Parton Distributions in the SMEFT

James Moore

27th May 2021

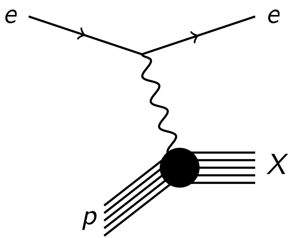
Structure of talk

1. Background: PDFs, EFTs and the SMEFT
2. ‘Standard’ simultaneous determination of PDFs and SMEFT couplings
3. Efficient simultaneous determination of PDFs and SMEFT couplings

Background: PDFs, EFTs and the SMEFT

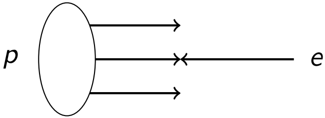
What is a PDF?

- Hadrons are bound states in QCD - we cannot understand their structure perturbatively with current methods.
- *Question: How do we make predictions for experiments involving hadrons?*
- Consider this problem in the ‘model’ case: *deep-inelastic scattering* (DIS), pictured below. How can we obtain the cross-section without a perturbative description of the hadronic state $|p\rangle$?



What is a PDF?

- *Idea:* Feynman (1969) came up with the *parton model* to answer this question. In a frame where the proton is *ultra-relativistic*, time dilation causes the proton's constituents to interact very slowly - they appear free.
- Suggests that electrons instantaneously scatter off individual hadron constituents called *partons* (= part of a proton), now known to be quarks and gluons.



What is a PDF?

- Feynman's parton model implies that total cross-section can be written in the form

$$\sigma = \sum_{\substack{\text{parton species} \\ q \text{ in proton}}} \int_0^1 dx\ f_q(x)\hat{\sigma}_q(x)$$

where:

  - x is the fraction of the proton's momentum carried by the struck parton.
  - $\hat{\sigma}_q(x)$ is the *partonic cross-section* - the cross-section for electron-parton scattering, with the initial parton having momentum fraction x. This can be computed in perturbation theory.
  - $f_q(x)$ are *parton distribution functions*, representing the probability density that the struck parton is of species q and carries momentum fraction x. These are non-perturbative, but *universal* (only depend on proton structure).

What is a PDF?

- Eventually the parton model was codified into a fully-fledged theory (*perturbative QCD*) derived from the basic principles of QCD. The key result is the *QCD factorisation theorem*, which for DIS states:

$$\sigma = \sum_q \int_0^1 dx\, \hat{\sigma}_q(x) f_q(x, \mu^2) + \text{corrections suppressed by energy scale.}$$

- **Important observation:** full treatment in QCD implies that the PDFs acquire an additional dependence, $f_q = f_q(x, \mu^2)$, on an arbitrary scale called the *factorisation scale*. Similar to renormalisation scale, a simple equation (the *DGLAP equation*) governs the μ^2 dependence of PDFs:

$$\mu^2 \frac{\partial f_q}{\partial \mu^2}(x, \mu^2) = \sum_{q'} \int_x^1 \frac{dy}{y}\, P_{qq'}\left(\frac{x}{y}\right) f_{q'}(y, \mu_F^2).$$

Usually chosen to be energy scale, $\mu^2 = Q^2$.

How are PDFs determined?

- PDFs non-perturbative $\Rightarrow$ determined by *fits to data*.
- Basic outline:

1. PDFs written in some parametrisation at initial scale Q_0, e.g. NNPDF collaboration use *neural network* (advantage: unbiased).
2. Evolved to all scales using DGLAP equation.
3. Minimising the goodness-of-fit statistic to experimental data at each scale then allows PDF parameters to be determined:

$$\chi^2 = (\text{data} - \text{theory(PDFs)})^T \text{covariance}^{-1} (\text{data} - \text{theory(PDFs)}).$$

How are PDFs determined?

- Experimental error propagated by *Monte Carlo replica* approach.
- N_{rep} 'pseudodata' copies are made, and an ensemble of N_{rep} PDFs are created fitting to each copy of the pseudodata in turn, $\{\mathbf{f}_1, \mathbf{f}_2, ..., \mathbf{f}_{N_{\text{rep}}}\}$ (here $\mathbf{f} = (f_u, f_d, f_s, ...)$).
- Ensemble properties can then be derived, e.g.

$$\mathbf{f}_0 = \text{mean (central) PDF} = \frac{1}{N_{\text{rep}}} \sum_i \mathbf{f}_i.$$

How are PDFs determined?

- **Important observation:** Fitted PDFs *depend on the theory in which the hard cross-section was computed*:

$$\sigma = \sum_q \int_0^1 dx\ \hat{\sigma}_q(x) f_q(x, Q^2).$$

  Often the only *consistent* way of fitting is to determine both theory parameters and PDFs *simultaneously*.

- *Toy example:* To extract strong coupling $\alpha_S(m_Z^2)$:

$$\sigma = \sum_q \int_0^1 dx\ (\hat{\sigma}_{\text{LO}} + \alpha_S(Q^2)\hat{\sigma}_{\text{NLO}}) f_q(x, Q^2).$$

  Fix PDFs $\Rightarrow$ can scan $\alpha_S(m_Z^2)$ values. But PDFs were determined with some fixed value of $\alpha_S(m_Z^2)$!

Main question

- The above discussion applies also to parameters in *beyond-the-Standard-Model theories* (BSM theories).
- In BSM physics searches, researchers always assume PDFs are fixed to SM values ('**black box PDFs**') - this is inconsistent, but is it a problem?
- Care about this problem because important in *indirect searches for new physics*: small deviations from SM in high-energy observables.
- Motivates following key question:

To what extent does a consistent, simultaneous fit of PDFs and BSM parameters affect bounds on the BSM parameters?

Effective field theories and the SMEFT

- PDF fitting group in Cambridge work with *effective field theories*, namely SMEFT, as BSM model of choice.
- An EFT is a *low-energy limit* of a renormalisable quantum field theory.
- Result is a Lagrangian with infinitely many terms, ordered in increasing powers of $1/\Lambda$, where Λ is an energy scale where EFT breaks down - scale of '**New Physics**'.
- Importantly: still renormalisable at any fixed order in $1/\Lambda$.

Effective field theories and the SMEFT

- $\Rightarrow$ Can treat the SM as a low-energy limit of some unknown theory by adding on all possible non-renormalisable terms consistent with the SM symmetries and built from SM fields. The result is the *Standard Model effective field theory* (SMEFT):

$$\mathcal{L}_{\mathsf{SMEFT}} = \mathcal{L}_{\mathsf{SM}} + \sum_{i=1}^{N_6} \frac{1}{\Lambda^2} \mathcal{O}_i^{(6)} + \cdots$$

  The SMEFT is sometimes called 'unbiased' as it should account for all possible renormalisable field theories of which it is is the low-energy limit.

- Very difficult problem to classify which operators can appear in the expansion, however solved for dimension 6.

Effective field theories and the SMEFT

- Summary of four-fermion operators in the Warsaw basis given in table below (from arXiv:1008.4884).

$(\bar{L}L)(\bar{L}L)$		$(\bar{R}R)(\bar{R}R)$		$(\bar{L}L)(\bar{R}R)$	
Q_{ll}	$(\bar{l}_p\gamma_\mu l_r)(\bar{l}_s\gamma^\mu l_t)$	Q_{ee}	$(\bar{e}_p\gamma_\mu e_r)(\bar{e}_s\gamma^\mu e_t)$	Q_{le}	$(\bar{l}_p\gamma_\mu l_r)(\bar{e}_s\gamma^\mu e_t)$
$Q_{qq}^{(1)}$	$(\bar{q}_p\gamma_\mu q_r)(\bar{q}_s\gamma^\mu q_t)$	Q_{uu}	$(\bar{u}_p\gamma_\mu u_r)(\bar{u}_s\gamma^\mu u_t)$	Q_{lu}	$(\bar{l}_p\gamma_\mu l_r)(\bar{u}_s\gamma^\mu u_t)$
$Q_{qq}^{(3)}$	$(\bar{q}_p\gamma_\mu \tau^I q_r)(\bar{q}_s\gamma^\mu \tau^I q_t)$	Q_{dd}	$(\bar{d}_p\gamma_\mu d_r)(\bar{d}_s\gamma^\mu d_t)$	Q_{ld}	$(\bar{l}_p\gamma_\mu l_r)(\bar{d}_s\gamma^\mu d_t)$
$Q_{lq}^{(1)}$	$(\bar{l}_p\gamma_\mu l_r)(\bar{q}_s\gamma^\mu q_t)$	Q_{eu}	$(\bar{e}_p\gamma_\mu e_r)(\bar{u}_s\gamma^\mu u_t)$	Q_{qe}	$(\bar{q}_p\gamma_\mu q_r)(\bar{e}_s\gamma^\mu e_t)$
$Q_{lq}^{(3)}$	$(\bar{l}_p\gamma_\mu \tau^I l_r)(\bar{q}_s\gamma^\mu \tau^I q_t)$	Q_{ed}	$(\bar{e}_p\gamma_\mu e_r)(\bar{d}_s\gamma^\mu d_t)$	$Q_{qu}^{(1)}$	$(\bar{q}_p\gamma_\mu q_r)(\bar{u}_s\gamma^\mu u_t)$
		$Q_{ud}^{(1)}$	$(\bar{u}_p\gamma_\mu u_r)(\bar{d}_s\gamma^\mu d_t)$	$Q_{qu}^{(8)}$	$(\bar{q}_p\gamma_\mu T^A q_r)(\bar{u}_s\gamma^\mu T^A u_t)$
		$Q_{ud}^{(8)}$	$(\bar{u}_p\gamma_\mu T^A u_r)(\bar{d}_s\gamma^\mu T^A d_t)$	$Q_{qd}^{(1)}$	$(\bar{q}_p\gamma_\mu q_r)(\bar{d}_s\gamma^\mu d_t)$
				$Q_{qd}^{(8)}$	$(\bar{q}_p\gamma_\mu T^A q_r)(\bar{d}_s\gamma^\mu T^A d_t)$

$(\bar{L}R)(\bar{R}L)$ and $(\bar{L}R)(\bar{L}R)$		B-violating	
Q_{ledq}	$(\bar{l}_p^j e_r)(\bar{d}_s q_t^j)$	Q_{duq}	$\varepsilon^{\alpha\beta\gamma}\varepsilon_{jk}\left[(d_p^\alpha)^T C u_r^\beta\right]\left[(q_s^{\gamma j})^T C l_t^k\right]$
$Q_{quqd}^{(1)}$	$(\bar{q}_p^j u_r)\varepsilon_{jk}(\bar{q}_s^k d_t)$	Q_{qqu}	$\varepsilon^{\alpha\beta\gamma}\varepsilon_{jk}\left[(q_p^{\alpha j})^T C q_r^{\beta k}\right]\left[(u_s^\gamma)^T C e_t\right]$
$Q_{quqd}^{(8)}$	$(\bar{q}_p^j T^A u_r)\varepsilon_{jk}(\bar{q}_s^k T^A d_t)$	Q_{qqq}	$\varepsilon^{\alpha\beta\gamma}\varepsilon_{jn}\varepsilon_{km}\left[(q_p^{\alpha j})^T C q_r^{\beta k}\right]\left[(q_s^{\gamma m})^T C l_t^n\right]$
$Q_{lequ}^{(1)}$	$(\bar{l}_p^j e_r)\varepsilon_{jk}(\bar{q}_s^k u_t)$	Q_{duu}	$\varepsilon^{\alpha\beta\gamma}\left[(d_p^\alpha)^T C u_r^\beta\right]\left[(u_s^\gamma)^T C e_t\right]$
$Q_{lequ}^{(3)}$	$(\bar{l}_p^j \sigma_{\mu\nu} e_r)\varepsilon_{jk}(\bar{q}_s^k \sigma^{\mu\nu} u_t)$		

Effective field theories and the SMEFT

- Total number of operators in Warsaw basis: 59 with additional flavour symmetry assumptions, 2599 without.

- Lots of parameters to fit! Ideally a *global simultaneous fit* of all couplings and PDFs at the same time, but this is impossible with current technology - instead, we focus on small numbers of couplings drawn from the SMEFT fitted simultaneously with PDFs.

‘Standard’ simultaneous determination of PDFs and SMEFT couplings

Existing studies on PDF and SMEFT interplay

- So far, there have been two studies into the simultaneous determination of PDFs and SMEFT couplings:
  - *Can New Physics Hide Inside the Proton?*, 2019, arXiv:1905.05215 (Carrazza, Degrande, Iranipour, Rojo, Ubiali). Proof-of-concept study based on four four-fermion operators in DIS.
  - *Parton distributions in the SMEFT from high-energy Drell-Yan tails*, 2021, arXiv:2104.02723 (Greljo, Iranipour, Madigan, Moore, Rojo, Ubiali, Voisey). Study based on $\hat{W}$, $\hat{Y}$ operators (and an additional operator, which we omit for time reasons) and high-energy Drell-Yan data, including projections for bounds when new high-luminosity data is available.
- Both studies based on the same 'standard' methodology (with small technical differences in how SMEFT sector is implemented).

Existing studies on PDF and SMEFT interplay

- To simultaneously fit PDFs and SMEFT parameters with the 'standard method', we do the following:

1. Pick a grid of 'benchmark points' in SMEFT parameter space, $\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_n$.

2. For each benchmark point $\mathbf{a}_i$, perform a PDF fit using the standard NNPDF methodology with the SMEFT parameters fixed to the values $\mathbf{a}_i$.

3. Record the χ^2 goodness-of-fit statistic of the PDF to the data at each point. Interpolate the χ^2 using an appropriate hypersurface (this is just a curve for one SMEFT parameter) and use this surface to derive bounds on the SMEFT couplings.

PDFs in the SMEFT from high-energy DY tails

- In this study, the focus was instead on the $\hat{W}$, $\hat{Y}$ operators, which arise as EFT corrections to electroweak gauge-boson self-energy and have an enhanced effect in high-energy Drell-Yan data.

$(\bar{L}L)(\bar{L}L)$		$(\bar{R}R)(\bar{R}R)$		$(\bar{L}L)(\bar{R}R)$	
Q_{ll}	$(\bar{l}_p\gamma_\mu l_r)(\bar{l}_s\gamma^\mu l_t)$	Q_{ee}	$(\bar{e}_p\gamma_\mu e_r)(\bar{e}_s\gamma^\mu e_t)$	Q_{le}	$(\bar{l}_p\gamma_\mu l_r)(\bar{e}_s\gamma^\mu e_t)$
$Q_{qq}^{(1)}$	$(\bar{q}_p\gamma_\mu q_r)(\bar{q}_s\gamma^\mu q_t)$	Q_{uu}	$(\bar{u}_p\gamma_\mu u_r)(\bar{u}_s\gamma^\mu u_t)$	Q_{lu}	$(\bar{l}_p\gamma_\mu l_r)(\bar{u}_s\gamma^\mu u_t)$
$Q_{qq}^{(3)}$	$(\bar{q}_p\gamma_\mu\tau^I q_r)(\bar{q}_s\gamma^\mu\tau^I q_t)$	Q_{dd}	$(\bar{d}_p\gamma_\mu d_r)(\bar{d}_s\gamma^\mu d_t)$	Q_{ld}	$(\bar{l}_p\gamma_\mu l_r)(\bar{d}_s\gamma^\mu d_t)$
$Q_{lq}^{(1)}$	$(\bar{l}_p\gamma_\mu l_r)(\bar{q}_s\gamma^\mu q_t)$	Q_{eu}	$(\bar{e}_p\gamma_\mu e_r)(\bar{u}_s\gamma^\mu u_t)$	Q_{qe}	$(\bar{q}_p\gamma_\mu q_r)(\bar{e}_s\gamma^\mu e_t)$
$Q_{lq}^{(3)}$	$(\bar{l}_p\gamma_\mu\tau^I l_r)(\bar{q}_s\gamma^\mu\tau^I q_t)$	Q_{ed}	$(\bar{e}_p\gamma_\mu e_r)(\bar{d}_s\gamma^\mu d_t)$	$Q_{qu}^{(1)}$	$(\bar{q}_p\gamma_\mu q_r)(\bar{u}_s\gamma^\mu u_t)$
		$Q_{ud}^{(1)}$	$(\bar{u}_p\gamma_\mu u_r)(\bar{d}_s\gamma^\mu d_t)$	$Q_{qu}^{(8)}$	$(\bar{q}_p\gamma_\mu T^A q_r)(\bar{u}_s\gamma^\mu T^A u_t)$
		$Q_{ud}^{(8)}$	$(\bar{u}_p\gamma_\mu T^A u_r)(\bar{d}_s\gamma^\mu T^A d_t)$	$Q_{qd}^{(1)}$	$(\bar{q}_p\gamma_\mu q_r)(\bar{d}_s\gamma^\mu d_t)$
				$Q_{qd}^{(8)}$	$(\bar{q}_p\gamma_\mu T^A q_r)(\bar{d}_s\gamma^\mu T^A d_t)$

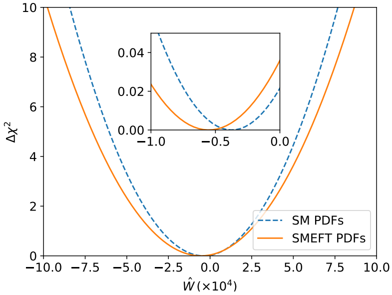
$(\bar{L}R)(\bar{R}L)$ and $(\bar{L}R)(\bar{L}R)$		B-violating	
Q_{ledq}	$(\bar{l}_p^j e_r)(\bar{d}_s q_t^j)$	Q_{duq}	$\varepsilon^{\alpha\beta\gamma}\varepsilon_{jk}\left[(d_p^\alpha)^T C u_r^\beta\right]\left[(q_s^{\gamma j})^T C l_t^k\right]$
$Q_{quqd}^{(1)}$	$(\bar{q}_p^j u_r)\varepsilon_{jk}(\bar{q}_s^k d_t)$	Q_{qqu}	$\varepsilon^{\alpha\beta\gamma}\varepsilon_{jk}\left[(q_p^{\alpha j})^T C q_r^{\beta k}\right]\left[(u_s^\gamma)^T C e_t\right]$
$Q_{quqd}^{(8)}$	$(\bar{q}_p^j T^A u_r)\varepsilon_{jk}(\bar{q}_s^k T^A d_t)$	Q_{qqq}	$\varepsilon^{\alpha\beta\gamma}\varepsilon_{jn}\varepsilon_{km}\left[(q_p^{\alpha j})^T C q_r^{\beta k}\right]\left[(q_s^{\gamma m})^T C l_t^n\right]$
$Q_{lequ}^{(1)}$	$(\bar{l}_p^j e_r)\varepsilon_{jk}(\bar{q}_s^k u_t)$	Q_{duu}	$\varepsilon^{\alpha\beta\gamma}\left[(d_p^\alpha)^T C u_r^\beta\right]\left[(u_s^\gamma)^T C e_t\right]$
$Q_{lequ}^{(3)}$	$(\bar{l}_p^j\sigma_{\mu\nu} e_r)\varepsilon_{jk}(\bar{q}_s^k\sigma^{\mu\nu} u_t)$		

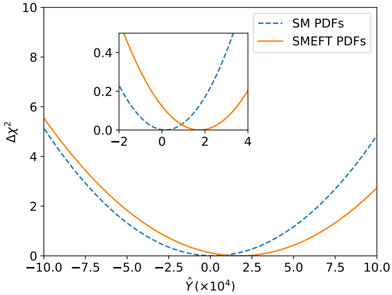
PDFs in the SMEFT from high-energy DY tails

- Bounds initially derived on $\hat{W}$, $\hat{Y}$ using existing data:
  - DIS-only data
  - Drell-Yan data standard to PDF sets
  - New high-mass Drell-Yan data implemented for this study
- Shown explicitly that SMEFT corrections to high-mass DY predictions dominated, but SMEFT effects treated consistently in DIS data too.

PDFs in the SMEFT from high-energy DY tails

- Resulting χ^2 parabolas given below for fixed SM PDFs and simultaneous fits:

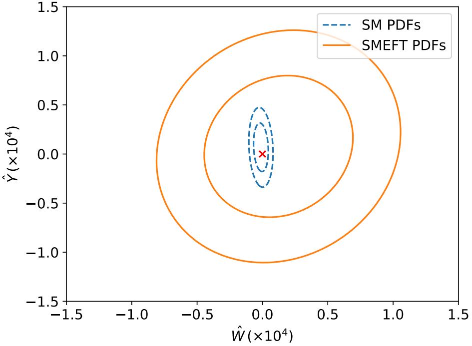




- $\Rightarrow$ bounds change! Roughly $\sim 15\%$ change in size of bounds.

PDFs in the SMEFT from high-energy DY tails

- More pronounced effect when projections are taken into account for the high-luminosity phase of the LHC, for energies of 14 TeV and luminosities of 6 ab^{-1}.
- Below: change in 68%, 95% bounds.



- $\Rightarrow$ huge change! Around $\sim 700\%$ for $\hat{W}$, $\sim 100\%$ for $\hat{Y}$!

Efficient simultaneous determination of PDFs and SMEFT couplings

Efficient simultaneous determination?

- Standard approach $\Rightarrow$ BSM bounds can be affected by consistent simultaneous fits with PDFs, effect will grow in future.

- **Problem:** Standard approach very inefficient! Leads to new question:

  Is there an efficient method to simultaneously determine PDFs and BSM parameters?

Efficient simultaneous determination?

- **Proposal:** *Linearise* the deviation of the SMEFT PDF from the naïve SM PDF:

$$\Delta \mathbf{f}(x, Q^2) = \mathbf{f}^{\text{SMEFT}}(x, Q^2) - \mathbf{f}^{\text{SM}}(x, Q^2) = \sum_{i=1}^{N} w_i \mathbf{h}_i(x, Q^2),$$

  where $w_i \in \mathbb{R}$ are parameters called *weights* and $\mathbf{h}_i$ are some suitable basis functions.

- The basis functions should be chosen to satisfy some key theory properties:
  1. Both $\mathbf{f}^{\text{SMEFT}}$ and $\mathbf{f}^{\text{SM}}$ satisfy the DGLAP equations, so $\mathbf{h}_i$ should also satisfy DGLAP equations by linearity.
  2. PDF sum rules imply that $\mathbf{h}_i$ should obey some non-trivial integral relations.

  Conditions (1) and (2) are met by taking $\mathbf{h}_i$ to be a *difference of existing PDF replicas*.

Efficient simultaneous determination?

- For example, we can take the functional form:

$$\mathbf{f}_j^{\text{SMEFT}} = \mathbf{f}_j^{\text{SM}} + \sum_{i=1}^{N} w_{i,j}(\mathbf{f}_i^{\text{SM}} - \mathbf{f}_j^{\text{SM}}),$$

  for the jth replica of the SMEFT ensemble. This should be thought of as an 'expansion of the jth SMEFT replica about the jth SM replica in a basis of PDF differences'.

- Using above, can be shown predictions take the form:

$$\sigma = \sigma^{\text{SM}} + \mathbf{Pw} + \mathbf{Qa},$$

  where $\mathbf{P}, \mathbf{Q}$ are constant matrices, $\mathbf{w}$ is the vector of weights for that replica, and $\mathbf{a}$ is the vector of SMEFT couplings.

- Linearisation requires neglecting terms of order $O(\mathbf{a} \cdot \Delta\mathbf{f})$.

Efficient simultaneous determination?

- Thus we have linearised the problem of simultaneous determination. The form:

$$\sigma = \sigma^{\text{SM}} + \mathbf{P}\mathbf{w} + \mathbf{Q}\mathbf{a},$$

makes it clear that this is a simultaneous determination of PDFs (through weights **w**) and SMEFT parameters (**a**), where a change in one can be compensated by a change in the other.

- When inserted into the χ^2 formula, all we need to do is to minimise a quadratic, which can be done *analytically* - extremely fast!

Efficient simultaneous determination?

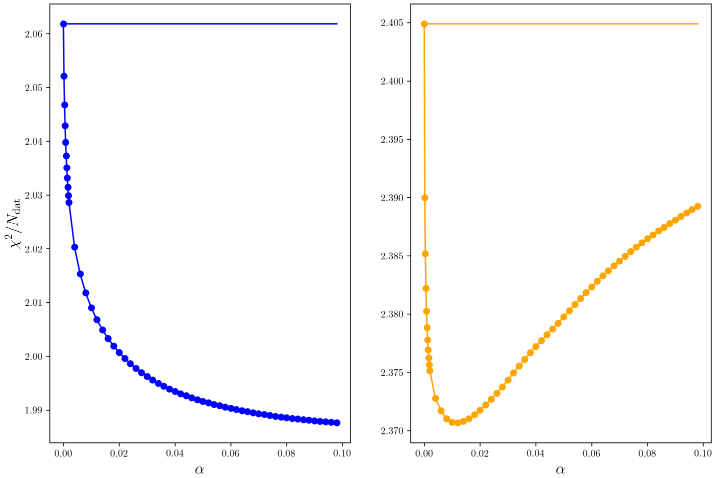
- *However*, naïve analytic minimisation can result in *overfitting* of PDFs.
- More weights ⇒ more PDF freedom ⇒ can overfit. Need to constrain size of weight space to avoid this.
- This can be achieved by a *hyperoptimisation procedure*. We introduce a regulator α into the χ^2 statistic given by:

$$\chi^2 \mapsto \chi^2 + \frac{1}{\alpha}\mathbf{w}^T\mathbf{w}.$$

  As the regulator α decreases close to 0, the weights become increasingly penalised if they are too large. Thus the regulator α limits the effective size of the space that the weights span.

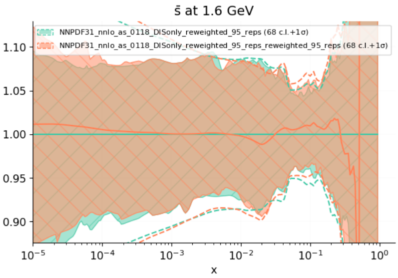
Efficient simultaneous determination?

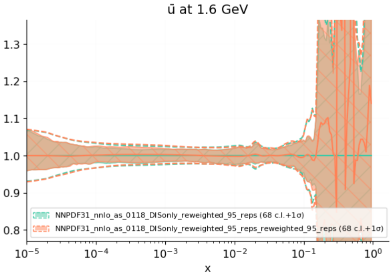
- Optimal value of α found by *hyperoptimisation*. Pseudodata split into training/validation sets and χ^2 monitored on both:



Results so far

- This method has undergone significant revision since its initial proposal. We have confirmed so far that:
  - When the SMEFT couplings are set to zero, the method reproduces the SM PDFs, so is self-consistent (note that this is not guaranteed without input from the hyperoptimisation procedure).

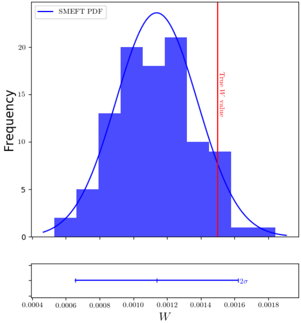




Preliminary plots: not necessarily final.

Results so far

- This method has undergone significant revision since its initial proposal. We have confirmed so far that:
  - When we make fake data based on fixed, known SMEFT parameters, the method is able to return bounds enclosing the known values.



Preliminary plot: not necessarily final.

Still to come

- Benchmark new method against old studies - see if bounds are consistent with those found previously.

- After that, can consider much more ambitious PDF-EFT interplay studies, with much larger numbers of operators!

Questions?