On determinations of the strong coupling constant from hadronic data

High Energy Physics Seminar, Cavendish Laboratory

Zahari Kassabov

June 12th 2018

Cavendish Laboratory, University of Cambridge

Introduction: Theoretical uncertainties at the LHC

The main uncertainties in a theoretical calculation of a process at the LHC:

- Missing higher orders uncertainties (MHOU).
- Uncertainties on the PDF.
- Uncertainties on α_S.

Example: *ggH* :

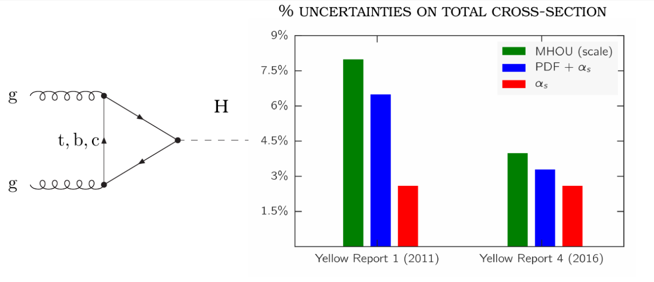


Figure 1: (S.Forte, Lattice 2017)

State of the affairs

- Difficult to provide a statistical interpretation to all three of these uncertainties. Rely on a decision *by committee*.
- Improvements driven by:
  - Better calculations: Generally decrease uncertainty.
  - More reliable methodologies: Can decrease or increase uncertainty.
- In general, recent improvements in MHOU and PDF, not in α_S.
- Reliable theory (and th uncertainties) crucial for LHC programme.

Estimation of Missing Higher Order Uncertainties

MHOU: Usually estimated trough scale variations.

- Normative source: *Yellow Report 4*.
- Central scale setting can make a large difference [e.g jets at NNLO, Currie, et al arXiv:1704.00923].
- Size of the scale variation taken by convention (usually 2 and $\frac{1}{2}$ for μ_r and μ_f).
- Procedure to combine the two scale variations.
  - E.g. $\Delta_{\text{MHOU}} = \left| \frac{[\sigma(2\mu_f, 2\mu_r) - \sigma(\frac{1}{2}\mu_f, \frac{1}{2}\mu_r)]}{2} \right|$ vs max – min of 7-point variation.
  - Uniform vs Gaussian distribution.
- Correlations in multi-process problems: Not well studied.

Improvements driven by calculating higher orders [e.g. *ggH* at N^3LO, Anastasiou et al, arXiv:1602.00695].

PDFs and PDF uncertainties

- Normative source: PDF4LHC Recommendation [arxiv:1510.03865].
- PDF uncertainties propagate experimental uncertainties to best fit of unbiased interpolants. MHOU not considered at all!
- Improvements driven by methodology, new data and theoretical calculations. Leads to agreement between groups.

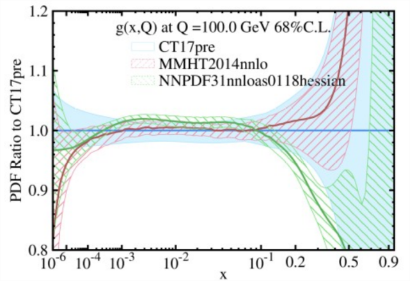


Figure 2: Tie-Jiun Hou, DIS 2018 (Unpublished)

Status of α_S

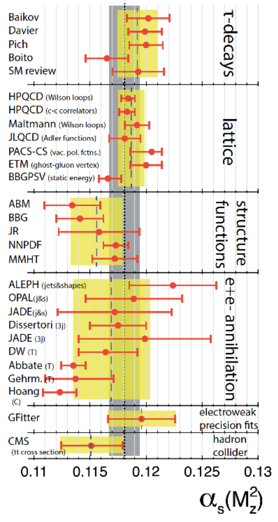
- Normative source: PDG combination [PDG review 2016, Bethke, Dissertori, Salam].
- *World average* on several determinations. Philosophy:

  "...be as neutral as possible with regard to disputes in the community about different determinations, with uniform prescriptions applied to all reasonable determinations."

  Salam, arxiv:1712.05165

- Current value
$$\alpha_S(M_Z^2) = 0.1181 \pm 0.0013$$
- No updates in 2017 and 2018.

PDG combination



- Only NNLO or better determinations considered.
- "Pre averaging": Take the unweighted mean and the mean error from each process.
- Final number obtained as a weighted "χ^2 average" over the processes.

Other viewpoints on α_S determination

"In my opinion one should select few theoretically simplest processes for measuring α_S and consider all other ways as tests of the theory."

G.Altarelli, 2013

Other viewpoints on α_S determination

"In my opinion one should select few theoretically simplest processes for measuring α_S and consider all other ways as tests of the theory."

G.Altarelli, 2013

- Identifying most reliable determinations largely a sociological problem.

Other viewpoints on α_S determination

"In my opinion one should select few theoretically simplest processes for measuring α_S and consider all other ways as tests of the theory."

G.Altarelli, 2013

- Identifying most reliable determinations largely a sociological problem.
- PDG world average can be improved by:
    - Updates to individual determinations.
    - Identifying flawed determinations (potentially bigger effect).

Other viewpoints on α_S determination

"In my opinion one should select few theoretically simplest processes for measuring α_S and consider all other ways as tests of the theory."

G.Altarelli, 2013

- Identifying most reliable determinations largely a sociological problem.
- PDG world average can be improved by:
  - Updates to individual determinations.
  - Identifying flawed determinations (potentially bigger effect).
- How reliable is the determination of α_S from PDFs?

α_S from PDFs

- Basic concept: Find best fit in the combined space of (α_s, PDF) of a large body of experimental data.
- Compared to Electroweak precision fit [Blas et al, arXiv:1608.01509], the larger dataset implies:
  - More challenging to achieve a good description of all processes (e.g. low scale DIS affected by both MHOU and higher twists).
  - Theoretical problems likely to average out to some extend [Carrazza, Forte, ZK, Rojo, Rottoli, arxiv:1803.07977].
  - More data and more dependence on α_S implies more precision.

Challenges extracting α_S from PDFs

- PDF parametrization may bias the α_S value.
- Correct treatment of experimental systematics (particularly normalization uncertainties).
- Hidden uncertainties in theoretical description of PDFs (e.g. heavy quark treatment).
- Inclusion of PDF uncertainty in the α_S determination.

- Accurate estimation of missing higher order uncertainties

Challenges extracting α_S from PDFs

- PDF parametrization may bias the α_S value. [NNPDF]
- Correct treatment of experimental systematics (particularly normalization uncertainties). [NNPDF]
- Hidden uncertainties in theoretical description of PDFs (e.g. heavy quark treatment). [NNPDF 3.1]
- Inclusion of PDF uncertainty in the α_S determination. [NNPDF 3.1-α_S]
- Accurate estimation of missing higher order uncertainties [TO DO]

Made notable progress since latest NNPDF determination in 2011.

This talk

- α_S determination based on NNPDF3.1
- Can collider determinations of α_S be really independent on those from PDFs?
- Can we *measure* theoretical uncertainties?

α_S determination based on NNPDF3.1 in summary

Precision determination of the strong coupling constant within a global PDF analysis

NNPDF Collaboration, arxiv:1802.03398

- 3979 data points.
- Including simultaneously differential top, *Z pT* and inclusive jet data for the first time.
- Exact NNLO theory
- Method to effectively fit α_S and the PDF simultaneously (PDFs effectively nuisance parameters).

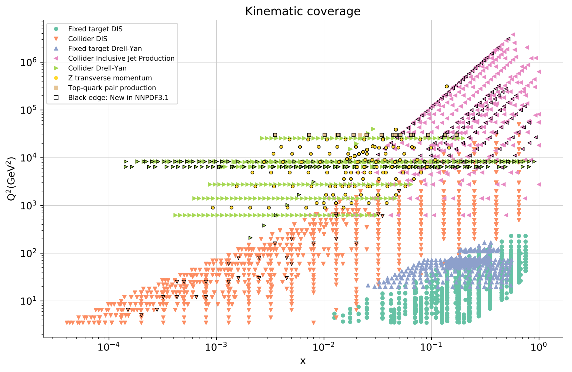
α_S determination based on NNPDF3.1 in summary

Precision determination of the strong coupling constant within a global ~~PDF~~ QCD analysis

NNPDF Collaboration, arxiv:1802.03398

- 3979 data points.
- Including simultaneously differential top, *Z pT* and inclusive jet data for the first time.
- Exact NNLO theory
- Method to effectively fit α_S and the PDF simultaneously (PDFs effectively nuisance parameters).

Dataset in NNPDF 3.1



α_S dependence though:

- PDF evolution (large scale differences advantageous).
- Direct dependence (e.g. $t\bar{t}$ data)
- Higher order QCD corrections.

Propagating uncertainties in NNPDF

Our data has experimental uncertainties.

- We view the data as random variables from the distribution $\mathcal{N}(d_i, \Sigma_{i,j})$
  - d_i is the experimentally measured central value for the point i.
  - Σ_{ij} *a* covariance between the points i, j.
- We sample N_{rep} datasets (Monte Carlo pseudodata) from the distribution, and train neural networks "replicas"s to each dataset to minimize an error function χ^2.
- PDF dependent quantities are calculated from statistics over the ensemble of replicas. E.g. "*PDF uncertainty*" is usually the standard deviation over the replicas.

Simultaneous minimization of (α_S, PDF)

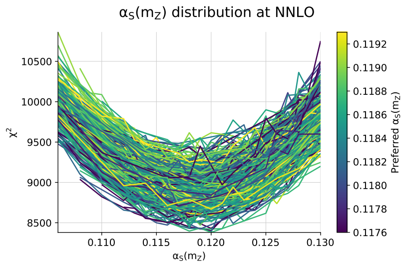
- Ideally would minimize simultaneously α_s and the PDF parameters in each replica.
  - Can't do easily because we use precomputed tables that depend on α_s.

Simultaneous minimization of (α_S, PDF)

- Ideally would minimize simultaneously α_s and the PDF parameters in each replica.
  - Can't do easily because we use precomputed tables that depend on α_s.
  - Solution: Repeat the fit for discrete values of α_S to the same data replica$\longrightarrow$ c-replica.

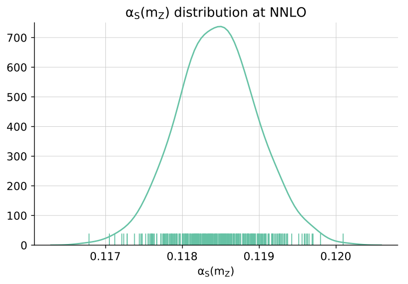
Simultaneous minimization of (α_S, PDF)

- Ideally would minimize simultaneously α_s and the PDF parameters in each replica.
  - Can't do easily because we use precomputed tables that depend on α_s.
  - Solution: Repeat the fit for discrete values of α_S to the same data replica$\longrightarrow$ c-replica.
- Each c-replica has a $\chi^2(\alpha_S)$ profile.



Simultaneous minimization of (α_S, PDF)

- Ideally would minimize simultaneously α_s and the PDF parameters in each replica.
  - Can't do easily because we use precomputed tables that depend on α_s.
  - Solution: Repeat the fit for discrete values of α_S to the same data replica$\longrightarrow$ c-replica.
- Each c-replica has a $\chi^2(\alpha_S)$ profile.
- Each minimum yields one sampled α_S value.



The result

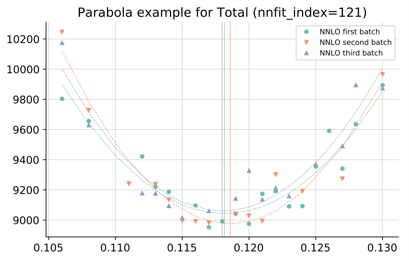
We obtain

$$\alpha_S^{\mathrm{NNLO}}(M_Z) = 0.11845 \pm 0.00052^{(\mathrm{exp})}(0.4\%)$$

- Experimental uncertainty is the standard deviation over the ensemble of c-replicas of the $\arg\min_{\alpha_S} \chi^2(\alpha_S)$.
- Very small uncertainty (compared to other determinations in the PDG). But need to assess:
  - Methodological systematics.
  - Theory uncertainties.

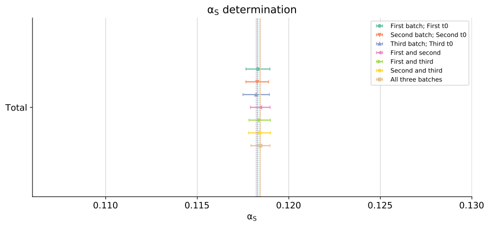
Improvement by batch minimization

- The NNPDF methodology discards replicas that do not meet certain convergence criteria (25-30%).
- $\chi^2(\alpha_S)$ noisy for each c-replica. Fit quality depends on:
  - Initial conditions.
  - Cross validation split
- Solution: Repeat each fit three times, and take the minimum (requiring at least 2 successful replicas).



Effect of batch minimization

- Small on central value $\sim O(0.4) \times \Delta^{(\mathrm{exp})}$.
- Significative on uncertainty. Reduction to up to 27%.
- Convergence already good with two batches.



Finite-size uncertainties

We estimate the uncertainties due to fitting a finite number of replicas by bootstrapping.

1. Take the set of N minima.
   1.1 Sample with replacement from it M sets of N values, with M large.
   1.2 Compute the M means of each of the M sets.
   1.3 Compute the standard deviation of the M means.

$$\Delta_{\text{stat}} = 3 \times 10^{-5}$$

Effect negligible compared to the PDF uncertainty.

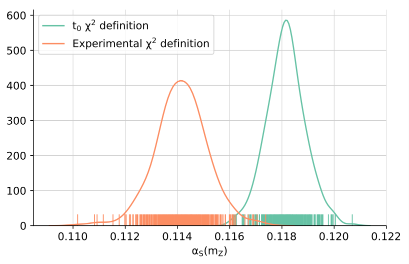
Curve selection

- Select c-replicas that contain at least N_{min} successfully fitted values.
- Choose N_{min} to minimize the bootstrapping uncertainty.
    - Tradeoff between statistics and quality of the parabolas.
- Result stable for reasonable values N_{min}.

N_{min}	$\alpha_s(m_Z)$	N_{rep}	Δ_{α_s}
18	0.11842 ± 0.00031 (0.3%)	12	0.00009
15	0.11844 ± 0.00044 (0.4%)	92	0.00005
6	$\mathbf{0.11845 \pm 0.00052}$ **(0.5%)**	**379**	**0.00003**
3	0.11844 ± 0.00056 (0.5%)	400	0.00003

Treatment of normalization uncertainties

- It is well known [D'Agostini, 2003] that minimizing the experimental χ^2 is biased in the presence of normalization uncertainties.
  - Maximum likelihood estimator is not an unbiased estimator.
- NNPDF uses the t_0 procedure [arxiv:0912.2276] (fix normalization from the result of a previous fit).
- α_S fit heavily biased when the experimental definition of χ^2 is used inconsistently to minimize α_S.



Final methodological uncertainty

- Checked many other possible sources of systematics.
  - Assumption that $\chi^2(\alpha_s)$ is parabolic.
  - Dependence on the extreme values.
  - Myriad of variations of replica selection.
  - Effect of t_0 procedure.
- Overall all effects much smaller than experimental uncertainty. Methodological effects conservatively estimated at $\sim 10^{-4}(0.09\%)$.

Theoretical uncertainties that are under control

- Theoretical effects necessarily have to affect the PDFs in order to affect our determination.
- But not sufficient: Dependence on $\alpha_S(M_Z)$ required.
- Many have been studied and found to be smaller than experimental uncertainties:

Higher twist found to be small compared to experimental uncertainties [arxiv:1303.1189].

Charm mass greatly improved by parameterizing the charm PDF [arXiv:1605.06515].

Electroweak corrections Kept under control with suitable cuts [arxiv:1706.00428].

Nuclear corrections Studied in [arxiv:1706.00428] and found to be small compared to experimental uncertainties.

In conclusion MHOU highly likely to be the dominant theoretical uncertainty.

Estimates of MHOU

No reliable method known for PDF based quantities. We have:

$$\alpha_S^{\mathrm{NNLO}}(M_Z) = 0.1184 \pm 5 \times 10^{-4}$$

$$\alpha_S^{\mathrm{NLO}}(M_Z) = 0.1206 \pm 6 \times 10^{-4}$$

- PDF fits with scale variations currently in early stage.
- Methods based on continuation of perturbative series such as Cacciari-Houdeau [arXiv:1105.5152] hampered by:
  - Poor fit quality at LO and even at NLO.
  - Lack of unique series expansion (many processes involved).
- CH yields a theoretical uncertainty of 4×10^{-4}, smaller that the experimental ones. Likely too optimistic.

Final MHOU estimate

For lack of better options, make it be crude and conservative.

$$\Delta\alpha_S^{\text{th}} = \frac{1}{2}\left|\alpha_S^{\text{NLO}} - \alpha_S^{\text{NNLO}}\right| = 0.0011(0.9\%)$$

- Uncertainty likely overblown by the poor fit quality at NLO.

	$\min_{\alpha_S} \chi^2(\alpha_S^{(\text{central})})/N_{\text{data}}$
NLO	5014/3979=1.26
NNLO	4814/3979=1.21

- Two times bigger than the experimental uncertainty.

There must be a better way! A data driven way?

Let us make the following assumptions:

- Data entering the fit can be separated in *processes* where MHOU are largely uncorrelated.
  - In practice correlations expected because of the PDF.
  - Also must see how to deal with experimental correlations.
- One can define a preferred value of α_S for each process somehow.
- There are sufficient number of processes to make statistics.
- Differences in preferred values above experimental uncertainties are due to theoretical differences.

It follows that theoretical uncertainties can be estimated trough the dispersion of the preferred values.

- Strong assumptions, but is it really worse than scale variations?

How NOT to define preferred values: The partial χ^2

The global χ^2 minimizes to fit (α_S, PDF) looks like

$$\chi^2\left[\{\theta\}, \alpha_S, \mathcal{D}\right] = \sum_{I,J=1}^{N_{\mathcal{D}}} \left(T_I[\{\theta\}, \alpha_S] - D_I\right) C_{IJ}^{-1} \left(T_J[\{\theta\}, \alpha_S] - D_J\right)$$

where

- $\{\theta\}$ are the parameters of the PDF.
- $\mathcal{D}$ is the set of $N_{\mathcal{D}}$ data points entering the fit.
- T_I is the theoretical prediction for the data point indexed by I.
- D_I is the experimentally measured value.
- C_{IJ} measures the experimental covariance.

The profile of a c-replica is

$$\chi^2(\alpha_s) = \min_{\{\theta\}} \chi^2\left[\{\theta\}, \alpha_S, \mathcal{D}\right]$$

Definition of partial χ^2

The partial χ^2 is:

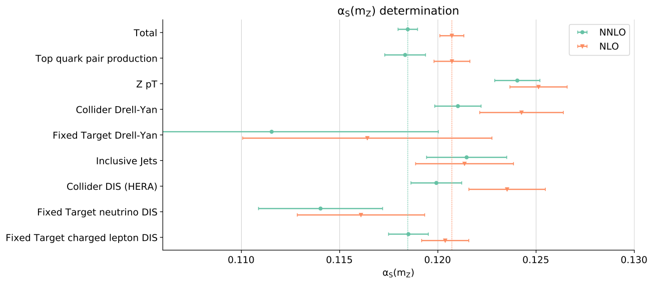
$$\chi^2_p\left[\{\theta\}, \alpha_S, \mathcal{P}\right] = \sum_{I,J=1}^{N_{\mathcal{P}}} \left(T_I[\{\theta\}, \alpha_S] - D_I\right) C^{-1}_{IJ} \left(T_J[\{\theta\}, \alpha_S] - D_J\right)$$

where we have replaced $\mathcal{D}$ with $\mathcal{P}$, a subset of $\mathcal{D}$ with $N_{\mathcal{P}}$ points, and neglected the correlations between the points in $\mathcal{P}$ and $\mathcal{D}$.

- Up to missing correlations, for a set of processes that cover all $\mathcal{D}$

$$\chi^2(\alpha_s) = \sum_p \chi^2_p(\alpha_s)$$

Results from partial χ^2



- Define

$$\alpha_S^p = \arg\min_{\alpha_S} \chi_p^2(\alpha_S)$$

- Useful to estimate pulls qualitatively. E.g. can say that the LHC data likely contributes to increase $\alpha_S(M_Z)$.
- But quantitative calculations are seriously off.

Partial χ^2 and α_S determination from hadronic processes

> *By "collider determinations" I mean determinations of α_S based on cross sections measured at hadron–hadron and hadron–lepton colliders that are used to constrain the strong coupling independently of a PDF fit.*
>
> *G.Salam [arxiv:1712.05165]*

However all "collider determinations" in practice amount to the partial χ^2 minimization above. Collider determinations depend on the PDF in two ways:

- Best fit PDF changes strongly with α_S, i.e. $\{\theta\}(\alpha_S) = \arg\min_{\{\theta\}} \chi^2(\alpha_s, \{\theta\}, \mathcal{D})$.
- Global fit quality $\chi^2(\alpha_S)$ changes strongly with α_S.

Collider determinations

- Notable example CMS measurement of $t\bar{t}$ [arXiv:1307.1907] computed at NNLO [Czakon, Fiedler, Mitov, arXiv:1303.6254] included in the PDG average as the only item in the collider determination category.
- Many others have appeared recently. For example, a determination using jets in DIS at NNLO [H1 Collaboration arxiv:1709.07251] at. The result, using NNPDF3.1 sets, is $\alpha_S = 0.1157 \pm 0.002^{\mathrm{exp}} \pm 0.003^{\mathrm{th}}$. The central value is discarded by the NNPDF3.1 determination at 5σ.

Problems with the partial χ^2

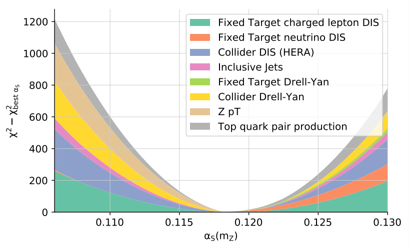
- Neglecting correlations is a bad approximation.

$$\alpha_S^{(\mathrm{global})} - \arg\min_{\alpha_S} \sum_p \chi_p^2(\alpha_S) \sim \mathrm{std}_p \alpha_S^p$$

  Deviation of the order of the quantity we wanted to estimate.
- By construction it doesn't take into account the global fit quality [Z.K, arxiv:1802.05236].
- Preferred values extracted this way amount to a logical contradiction:
  - Choose a point in the phase space $(\alpha_S^p, \underbrace{\arg\min_{\{\theta\}} \chi^2(\{\theta\}, \alpha_s^p, \mathcal{D})}_{\text{best fit PDF at } \alpha_S^p})$,
  - discarded by the data $(\mathcal{D})$,
  - used to construct constrain the PDF parameters $\{\theta\}$,
  - on which α_s^p relies in the first place.

Fit results outside the global best fit

- χ^2_p depends strongly on the rest of the data entering the fit.
- χ^2_p inversely correlated to relative weight of $\mathcal{P}$ (given by the number of points). The more weight $\longrightarrow$ the more advantageous it is to optimize for it at the expense of some other data $\longrightarrow$ the smaller χ^2_p.
- For example, note $t\bar{t}$. 26 points (0.6% of the total) are only described simultaneously with the rest of the data in a small range of α_S values.



Importance of simultaneous minimization

Can it be that for processes like $t\bar{t}$, $\chi_p(\alpha_S)$ changes much more quickly than the global $\chi^2(\alpha_S)$. Made an experiment to test it:

- The further away from the global minimum, the larger the space of PDs with the same $\chi^2(\alpha_S)$.
- Can alter the fit in a way that changes substantially χ_p^2 and very little the global χ^2.
- Set up:
  - Take the 26 points of $t\bar{t}$ data and copy them 15 times in a fit at $\alpha_S(M_Z) = 0.121$. The target function becomes $\chi'^2 = \chi^2 + 14\chi^2_{t\bar{t}}$.
  - Now $t\bar{t}$ has more weight in the fit.

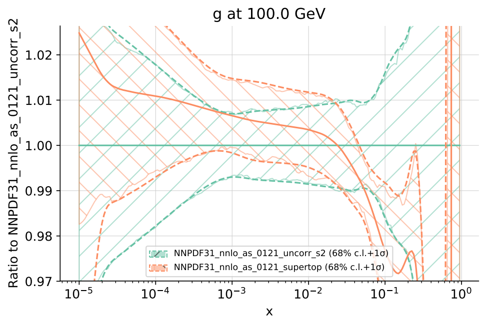
Results of reweighed fit

$\chi^2/d.o.f.$	$\alpha_s = 0.118$	$\alpha_s = 0.121$ default	$\alpha_s = 0.121$ weighted $\bar{t}t$
Total	1.162	1.212	1.228
$t\bar{t}$	1.07	1.42	1.02

- Very small change in the total χ^2.
- Can bend the PDF to describe the top data perfectly at a larger value of α_S.

Changes in the PDFs

- Bigger α_S compensate with smaller gluon PDF in the relevant kinematic region.



- Still compatible within uncertainties.
- Rest of PDFs largely unchanged.
- Conclusion: $t\bar{t}$ determination not independent on PDFs.

An even more convincing example

- It turns out that in our analysis the minimum for $\chi^2_{t\bar{t}}$ $(\alpha_S(M_Z) = 0.1183)$ is very close to the global minimum $\chi^2(\alpha_S(M_Z) = 0.11845)$.
- Selecting a different process, we can engineer a fit where both χ^2_p and χ^2 are lower than for the minimum of the partial χ^2.
- For $Z\,pT$ we have $\alpha_S^{Z\,pT} \approx 0.124$ in the default fit.
- Assign the 120 points of $Z\,pT$ data a weight of 32 (this makes it weight as much as the rest of the data).

Results of $Z\,pT$ weighted fits

$\chi^2/d.o.f.$	$\alpha_s = 0.120$ weighted $Z\,pT$	$\alpha_s = 0.124$ default
Total	1.226	1.281
$Z\,pT$	0.94	1.11

The weighted fit agrees better both with the whole ensemble $\mathcal{D}$ and the $Z\,pT$ data. It is therefore a better value of α_S from $Z\,pT$, no matter how you look at it.

- I consider this a proof that the partial χ^2 minimization entails a logical contradiction, since in principle it is possible to find a better fit.

Coming back to theoretical uncertainties

- We could exploit similar weighting strategy to define dataset dispersion [Z.K. arxiv:1802.05236].
- Associate a large enough weight to each dataset to force $\chi^2_p(\alpha_S)/d.o.f. \simeq 1$ (statistical minimum when cross validation is applied) in a large range of α_S.
- Define preferred value as the one where the rest of the data agrees the best, i.e. $\alpha_S^{(\text{preferred } p)} = \min_{\alpha_s} \chi^2(\alpha_S)^{(\text{weighted } p)}$.

Advantages of definition of preferred value

- Ideally defined as best fit to all the data, restricted to describing perfectly the specific dataset (weight is an implementation detail).
- Explicitly depends on all the data in the problem $\mathcal{D}$.
- Dispersion of preferred values over datasets (above experimental uncertainties) can be interpreted as mainly coming from theory.
- Possible formula

$$\Delta^{th} = \Delta^{exp} \frac{1}{N_p - 1} \sqrt{\sum_p \frac{\left(\alpha_S - \alpha_S^{(\text{preferred } p)}\right)^2}{\frac{1}{2}\left(\Delta^{2,exp} + \Delta^{2(exp,\text{preferred},p)}\right)}}$$

Conclusions

- Determinations of α_S from global QCD fits (i.e. PDFs) have interesting characteristics to constrain its value.
- “Collider determinations” are flawed, and not independent on PDFs. Should not be used for World averages.
- It is interesting to explore data driven methods to study theory uncertainties, e.g. the dataset dispersion outlined here.

Thank you!