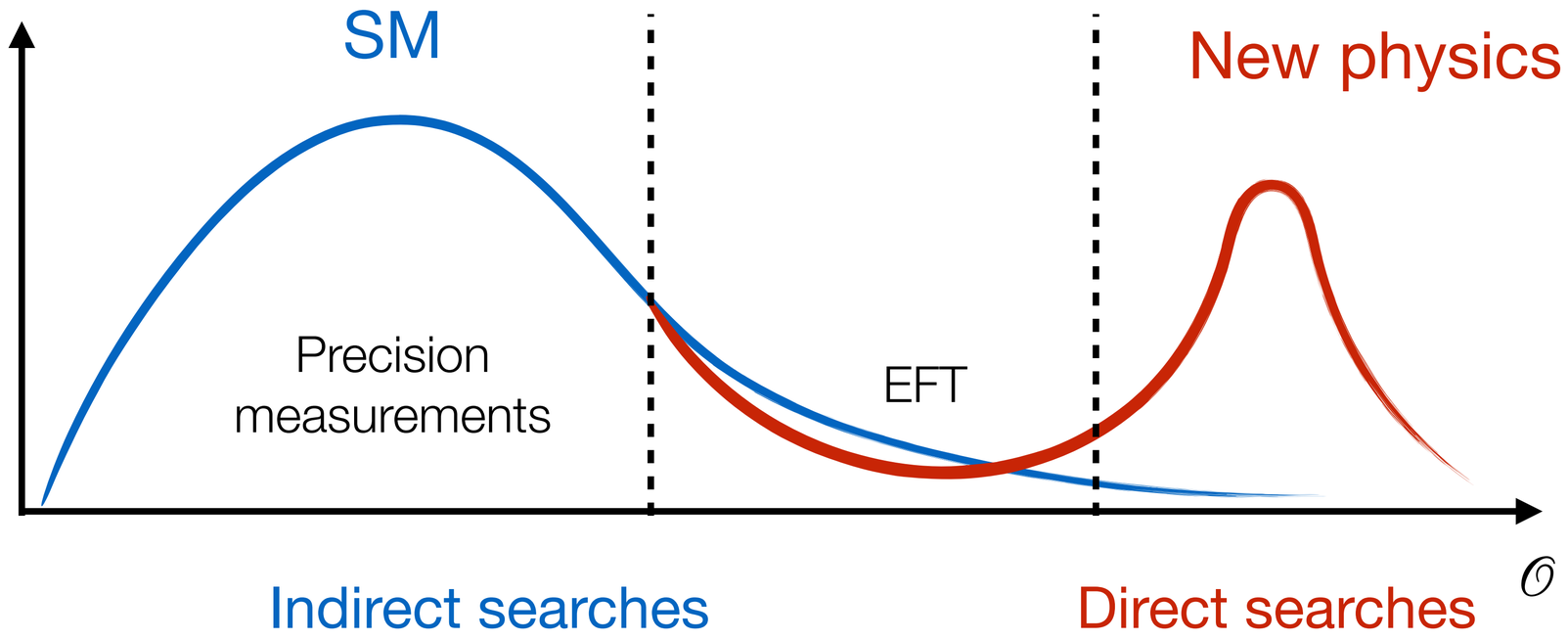
BACKUP

Documentation and support

- As we discussed a couple of meetings ago, I find the current code development/issue tracker not very nice.
- Following what every other software project in HEP is doing, I would suggest the following:
  - Move the code to a public repo (not limited to cern users)
  - Make the repository itself the main xFitter webpage
  - Allow everybody to open issues and merge requests
  - Move the issue reporting (both for developers and users) from JIRA to gitlab/github
    - Meaning the issues list will become a browsable FAQ (technical and physics)
  - Create a real documentation page including both technical and physics explanations, examples, tutorials …
  - Prepare containers for each new release (for newcomers, schools, lectures)

Searching for new physics

- The LHC mass reach has hit an asymptote, very likely new physics is heavy and beyond the LHC direct reach



- Model independent
- Can be reinterpreted in any UV completion and with the availability of better theory
- Sensitive to O(0.1) - O(0.01) effects

- Model dependent
- Challenging reinterpretation, requires detector simulation
- Sensitive to O(1) effects

The SM Effective Field Theory

- The SM EFT provides a framework to parametrise a generic theory for $E \ll \Lambda$

$$\mathcal{L}_{\text{SMEFT}} = \mathcal{L}_{\text{SM}} + \sum_i^{N_{d6}} \frac{c_i}{\Lambda^2} \mathcal{O}_i^{(6)} + \sum_j^{N_{d8}} \frac{b_j}{\Lambda^4} \mathcal{O}_j^{(8)} + \dots ,$$

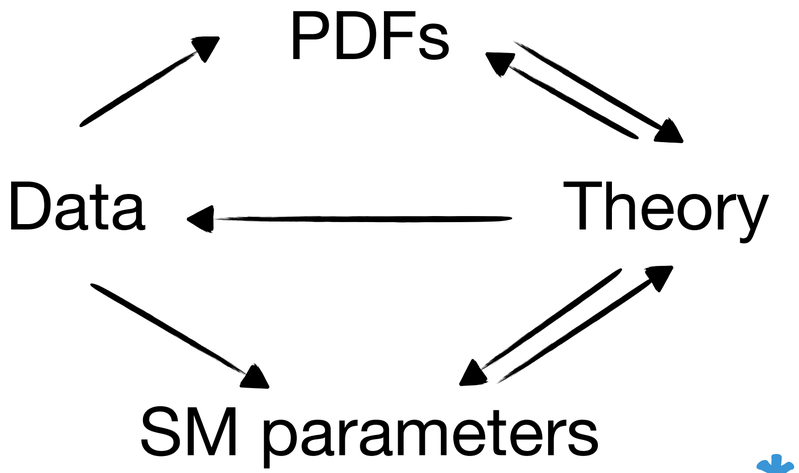
- Contribution of certain operators grows with partonic energy
  - Increased **sensitivity in TeV region**: unique feature at the LHC

$$\sigma(E) = \sigma_{\text{SM}}(E) \left(1 + \sum_i^{N_{d6}} \omega_i \frac{c_i \, m_{\text{SM}}^2}{\Lambda^2} + \sum_i^{N_{d6}} \widetilde{\omega}_i \frac{c_i \, E^2}{\Lambda^2} + \mathcal{O}\left(\Lambda^{-4}\right) \right)$$

- Large number of operators, complex phase space to explore
  - dim-6: **59 operators with MFV** (but > 2000 if no flavour assumption)

The issue with PDFs

- Knowledge of Parton Distribution Functions is a prerequisite for making precise predictions at hadron colliders



- **Circular interplay** between PDF fits, determinations of SM parameters and of EFT operators
- LHC high-energy data routinely included in global PDF fits

- A few important questions arise:

- **PDFs:** Are we **fitting away new physics** in global PDF determinations?
- **SM parameters:** Is there a **hidden bias** due to correlation with PDFs?
- **SMEFT:** Would **bounds change** with PDFs fitted under changes of the operators considered?

- The way forward: **simultaneous determination of the PDFs** and of SM parameters or SMEFT Wilson coefficients